A History of Engineering and Science in the Bell System

Transmission Technology (1925-1975)

Prepared by Members of the Technical Staff,
AT&T Bell Laboratories.

E. F. O'Neill, Editor.

AT&T Bell Laboratories

Credits for figures taken from
other than Bell System sources
appear on page 795.

Inquiries may be made to the following address:
AT&T Technologies
Commercial Sales Clerk
Select Code 500-473
PO Box 19901
Indianapolis, Indiana 46219
1-800-432-6600

First Printing, 1985

International Standard Book Number: 0-932764-08-8

Library of Congress Catalog Card Number: 85-70382

Printed in the United States of America

Contents

Foreword

Society's ability to transport its needs from a source to the point of use is an important measure of its progress. Our standard of living has been improved through advances in speed, safety, and economy of the transport of people, food, fuel, and information. This volume focuses on the last of these—transmission, the transport of information, as it evolved in the Bell System during the middle years of the 20th century.

This is a chapter in an ongoing story. The first half-century of telephone transmission witnessed the conquest of distance; by the late 1920s calls could be placed to all points of the United States and to Europe and South America as well. The period covered in this volume was the era of the conquest of expense. The cost of transmission was so dramatically reduced by a variety of technologies that calls could be placed to almost any location within the means of most Americans. There is every reason to expect that the future of transmission will be equally exciting and important as the technology continues to advance and the type and amount of information to be carried expands.

The history of transmission development is typical of the progress of technology in general. It is significant that no single transport method completely dominated, but rather that each method was adapted to optimize the need it was to fulfill. Also significant is the fact that some technologies did not achieve deployment because they were overtaken by the rapid advance of more effective techniques. An outstanding example is the millimeter waveguide, almost completely developed after decades of work, being superseded by fiber optics. Technology does not progress smoothly in a completely predictable and inevitable sequence of steps, but still involves some trial and error. Each attempt, however, leaves a legacy of knowledge gained as a firmer basis for the next advance.

While the transport of information was not dominated by a single technology in the past, the future form of transmission appears certain to be predominately digital. Although for more than two-thirds of the period covered by this volume analog transmission was the exclusive method, beginning in 1962 with the introduction of T-Carrier, the digital revolution started. Until the next breakthrough in transmission methods (which, based on the record, seems sure to come), fiber optics and digital transmission will dominate.

Many organizations in different parts of the world contributed to the progress of transmission technology and the realization of its benefits to society, but an extraordinary portion of that advance came from the research and development work in Bell Telephone Laboratories, as the story told in this volume will make clear.* The success at Bell Laboratories was due to a close-knit team of research, systems engineering, and development people. The research organization was free to follow any line of inquiry that seemed likely to advance basic understanding and to apply their knowledge in convincing demonstrations when these were judged appropriate. The systems engineering and development groups planned and designed systems based on day-to-day contacts with the research people. This working relationship of research, engineering, and development people was closer in transmission, it is believed, than in any other branch of communications technology. The equally close working relationship of the systems designers with Western Electric as the manufacturer and, through AT&T, with the operating telephone companies ensured that the systems were designed with the end user's requirements always in mind. Finally, the confidence of the managers of AT&T and Western Electric, the owners of Bell Laboratories, that a sustained program of research and development would ultimately pay high dividends was vital to the success of the work.

Irwin Welber
Vice President, Transmission Systems
August 1985

* As of January 1, 1984 Bell Laboratories adopted the designation AT&T Bell Laboratories in order to more clearly show its relationship with AT&T.

Acknowledgments

Transmission systems development was, and is, very much a team effort. In recognition of this, there are few personal citations in this volume. Outstanding individual contributors with original ideas and motivators or organizers of exceptional ability and vision are occasionally noted, but the team workers whose skilled efforts translated concepts into working systems generally remain anonymous. Their vital collective contribution is hereby acknowledged.

The story told in this volume also ignores organizational boundaries. Projects often advanced from discoveries in pure and applied research, through system studies and analysis, to exploratory development, to final design and the preparation of manufacturing information, with each stage of the work carried out in a different organization. The approach taken in this book results in a certain amount of redundancy, with parts of the story also covered in companion volumes, but the disadvantages are more than outweighed by having a complete story in one place.

Terminology is a problem in the history of technology; it was loose at all times and changed, often drastically, over the years. In an effort to convey a sense of things as they were, the writers and editor have generally used terms that were in use at the time, but have not hesitated to use later terminology where the earlier usage might be confusing. (Presumably, a generation has now grown up to whom *hertz* comes naturally and *cycles per second* sounds archaic.) Though contemporary illustrations are used where appropriate, some figures represent a particular period of development; the labels on these figures are especially likely to include terms no longer in use. In all cases it is hoped that the meaning will be sufficiently clear from the context.

This book was written with the assistance of many members of the technical staff of AT&T Bell Laboratories—too many, indeed, to acknowledge all their contributions individually. I am especially indebted to the following, who wrote initial manuscripts of complete chapters or major chapter sections, or who otherwise rendered help without which the book could not have been completed:

C. C. Cutler wrote the material on high frequency radio on which Chapter 2 is based, using, in part, an earlier unpublished document by F. A. Polkinghorn. W. H. Doherty wrote a comprehensive account of the Bell Laboratories involvement in broadcast radio, both AM and FM, from which Chapter 3 is adapted. The early history of microwave radio relay (Chapter 7) borrows freely from material prepared by A. C. Dickieson. The sections of Chapters 8 and 17 (Supporting Technology) on measurements is based on material furnished by G. D. Haynie and on network design by F. J. Witt. The information on later microwave radio (Chapter 11), was contributed by W. C. Jakes and R. L. Markle and is also based, in part, on published material by A. C. Dickieson and U. S. Berger. S. T. Brewer was the principal contributor for Chapter 12 on submarine cable, H. W. Evans for Chapter 13 on satellites, and P. T. Porter, S. Rhee, and their associates for Chapter 14 on mobile radio. W. H. Von Aulock wrote a comprehensive account of subscriber loop technology (Chapter 16). Material on cable and wire is based, in part, on an unpublished account by L. L. Swift and D. E. Wachter.

M. R. Aaron wrote the history of the early research and development on digital transmission for Chapter 18, using, in part, an unpublished account by W. R. Bennett. Chapter 19, which covers the further development of digital systems on wire pairs, was the work of J. J. Lang and associates. In Chapter 20 on higher rate digital systems, V. I. Johannes contributed material on high rate codecs and the T4M system; D. A. Alsberg wrote the millimeter waveguide story, which also benefitted from discussions with W. D. Warters; and W. C. Jakes and A. C. Longton furnished the material on digital radio.

R. L. Easton assisted with all sections of Part IV and wrote an account of the development of transmission standards and objectives (Chapter 21), basing it on an unpublished memo by R. W. Hatch and J. L. Sullivan. K. V. Annambhotla wrote the material in Chapter 22 on services, R. P. Guenther prepared Chapter 23 on operations and maintenance, and T. Walzman contributed Chapter 24 on economics.

Numerous other people assisted those named above, and many made contributions by phone calls and in person. J. Gilmore and F. J. Herr were helpful in identifying sources and promoting the flow of information. Chapters not otherwise attributed were written by the editor. Much of the material written by others was freely adapted for the purposes and size constraints of the book. The help of all contributors is, therefore, gratefully acknowledged, but they cannot be held responsible for any errors in the published text. The encouragement of I. Welber and the constructive comments of G. E. Schindler, Jr., and S. Millman were valuable throughout.

The manuscript drafts were produced with the assistance of many people. R. Ludwig did most of the original typing, and J. McLaughlin entered copy into the UNIX* text processing system, while B. Tursky input massive changes

---

* Trademark of AT&T Bell Laboratories.

made by the reviewers. The text benefitted from the editing of M. Heimerdinger, D. M. Solomon, and C. Biczak. C. Biczak also checked all the references. P. Greenberg and M. Cohen proofread the text. D. M. Solomon and C. Biczak prepared the illustrations while C. Yerger, J. P. Wyszkowski, M. Sanok, and A. Schillinger did the artwork. D. Mann and others at Tapsco, Incorporated, were extremely cooperative and did an excellent job of typesetting the text and preparing mechanicals. W. Feik prepared the index, and E. T. Foreman of Design in Mind designed and created the dust jacket. The organization of this mass of detail was initially the responsibility of D. McGrew and, through most of the preparation, S. Annunziata. The capable and thoroughly professional contribution of all these people is deeply appreciated by the authors and the editor.

E. F. O'Neill
Editor

PART I
THE DEVELOPMENT OF ANALOG TECHNOLOGY (1925–1950)

The improvement and general availability of vacuum tubes in the 1920s had a major impact on telecommunications. With reliable vacuum tube repeaters it became possible to interconnect widely separated cities in the United States network with circuits of low loss and good quality. A policy of making toll trunks available on demand was a spur to growth. Vacuum tube repeaters were first used on open-wire lines, but in the late 1920s, repeaters on voice-frequency four-wire circuits made up of small gauge wires in weatherproof cables were used to expand intercity trunk capacity. Vacuum tubes and frequency-selective wave filters also made possible the transmission of multiple circuits on a single pair by carrier techniques. Carrier was first used in the last years of World War I and went through successive developments in the early 1920s. The highly successful three-channel Type C carrier system for open-wire lines was first used in 1924 and was widely installed from the late 1920s through the 1940s.

Vacuum tubes also made radio transmission of voice signals possible. Radio appeared to be especially attractive for overseas telephony where wire transmission was not practical. The first transatlantic radio telephone circuit was established in 1927 with a carrier frequency at about 60 kHz. This was supplemented in 1929 by high frequency (HF) circuits operating in the range from about 7 to 20 MHz. In the years that followed, HF radio links were established worldwide, and these remained the basis for intercontinental telephony until replaced by submarine cable and satellite circuits in the 1950s and 1960s. Bell Laboratories also played a leading role in the development of broadcast radio equipment. Development groups produced high quality audio and studio control equipment and the most linear and efficient transmitters of the early AM era. Later, they designed FM transmitters to an equally high standard and pioneered in the design of novel FM antennas.

In 1927, H. S. Black discovered the gain stabilizing and distortion reducing properties of negative feedback. His work was extended in the 1930s by H. Nyquist, who established the criterion for stability, and H. W. Bode, who developed design methods for stable amplifiers. Feedback was basic to the realization of large-capacity multiplexed analog systems. It was used in the late 1930s in amplifiers for 12-channel carrier systems on cable pairs (K carrier) and on open wire (J carrier). Coaxial cable was proposed as a transmission line for broadband telephony in 1929 and developed into a practical transmission medium in the mid-1930s. An experimental coaxial system with a bandwidth of 1 MHz was installed between New York and Philadelphia in 1936 and used to transmit television pictures as well as multichannel telephony. Further development extended the band to 3 MHz to keep pace with the rapidly advancing television technology. The Type L1 coaxial system, with an initial capacity of 480 telephone circuits or one television picture signal, went into service in 1941.

Radio research was extended to higher and higher frequencies in the 1930s as negative-grid vacuum tubes of advanced design and velocity-modulation tubes made work above 100 MHz possible. The research included the design of waveguide components and led to a recognition of the existence of the low-loss circular-electric mode in cylindrical waveguide. The work also led in 1941 to a proposal for a microwave radio-relay system and was basic to the advances in microwave radar during World War II. Development toward commercial microwave radio-relay was started immediately after the war and culminated in 1947 in a successful demonstration of television and multiplex telephony transmission over the TDX system, an experimental seven-hop link from New York to Boston.

Chapter 1

The State of the Technology (1925–1930)

I. INTRODUCTION

In 1925, it was possible to call almost everywhere in the continental United States over a circuit of good quality. In most areas, local service was provided for a flat monthly charge independent of usage. The charge for long-distance calls was a toll, assessed according to the distance and duration of the call. Because of this, in telephone usage, the term *toll* became essentially synonymous with long distance. Both local and long-distance transmission were entirely over wire pairs. Toll transmission lines were predominantly open-wire pole lines, but telephone cables were used in congested urban areas and for some short toll circuits. In addition to voice, the domestic network provided extensive capability for telegraph transmission on the same wires used for voice telephony. Other signals were carried as well. Telephotograph transmission was demonstrated in 1923 and 1924 and opened to commercial use in 1925. Transmission of a primitive television picture was demonstrated over an open-wire line between New York and Washington and over shorter radio links in 1927.

Radio was a new and extremely exciting field in the 1920s, and, in addition to point-to-point radiotelephone links, the Bell System became actively engaged in radio broadcasting. Station WEAF (which later became Station WNBC) was operated from the Bell Laboratories location on West Street in Manhattan, Western Electric was a leading manufacturer of transmitter components, and AT&T provided extensive program circuits to interconnect transmitters for networking. It was not yet possible to call overseas by radio in 1925, but experimental transmissions had been made to England as early as 1923 with equipment aimed at eventual commercial service. Transatlantic service was first established in 1927 by long-wave radio at frequencies around 60 kHz and extended in 1928 and 1929 by high-frequency radio at frequencies in the 5- to 25-MHz range. The latter was an important advance because of its much greater capacity for expansion. It appeared that the whole world was about to be brought within reach of the telephone.

To many people, the achievements of the time must have appeared to be a climax in the technology of communications, launched by the invention of the vacuum tube. (See the first volume in this series, *The Early Years (1875–1925)*, for a complete discussion of early developments.) But, in many respects, 1925 marked a beginning rather than the culmination of an era. The application of electronics to telephony had hardly begun. Dialing local calls in large cities was becoming common, but long-distance calls were still placed by passing the information to a toll operator, who called back when the connection was established. The average time required to accomplish a connection had been steadily reduced from 7 minutes in 1920 to about 2-1/2 minutes in 1925. At that time, the Bell System management adopted a policy to reduce the time even further by enlarging the intercity circuit groups, providing a tremendous spur to network expansion.

It was, indeed, in many ways the culmination of an era—an era in which the principal objective had been the conquest of distance. But it was also the threshold of an era—an era to be marked by the multiplication of facilities and reduction of costs. By 1975, a goal for the Bell System, set in 1907 by T. N. Vail, then president of AT&T, had been met. The United States had essentially universal service at affordable rates.

II. OPEN-WIRE TRANSMISSION

In the early 1920s, intercity transmission was almost entirely at voice frequency on open-wire pairs. These pairs consisted of copper wires up to 165 mils (1/6 inch) in diameter, suspended about 12 inches apart on insulators on the crossarms of a pole line. Up to five pairs could be carried on each crossarm, and, depending on the traffic, the poles could be equipped with four or more crossarms. The conductors of each pair were periodically transposed to balance the capacitive and inductive coupling to other pairs on the same route. In the absence of such balance, the signals on one pair would induce excessive interference and possibly intelligible crosstalk in the other pairs.

From the early 1900s, the loss of open-wire pairs was often reduced for voice frequencies by inductive loading in series with the line. With the insertion of loading coils of an inductance value and at spacings determined by the capacitance of the pair, the transmission loss was transformed to a characteristic resembling the passband of a low-pass filter, with lower and more uniform loss in the useful voice-frequency range than the loss of an untreated line. In addition to loading to reduce loss, it was common to derive a third circuit from two 2-wire circuits. With both wires of each physical pair (the side circuit) used in effect as a single conductor in a balanced bridge arrangement, a third, or *phantom*, circuit could be obtained on every four wires.

The limit of long-distance telephony in the pre-electronic era was reached when a heavy-gauge, loaded, open-wire line was extended from the East Coast to Denver in 1911. Talking was possible, but just barely possible, over the line; to go further clearly required some form of signal amplifier. The invention and subsequent improvements in the high-vacuum tube provided just such an effective amplifier for the first practical telephone repeater and removed the last barrier to long-distance transmission within the continental United States. This event was dramatized by the construction and opening to service of the first transcontinental line to San Francisco in 1915.[1] (The terms *repeater* and *amplifier* are used somewhat interchangeably in telephone parlance. Generally, *amplifier* refers to a one-way device supplying signal gain, while *repeater* is more accurately used for the arrangement of transformers (hybrids) and one or more amplifiers, capable of amplifying speech in both directions. In time, and especially in carrier systems where the directions of transmission are separated, the term *repeater* was often used, even though only one direction of transmission was involved.)

The first coast-to-coast line provided three open-wire circuits consisting of two physical pairs and a phantom. Less than one-half of one percent of the distance was in cable. Three vacuum-tube repeaters were used at first, with previously developed mechanical amplifiers available on a standby basis, reflecting the somewhat shaky confidence in the new high-vacuum triode. The vacuum tube quickly demonstrated its superiority; within the first year, three additional repeaters were added, and, in 1918, two more were added for a total of eight. Knowledge of the properties of the new device was growing, as was confidence that it could be designed into an overall circuit and used as an integral part of the transmission design, rather than as a prop to rescue a faint signal from total oblivion.

An important additional step was taken in 1920 when the transcontinental line was unloaded and equipped with 12 improved repeaters using 3000-Hz filters. Compared to the original heavily loaded line, the transmitted band of frequencies was doubled from 900 to 1800 Hz, the loss was halved, and the propagation velocity was increased by a factor of 3.5. All of these factors contributed to a marked improvement in quality. Additional open-wire transcontinental lines along widely separated routes were built in 1923 (Denver, El Paso, Los Angeles) and in 1927 (Chicago, Minneapolis, Seattle). These, along with many other long circuits, were not loaded, but were designed and equipped with repeaters to the new standards from the start. In 1925, AT&T recommended that loading not be applied to open-wire lines where it was not already in place. This was intended to improve the flexibility in plant rearrangements and to make the pairs more readily available for carrier.

III. CARRIER ON OPEN WIRE

It had long been recognized that the widely spaced open-wire pairs were capable of transmitting a much broader band of frequencies than the 3 kHz or so required for good-quality voice transmission. Vacuum tubes not only made voice-frequency repeaters practical, but, along with the electric-wave filter, opened the door to the wider band at higher frequencies by carrier techniques. System developers had, in effect, cut their teeth on the earlier Type A (1918) and Type B (1920) carrier systems. They had learned a great deal about the pros and cons of single- and double-sideband transmission, carriers transmitted and suppressed, crosstalk, the problems of using the same or different frequencies for opposite directions of transmission, and a host of other problems related to the use of higher frequencies on existing open-wire pairs. The lessons learned from all this early experience were embodied in the Type C carrier system, first installed between Pittsburgh and St. Louis in 1924.

The Type C carrier provided three 2-way voice channels above the voice frequencies in the band from 5 to 30 kHz on a single open-wire pair. Transmission for each channel was by single sideband with suppressed carrier, and different frequency bands were used for opposite directions of transmission. A pilot tone, separate from the voice channels, was transmitted so that the transmission gain or loss for a constant-amplitude signal could be observed and adjusted at repeaters and terminals.[2] C carrier embodied no radically new concept, but it was carefully designed and engineered, and dealt with the real-world problems of carrier on open wire in a thoroughly practical way. It was immediately successful, and other installations followed rapidly. C carrier became a workhorse system on open wire and was installed in increasing quantities and with many improvements, until 1.5 million voice-circuit miles were in service in 1950. The last system was removed from service about 1980, a tribute to its durability. A more detailed account of the early development of open-wire lines, loading, phantoms, and the first carrier systems is given in *The Early Years (1875–1925).*

IV. TRANSMISSION ON LONG CABLES

Open-wire lines date back to the first years of telephony; but, at an early date, the large space required, the need to meet special situations (such as underwater river crossings), and the unsightliness of heavily equipped open-wire pole lines led to the development of telephone cables. In cables, a large number of wire pairs of much smaller diameter (finer gauge) than those used for open wire were assembled in a compact bundle within a weatherproof sheath. Although the higher resistance and high capacitance between the wires of these small-diameter, closely spaced pairs caused much higher losses than the same length of open-wire line, this was tolerable for short distances, and cables were widely used between central offices in cities and as entrance links to long open-wire lines. The wires of each cable pair were twisted together to reduce crosstalk coupling to other pairs, and the entire sheathed bundle could be suspended from poles or drawn into an underground conduit.

From an early date, parts of long toll circuits were placed in cable to provide entrance links to congested urban areas and for river crossings. Widespread use for longer links was prevented by the high loss of the closely packed pairs. For the same gauge, at 1000 Hz, a cable pair had about 10 times the loss compared to open wire; for the finer gauges commonly used, the loss could be 25 to 30 times as great. In addition, before effective loading was developed, the high capacitance caused a pronounced change in loss with frequency, even across the few kilohertz of a single voice signal. Because of the high loss and its variation over the voice frequencies, loading was especially beneficial in cable, where the pair inductance was much less than the optimum for low-loss transmission on the high-capacitance pairs. But loading was not a panacea. Adding substantial inductance (heavy loading) while decreasing the loss narrowed the band and distorted transmission at the band edge. Despite these handicaps, cable was used on some intercity routes in the Northeast to provide protection from the weather, especially the sleet and freezing rain that proved so destructive to open-wire lines in that area.

The first long intercity toll cable was installed between Philadelphia and Washington in 1912 and was extended through New York to Boston in 1913. It was made up of loosely packed four-wire quads (twisted pairs twisted together to permit the derivation of a phantom on the four wires with crosstalk to other quads balanced out by the twists) of No. 13 gauge (1.8-mm diameter) and even No. 10 gauge (2.6-mm diameter) pairs and was heavily loaded. Despite the massive wires and the loading, the loss was very high, and a mechanical repeater was used at Philadelphia from the start. When vacuum tubes became generally available, they were promptly applied to the Boston-Washington and other early cables. Since the loss in cable was mainly determined by the high capacitance, loading continued to be used in conjunction with repeaters to provide a reasonably flat loss over the speech band and to make transmission possible over smaller-gauge pairs. In the early 1920s, however, even with additional repeaters, transmission over long cable circuits was not entirely satisfactory. It was one thing to use a few, or even several, repeaters on the low-loss open-wire lines; it was quite another to apply them at frequent intervals to the high-loss pairs within the close confines of a cable.

4.1 Four-Wire Circuits

The first (and most subsequent) amplifiers were unidirectional devices. They amplified low-level signals at their input to a higher amplitude at the output. When single wire pairs were used for both directions of transmission, it was necessary to separate the signals in each direction, amplify them, and recombine the oppositely directed signals on the single pair. The separation and recombination were done by means of transformers and hybrid coils, in a manner similar to the separation of transmitted and received signals in a telephone set. On open-wire lines, repeaters were needed only at about 250-mile intervals, and the separation and recombination of the directional signals were done at each repeater. In cable, however, the higher loss required amplifiers at much closer spacing, typically about every 50 miles. The frequent separation and recombination of signal directions and the high gain required introduced new and difficult problems in transmission. The problems of maintaining balance and crosstalk became overwhelming. As a result, for circuits of more than 100 miles or so in length, it became common to separate the transmission directions at the terminals of a transmission section and to transmit and amplify each direction over a separate pair in a four-wire circuit.

The problems of repeatered and loaded cables were difficult, and four-wire operation, while very helpful, seemed like an extravagant use of copper, even though the gauges could be smaller than for two-wire transmission. On the other hand, the need for toll cables was urgent. Not only was the use of long-distance telephony increasing rapidly throughout the 1920s, but the policy of upgrading the service added to the pressure. If transmission could be improved sufficiently, a single 278-pair cable was capable of providing ten times as many circuits as a fully equipped pole line. In heavily populated areas, the system had literally outgrown the open-wire line. In addition, the weather immunity of cable had become essentially a requirement for new facilities.

4.2 The New York–Chicago Cable

In 1921, the combination of effective vacuum-tube repeaters and improved loading, using compressed powdered-iron core coils, made the use of No. 19

gauge (0.9-mm diameter) conductors possible (at least on paper) in a cable from Philadelphia to Pittsburgh. Repeaters were spaced at about 50 miles, but, even with the cable designed on a four-wire basis, a host of new problems were encountered. The low level to which the signals were attenuated before amplification caused severe problems from crosstalk and noisy splices. Nonuniform spacing and changes in loading-coil inductance caused by lightning surges distorted transmission. Loading-coil spacings were more carefully observed, and the coils were redesigned with closer tolerances and air gaps to improve surge resistance. Interference from superimposed telegraph circuits was a constant problem, alleviated finally by abandoning ground-return telegraph on these facilities. All difficulties were finally overcome, and the cable was successfully extended to New York and Chicago by October 1925. This marked the beginning of an epoch, as significant as the achievement of the first transcontinental line ten years earlier.

4.3 Advances in Loading

The importance of finer gauge in cable can be appreciated by noting that a No. 6 gauge (165-mil diameter) open-wire transcontinental pair required over 1200 tons of copper. A No. 19 gauge pair required only 1/13 as much copper per pair mile and thus required much less copper, even in four-wire circuits. Loading was crucial to cable, and the cost of loading itself was not trivial. At a time when several loading coils can be held in one hand and weigh only a few ounces, it is difficult to realize what was involved in this earlier period. The first open-wire iron-core loading coils were over eight inches in diameter and weighed almost 30 pounds. Loading coils for cable were smaller, but, in 1922, the coil pot for the 108 coils needed to load 54 four-wire circuits was 30 inches in diameter by 52 inches high and weighed 2700 pounds! Furthermore, for cable, such loading had to be provided every 6000 feet instead of the 8-mile spacing on open wire. It required a considerable amount of massive hardware to provide and load all the pairs and phantoms in a large cable.[3] The total weight of the New York–Chicago cable was over 17,000 tons (almost 11,000 tons in the lead sheath). If all the pairs had been equipped and loaded, several thousand tons of loading apparatus would have been required in addition. Intensive economic studies were directed at the optimum trade-offs between repeaters, wire gauge, and loading in those years.

A significant improvement was achieved in 1928 when powdered-permalloy coils became available.[4] A nickel-iron alloy discovered about 1915 by G. W. Elmen of Western Electric, permalloy provided extraordinarily high permeability at the very low flux densities characteristic of the signal levels on telephone cables.[5] It had been used in tape form for the continuous loading of submarine telegraph cables in 1924 and 1925 with spectacular success, the signaling speeds of the loaded cables being increased by a factor of four or five. Compressed powdered-permalloy loading coils reduced the cost of telephone cable loading by as much as 40 percent and the size by an even larger factor. Further research culminated in the mid-1930s in the design and manufacture of molybdenum-permalloy coils.[6] These provided a second 40-percent reduction in size and cost.

4.4 Echo and Echo Suppressors

With four-wire circuits, improved and standardized repeaters, and effective loading, the cable network expanded rapidly. By 1927, cable facilities were available from the East Coast to Chicago and St. Louis, and were extended to Dallas by 1933. As the circuits grew longer, a significant new problem appeared. Propagation on the loaded pairs was relatively slow, at about 20,000 miles per second. This was only about one-eighth the velocity on unloaded open wire, which approached the velocity of light. The echo of a talker's own voice from the inevitably imperfect termination at the distant end becomes disturbing when the delay exceeds about 100 ms, about the round-trip interval over 1000 miles on the slow lines. (For shorter delays, the echo merges with, and is indistinguishable from, the normal sidetone.) Echo suppressors, voice-actuated devices that opened the return path when the speaker on the near end was talking, were installed to eliminate the echo on longer circuits. Echo suppressors, in turn, introduced impairments that were not entirely eliminated for many years, but they made the long circuits satisfactory by the standards of the day. The expansion of the voice-frequency cable network was finally inhibited by the collapse of business during the Great Depression and by the prospect for carrier on cable.

V. NONVOICE SIGNALS

5.1 Telegraphy

In the early years of telephony, telegraphy was already a well-established and widely used service. It was recognized that the lines required for telephone transmission could be used for telegraph transmission as well. Nobody thought it strange that the new long-distance company should include telegraph in the name it adopted in 1885. From the start, however, the Bell System companies restricted their telegraph service to leased private lines. The operating companies transported the coded signals, but did not furnish the specially trained operators needed to transmit and receive the signals in the days of manual operation.

Telegraph circuits over telephone lines were derived by either simplex or composite arrangements. In simplex circuits, both wires of a telephone pair were used to derive a single, ground-return telegraph circuit, as a pair was used for one-half of a phantom circuit. The balance between the wires at the pair terminal helped to separate telegraph and telephone signals. (A discussion of circuit balance appears in *The Early Years (1875–1925)*, Chapter 4, Section 2.2.) In the composite arrangement, a ground-return telegraph circuit was derived from each wire of the pair. Elementary low-pass/high-pass arrangements (series capacitors and shunt inductors) were used to separate the low-frequency telegraph signals from the higher-frequency voice signals. The control of Western Union by AT&T, temporarily achieved in 1910, spurred Bell System interest in public message service and led to the development of printers and, ultimately, to teletypewriter service, which became well established in the 1920s.

The common use of conductors for both telephony and telegraphy was not without technical problems. Telegraph thump and flutter were caused by cross-talk from the ground-return telegraph circuits and by interaction with the high-level telegraph signals in the somewhat nonlinear loading coils. When fine-gauge, loaded, repeatered cables were introduced in the 1920s, the low levels to which telephone signals were attenuated at the end of a repeater section made these problems acute. A major improvement was made early in the development of fine-gauge cables by the use of fully metallic telegraph circuits. The simplex arrangement was extended to occupy the four wires of two pairs, in a manner essentially identical to that used for a voice-frequency phantom, to derive one nongrounded telegraph circuit. Telegraph interference into the telephone channel was further diminished by a reduction in the telegraph line current from 60 to 5 ma, made possible by the metallic circuit. A large reduction in interference from power lines was realized at the same time by eliminating the ground returns. But metallic telegraph circuits were costly. The need for four wires compared to only one for the composite arrangement, or even two for a grounded simplex line, was a strong incentive for the development of multiplexed carrier telegraphy.

5.2 Carrier Telegraph

While cables were becoming common on the dense toll routes of the Northeast, they were not yet much used off these routes or in the more open country of the West. But voice and telegraph traffic was growing on the open-wire routes, too. At very nearly the same time as the first use of carrier for telephony, an open-wire carrier telegraph system was developed.[7] An early decision was made not to provide carrier telegraphy and either voice frequency or carrier telephony on the same open-wire pair. Using vacuum tubes and electric-wave filters in a way similar to that used for the contemporary carrier telephone system, 20 one-way channels (10 west-east and 10 east-west) were derived to make up 10 two-way channels on a pair in the frequency band from 3300 to 10,000 Hz. Carrier frequency spacing varied from 240 to 500 Hz, being roughly proportional to frequency, to ease the filter and frequency control problems. A trial system was tested between Pittsburgh and Harrisburg during 1919, and the trial was extended to Chicago in 1920. A standard system, the Type B carrier, based on the trial design, with repeaters at 250- to 300-mile intervals, was widely installed in the following years.

In the early 1920s, a carrier-telegraph system was also developed for cable using the same techniques as those used on open wire, but, by taking advantage of advances in the technology, designers created a system providing 10 to 15 (one-way) telegraph channels in the nominal 3-kHz voice-frequency band of a single telephone channel. (Ten channels were in use in 1925. The total that could be implemented depended on the weight of loading and hence the band available.) Channels were spaced at uniform 160-Hz intervals, the carriers being generated by a rotating machine with 12 carrier-generating rotors on a single shaft. This was done to assure the proper relative frequencies of the carriers, but the arrangement was replaced by stable vacuum-tube oscillators when these became available a few years later.[8]

The interesting thing about the voice-frequency telegraph carrier system was that it used the identical line for telegraph and telephone. This was one of the first instances in which an essentially unmodified telephone line was used, by means of appropriate terminal arrangements, to carry signals altogether different from the ones for which it was originally designed. It was essential, of course, to confine the frequencies of carrier telegraph in cable to the voice frequencies under 3 kHz. The wider band used on open-wire carrier telegraph could not be transmitted over the loaded cable pairs.

5.3 Radio and Program Circuits

In one of its enduring roles in broadcast radio, AT&T furnished, and continues to furnish, interconnections among the stations of the national networks. These program circuits were unique and had many special requirements compared to ordinary telephone connections. Satisfactory reproduction of music required both a wider band and the ability to transmit a greater range of volume than was needed for ordinary telephone circuits. On the other hand, these circuits were normally one-way circuits, and problems of balance and echo were not a factor.

Specially administered circuits with broadband filters and special amplifiers with the required load characteristics were in use on both open wire and cable by the end of the decade. Great care was taken that high-level music would not interfere through crosstalk into the adjacent voice circuits and that crosstalk from the voice circuits would not be audible during the low volume or silent intervals in the program channel. These objectives were met by special amplifiers with higher-power output and, in cable where the crosstalk problem was most severe, by dedicated No. 16 gauge pairs specially located in the cable bundle. The early systems provided about a 5-kHz band, and systems for 8 and even 15 kHz were also designed. However, the limitations of other radio components, especially home receivers, provided little incentive to incur the higher costs involved in transmitting the broader band, and 5 kHz remained a common standard for AM radio networking up through the mid-1980s, when this volume was prepared.

VI. SUMMARY

The impact of the technical developments on the transmission network of the late 1920s was profound [Table 1-1]. Total network capacity increased by over 250 percent in only five years. Three-quarters of this astonishing growth was by cable circuits, a large portion of which was in fine-gauge, loaded, repeatered cables. Cable capacity alone increased by a factor of four, and, by

TABLE 1-1: Total Voice-Circuit Miles (Thousands)

	1925	1930
Voice frequency on open-wire pairs	1,613	2,121
Voice frequency on cable pairs	1,054	4,310
Carrier on open wire	47	486
Total	2,714	6,917
Vacuum-tube repeaters	7,500	80,000

1930, it accounted for 62 percent of system capacity, compared to 39 percent at the start. Carrier on open wire, starting from a small base, increased by a factor of ten, almost equaling in growth the increase at voice frequency on the extensive open-wire lines. That the age of electronics had truly begun is perhaps best indicated by the tenfold increase in the number of vacuum-tube repeaters in service over those few years.

Underlying the large-scale changes in the external, measurable plant were a number of fundamental advances in science and technology that made them possible, and that set the stage for the even more spectacular advances that lay in the future. In the same few years, J. B. Johnson and H. Nyquist identified and quantified a basic physical limit to amplification in the thermal noise of electrons in conductors.[9,10] O. J. Zobel advanced and systematized the design of filters and equalizers;[11] quartz crystals were first applied to precise frequency controls[12,13] and were soon to be used in filters with a new order of discrimination;[14] and H. S. Black invented the negative-feedback amplifier, the basic discovery on which almost all later carrier development rested.[15]

An array of other technology and support activities kept pace with the system developments and deployments. Measuring equipment for both components and system transmission was improved in accuracy, speed, and convenience, and was made smaller and portable for use in the field as well as the laboratory. Vacuum tubes became stable, durable components. Standards were developed and extended. The "mile of standard cable" used as a standard of loss was replaced by the transmission unit, later renamed the decibel. Standards were established for noise and distortion. In 1924, the bandwidth objective for exchange trunks was increased from 2300 to 2800 Hz. In a series of tests in the late 1920s and early 1930s, H. Fletcher studied the relationships among bandwidth, noise, and the ability of listeners to distinguish words. As a result of recent developments in telephone instruments and the growing prevalence of carrier systems for toll transmission, the recommended speech band was increased to 3500 Hz in 1932. In 1925, essentially no toll connections were completed while the customer was on the line. By 1929, the average setup time was a little more than one minute, and 82 percent of connections were established with the calling customer on line. The system was not only getting larger, it was becoming faster and providing better speech quality as well.[16]

REFERENCES

1. T. Shaw, "The Conquest of Distance by Wire Telephony," *Bell Syst. Tech. J.* **23** (October 1944), pp. 337-421.
2. H. A. Affel, C. S. Demarest, and C. W. Green, "Carrier Systems on Long Distance Telephone Lines," *Bell Syst. Tech. J.* **7** (July 1928), pp. 564-629.
3. T. Shaw, "The Evolution of Inductive Loading for Bell System Telephone Facilities," *Bell Syst. Tech. J.* **30** (January 1951), pp. 149-204; (April 1951), pp. 447-472; (July 1951), pp. 721-764; and (October 1951), pp. 1221-1243.
4. W. J. Shackelton and I. G. Barber, "Compressed Powdered Permalloy, Manufacture and Magnetic Properties," *Trans. Amer. Inst. Elec. Eng.* **47** (April 1928), pp. 429-436.
5. H. D. Arnold and G. W. Elmen, "Permalloy, An Alloy of Remarkable Magnetic Properties," *J. Franklin Inst.* **195** (May 1923), pp. 621-632.
6. V. E. Legg and F. J. Given, "Compressed Powdered Molybdenum-Permalloy for High-Quality Inductance Coils," *Bell Syst. Tech. J.* **19** (July 1940), pp. 385-406.

7. E. H. Colpitts and O. B. Blackwell, "Carrier Current Telephony and Telegraphy," *Trans. Amer. Inst. Elec. Eng.* **40** (February 17, 1921), pp. 205-300.
8. J. A. Duncan, R. D. Parker, and R. E. Pierce, "Telegraphy in the Bell System," *Trans. Amer. Inst. Elec. Eng.* **63** (June 1944), pp. 1032-1044.
9. J. B. Johnson, "Thermal Agitation of Electricity in Conductors," *Phys. Rev.* **32** (July 1928), pp. 97-109.
10. H. Nyquist, "Thermal Agitation of Electric Charge in Conductors," *Phys. Rev.* **32** (July 1928), pp. 110-113.
11. O. J. Zobel, "Extensions to the Theory and Design of Electric Wave-Filters," *Bell Syst. Tech. J.* **10** (April 1931), pp. 284-341.
12. W. A. Marrison, "A High Precision Standard of Frequency," *Bell Syst. Tech. J.* **8** (July 1929), pp. 493-514.
13. F. R. Lack, "Observations on Modes of Vibration and Temperature Coefficients of Quartz Crystal Plates," *Bell Syst. Tech. J.* **8** (July 1929), pp. 515-535.
14. O. E. Buckley, "Evolution of the Crystal Wave Filter," *J. Appl. Phys.* **8** (January 1937), pp. 40-47.
15. H. S. Black, "Stabilized Feedback Amplifiers," *Bell Syst. Tech. J.* **13** (January 1934), pp. 1-18.
16. W. H. Martin, "Transmitted Frequency Range for Telephone Message Circuits," *Bell Syst. Tech. J.* **9** (July 1930), pp. 483-486.

Chapter 2

Overseas Radio

I. THE FIRST HIGH-FREQUENCY CIRCUITS

The success of early radiotelegraphy at low frequencies and the evidence that transmission was limited to 50 miles or so at frequencies above a few hundred kilohertz was such that little work was done above 1 MHz for many years. However, the intensive use and long distance of transmission at the low frequencies, coupled with a lack of antenna directivity, imposed severe limits on the expansion of services within this restricted band. In 1916, Marconi, working in England and keenly aware of these problems, directed his associate, C. S. Franklin, to begin experiments at frequencies above 2 MHz. This work was motivated in part by the desire for secrecy in military use. There was also some similar military work by others at much higher frequencies at about the same time, but Marconi was especially interested in antenna directivity and had broader applications in mind.

In 1921, several amateurs detected transatlantic signals at 1.5 MHz during the night hours. By 1923, working at 3 MHz during the night, amateurs in France and the United States were able to exchange messages. In the spring of that year, Marconi received some signals at 3 MHz up to 1400 miles from his transmitter during the day and strong signals over 2500 miles at night. The following year, 9-MHz signals sent from his transmitter at Poldhu, England were received throughout the day on his yacht harbored in Beirut, Lebanon, 2400 miles away.[1] Clearly, a new and important propagation mechanism was at work, and the properties of shortwaves deserved further investigation. Late in 1921, R. A. Heising and J. F. Farrington at Western Electric began to study methods of generating, receiving, and measuring high-frequency signals. In 1922, a program of experimental work at 3.5 to 5 MHz was started in the Western Electric group that was to become the Bell Laboratories Radio Research Department. By 1924, early work had led to the conclusion that a larger-scale program looking toward commercial applications was justified.

In the 1920s, radio research and development in the Bell System was divided among three groups. Considerable theoretical work was done by the AT&T Development and Research Department in New York, where laboratory analyses were supplemented with field testing at AT&T working radio stations as they came into operation. The more experimental research was done by Bell Laboratories, initially in New York, but later, in locations that were less troubled by radio noise—at Cliffwood, Deal, and Holmdel, New Jersey. The people

Fig. 2-1. Engineers in charge of the development of the shortwave radio transmitters and receivers. *Left to right:* A. A. Oswald, M. J. Kelly, W. Wilson, R. A. Heising, J. C. Schelleng, H. T. Friis.

heading the New Jersey research laboratories worked mainly on transoceanic radiotelephony [Fig. 2-1]. The third group concentrated on broadcast radio and the design of equipment for sale and commercial use. They, too, started in New York at the Bell Laboratories location at West Street, where Station WEAF was in operation. When work on an experimental station, called Station 2XB, involved increasing the power far above the level that could be tolerated in a location where sensitive measurements were also necessary, 2XB and a portion of this third group were moved to a new radio field laboratory in a dairy barn of stately proportions at Whippany, New Jersey. This became the center of commercial radio and broadcasting work in Bell Laboratories.

In 1925, many of the people in the research group who eventually contributed to shortwave radio were caught up in the establishment of long-wave telephony using the Radio Corporation of America (RCA) telegraph antenna at Rocky Point, Long Island. The demonstration of transoceanic speech transmission in 1915 and the known reliability of the propagation of long-wave radio* naturally led to a concentration of effort on these frequencies for the first transatlantic radiotelephone link. The long-wave work had stretched the group's ingenuity in getting sufficient bandwidth for a single speech channel

* The early terminology of frequency and wavelength had not been standardized, and terms used sometimes conflict with later usage. The term *long waves* generally referred to frequencies up to about 300 kHz. The first transatlantic telephone system used carriers at about 60 kHz. Before 1924, the term *high frequency* usually meant frequencies in the 500- to 3000-kHz range. In the early 1920s, the term *shortwaves* was often used for the 2- to 10-MHz range and later, for perhaps the 3- to 30-MHz range. All frequencies above about 30 MHz were grouped as ultrahigh frequencies (UHF) or ultrashort waves.

through amplifiers and into a mile-long, 410-foot-high, multiply tuned antenna on a (suppressed) carrier at 60 kHz and in solving the problem of how to handle the unprecedented 60,000 to 100,000 w of radio-frequency (RF) power. This work culminated in the establishment of the successful circuit to England in 1927.[2] (See another volume in this series, *The Early Years (1875–1925)*, Chapter 5, Section 4.3.) But, even before the long-wave circuit went into service, the developers were aware that long-wave transmission would be inadequate. A single circuit could not be expected to satisfy the communications needs of the world, and there was very little room for expansion in the long-wave spectrum. In addition, thunderstorm activity in the summer would create very bad static conditions. Consequently, as soon as it was realized that shortwaves might serve for communications between the United States and Europe, work was directed toward developing circuits in the new range.

When established, the shortwave systems quickly took most of the traffic, but long-wave radio proved surprisingly durable. The two systems complemented each other. The long waves were seriously impaired by thunderstorm static while the shortwaves were not; the shortwaves were interrupted by disturbances of the ionosphere during magnetic storms, which had little effect on the long waves. A project to build a second long-wave link was well started in 1929, but was suspended by the onset of the Great Depression. In 1940, the first circuit was moved off the RCA antenna to a new transmitter at Bradley, Maine, and a second circuit was added. Long-wave operation continued until 1957, when submarine-cable circuits were established.

For shortwaves, the problems of stability, efficiency, and RF heating were serious, but understanding the vagaries of shortwave propagation and designing high-directivity antennas were the major new challenges. By the middle of 1925, a shortwave transmitter using newly developed tubes had been constructed at Deal and installed in an outlying building [Fig. 2-2]. Originally, the transmitter operated at the lower end of the band, but, by October, it was operating at frequencies up to 6.8 kHz and a little later at 9.1 MHz. Within a year, it was operating at frequencies up to about 20 MHz. This transmitter had a crystal oscillator followed by a harmonic generator for the various radio frequencies to be transmitted. This was closely coupled to a pair of 250-w tubes operating in the push-pull configuration as a plate-modulated controlled oscillator.

In the early Deal transmitter, the 250-w stage drove water-cooled push-pull stages with two 10-kw tubes in the output. This transmitter was later replaced by a prototype of the commercial version with six 10-kw water-cooled tubes in the final stage [Fig. 2-3]. This approach, the generation of a modulated signal at a fairly low level, followed by linear amplification, was to be a feature of Bell Laboratories radio designs for many years.

Hourly field-strength measurements were made in New Southgate, England on transmissions from Deal one day a week for about two years [Fig. 2-4]. The results were cabled to New York, analyzed, and subsequently published. They were supplemented later by data collected at the receiving station at Netcong, New Jersey, and by measurements made on signals received at Buenos Aires from the United States. Transmission was also tracked by ship from New York to Bermuda and later, aboard the SS *Leviathan*, to England. In some of the

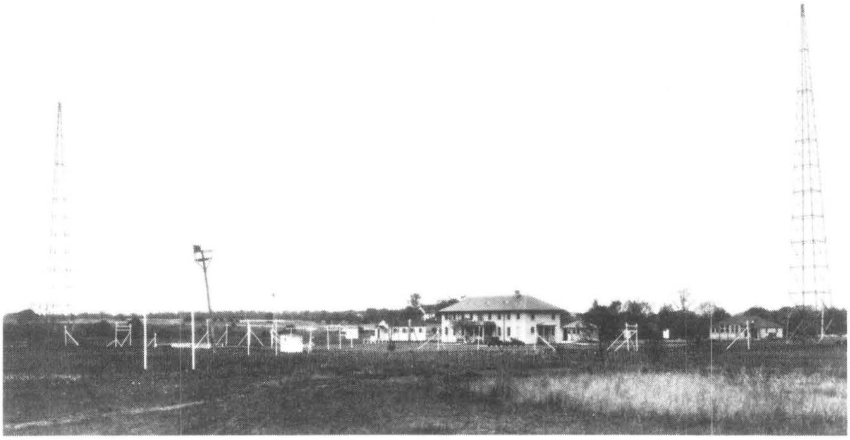


Fig. 2-2. A general view of the transmitting station and antenna at Deal in the 1920s.

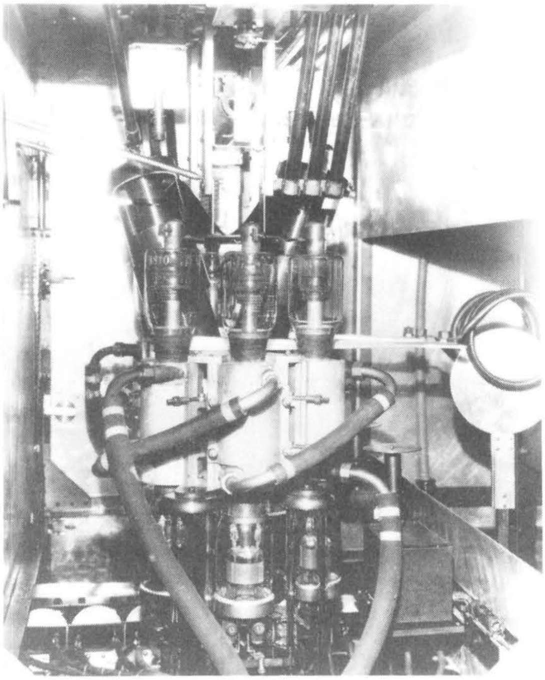


Fig. 2-3. Final stage of radio amplification for engineered model, shortwave transmitter, May 1930.

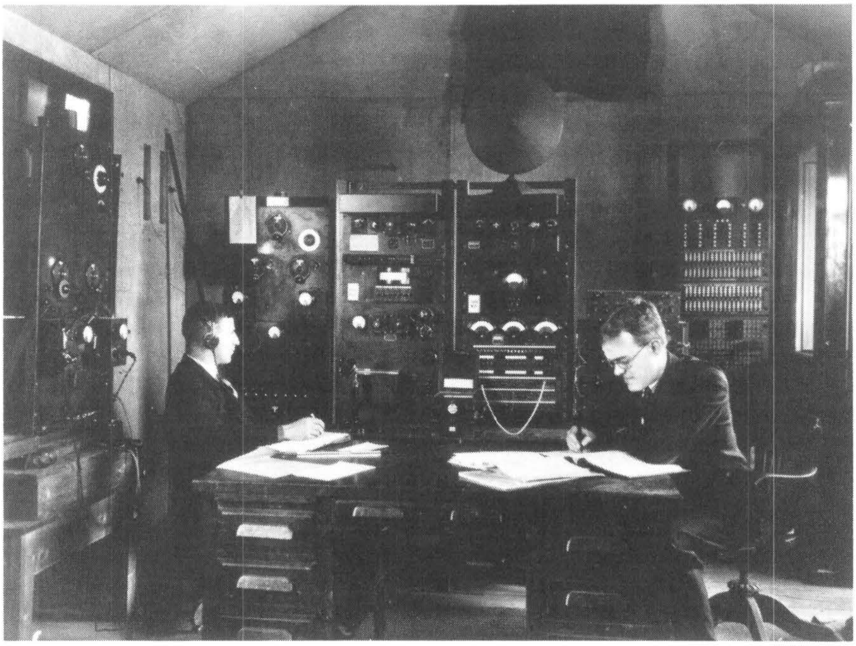
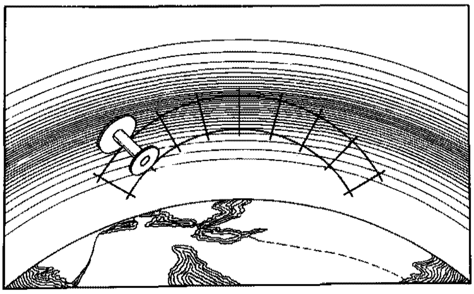


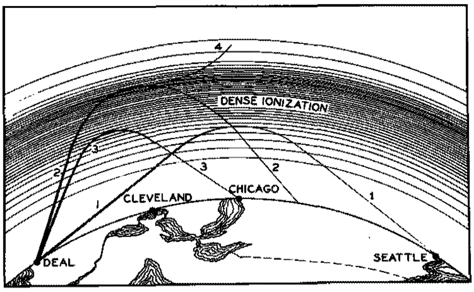
Fig. 2-4. The shortwave receiving laboratory at New Southgate, England.

tests, modulating the carrier with several audio tones showed clearly that selective fading within a voice band did indeed occur as a result of multipath transmission.[3]

As a result of these tests, and the work of the Marconi Co. and many others, a fairly clear picture emerged of the mechanisms governing high-frequency propagation. As a wave front entered the ionosphere at an angle, the portion that first entered the higher-altitude region of greater ion density was propagated more rapidly than the portion in the lower regions. As a result, the wave was refracted and redirected back to the earth. Only waves within a certain range of angle of approach would be returned, and that range depended in turn on the time-varying altitude and composition of the ionosphere [Fig. 2-5]. This mechanism explained the skip effect and the observed daily and seasonal changes in propagation.[4] To be useful, however, these results had to be reduced to an applicable form. At an early stage of the work, it had become apparent that, for reasonably continuous transmission, frequencies would have to be changed at least twice and perhaps as often as three or four times a day. Field-strength measurements of summer and winter shortwave-radio propagation characteristics, made in the late 1920s and early 1930s, provided the most comprehensive information available on the subject for some years [Fig. 2-6]. They indicated the diurnal and annual changes of frequency that were necessary, as well as the time when no transmission could be obtained.



(a)

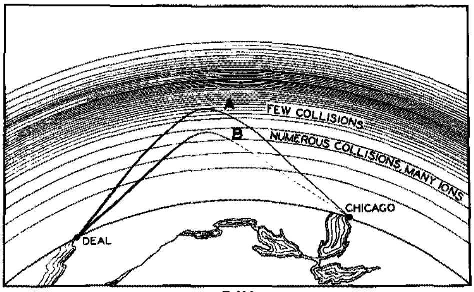


(b)

Fig. 2-5. Shortwave propagation mechanisms. (a) Wavefronts, symbolized by the axle, are refracted back to earth when the portion in a region of higher ion density is propagated more rapidly (larger wheel) than the rest of the wave. (b) The angle of return, and hence the skip distance, depends on the angle of incidence and composition of the ionosphere (1 and 3). Waves too close to the vertical may not return at all (2).

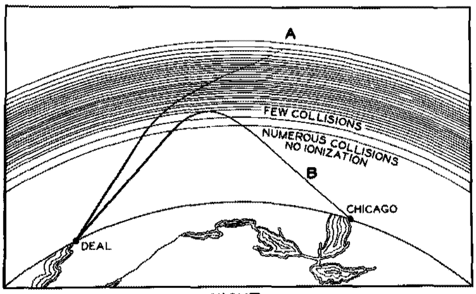
During the Deal tests, any of four frequencies (6.6, 9, 13.6, and 18.7 MHz) were used according to the momentary propagation conditions. The inductance coils (tank coils) in the high-power amplifiers had to resonate with the interelectrode, stray capacitance, and whatever trimming capacitors were added. This meant carrying hundreds of amperes in the coils, and, because of the high frequencies, the current was confined to a skin depth of about 20 μm. Polished-copper tubing, sometimes larger than 1 cm in diameter, was used, making the transmitter resemble an alcohol still. Changing wavebands in the early shortwave transmitters was a strenuous exercise involving bolting and unbolting a number of the heavy copper-tubing coils and manually retuning the high-power stages. This activity actually became a sort of competitive sport—a race against time with occasional odd results, as when one engineer left his loaded tobacco pipe in the high-power tank circuit, fumigating the entire station.

In the spring of 1927, a transportable 500-w shortwave transmitter was assembled in two large moving vans to explore sites for commercial stations.



DAY

(c)

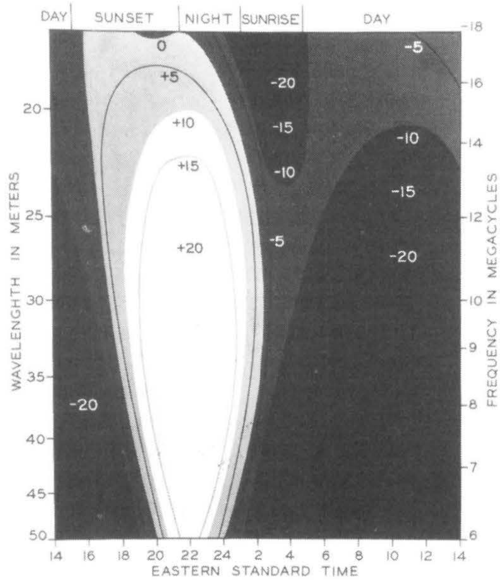


NIGHT

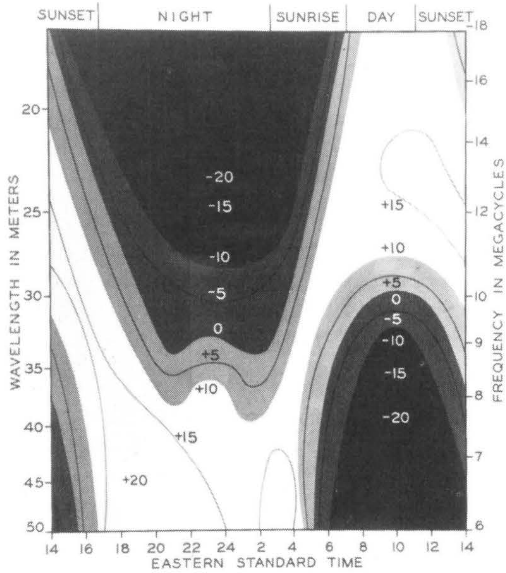
(d)

Fig. 2-5. (c) The day-to-night changes in the composition and structure of the ionosphere change the refraction and attenuation of different wavelengths. During the day, wave A of, for example, 25 m, is refracted to reach Chicago, but is not returned at all at night. (d) Wave B, of 100 m, is strongly attenuated during the day, but at night, when the ions in the lower atmosphere have disappeared, it is returned, with much less loss, to reach Chicago.

In the fall, this equipment was taken to Houlton, Maine, where it was used to transmit to collaborators in England for several weeks on an alternate schedule with Deal to determine their relative merits as transmitting sites. It was found that the Maine site gave signals that were on the average 9 dB stronger, but it was eventually decided that the advantage was not worth the cost of the 800 miles of wire lines necessary to connect the more remote location. During the same period, a number of possible sites for receiving stations were examined using portable field-strength measuring equipment [Fig. 2-7]. A plateau just south of Netcong was chosen as a receiving site because of its freedom from local noise sources and the absence of potential sources of interference along the great-circle radio path toward England. A temporary building was erected in May 1928 to house the equipment for the first channel, and a receiving antenna was erected in front of this building. A site west of Lawrenceville, New Jersey was selected for transmitting.



(a)



(b)

Fig. 2-6. Relative field strength, as measured from the United States to England, 1926 and 1927. (a) Summer. (b) Winter.

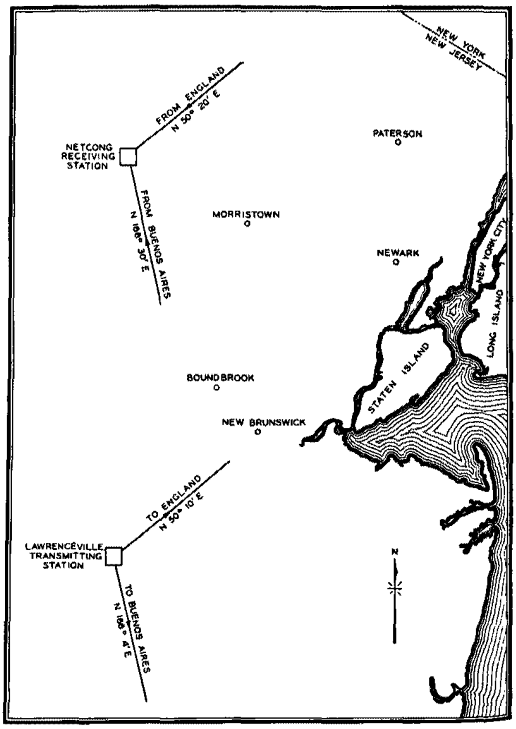


Fig. 2-7. Map of northern New Jersey showing transmission paths affecting location of stations.

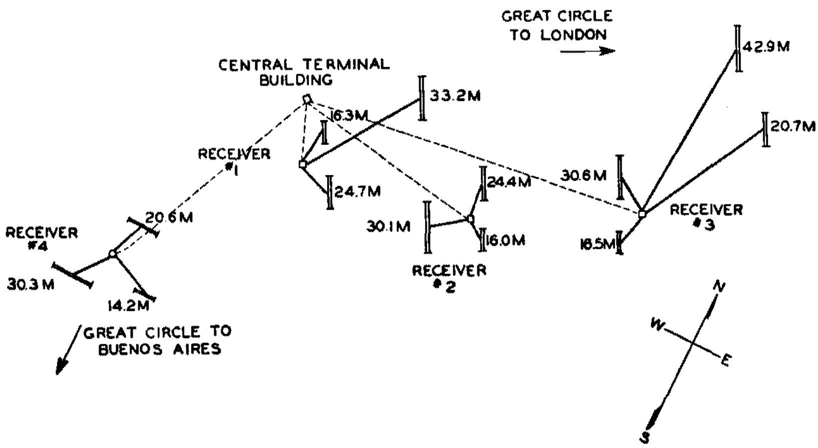
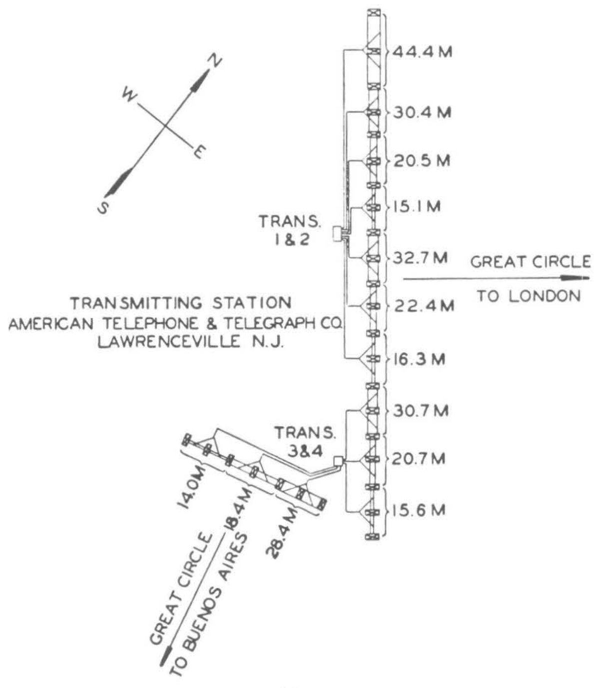
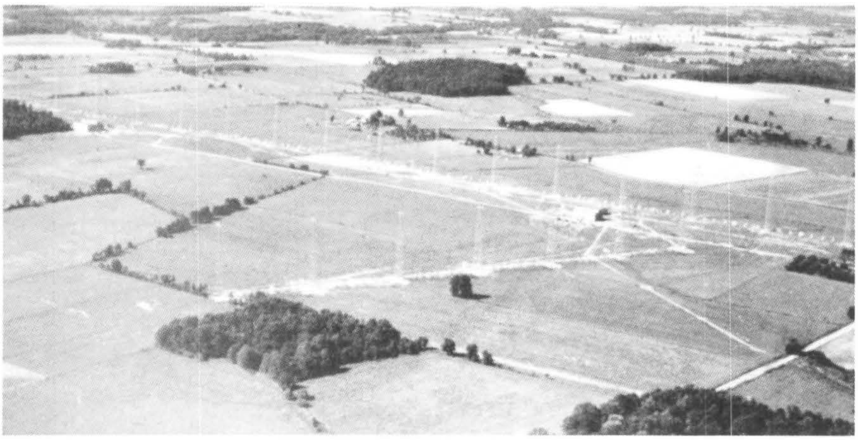


Fig. 2-8. Arrangement of antennas at Netcong receiving station.



(a)



(b)

Fig. 2-9. The Lawrenceville transmitting station. (a) Arrangement of antennas. (b) General view of antennas.

Meanwhile, events on the business front were moving rapidly. The success of the first long-wave circuit resulted in a rapid growth of traffic. Early in 1928, an agreement was reached with the British Post Office to proceed with commercial shortwave service across the North Atlantic. Frequency allocations were obtained from the Federal Radio Commission, the predecessor of the Federal Communications Commission, and weekly conferences were started among the engineers at Bell Laboratories, the AT&T Development and Research Department, and the Long Lines Department. It was established that the equipment was not to be considered experimental but should be the best feasible in the existing state of the art.

The initial objective was to establish three high-frequency (HF) links to England, to be followed immediately by one to Buenos Aires. The Deal transmitter was pressed into service as a backup to the long-wave circuit on the west-east link even before the station in Netcong was ready. When the first Netcong antenna and receiver were in place in June 1928, two-way shortwave service was established to England just in time for the summer season, when transmission on long waves was disturbed by high static. Since the antennas were tuned and therefore usable over only a small frequency range, each channel had to be provided with three separate antennas operating at the low, middle, and high end of the band and suitably spaced to prevent detrimental interaction. By February 1929, the three antennas for the first channel had been completed at Netcong. Separate buildings and antennas were erected later in the year at three other widely separated places on the property to handle the other channels. A permanent terminal building was later constructed just below the west edge of the plateau, where any noise generated would not interfere with reception [Fig. 2-8].

With Deal available as a transmitter site, work at Lawrenceville proceeded more deliberately. Permanent masonry buildings were designed, requiring some months longer than the Netcong station for completion. Twenty-eight steel towers 180 feet high were erected and antennas suspended from them. Each antenna consisted of an array suspended from three towers. One antenna array was required for each frequency and transmitter—a total of 13 antennas, each roughly 180 by 500 feet [Fig. 2-9]. (One channel was equipped with four antennas for operation at a frequency as low as 6.8 MHz, waves 44.4 m long.) Transmissions from the experimental station at Deal continued until the three channels were put in operation at Lawrenceville in June, September, and December 1929.

Work on high-frequency radiotelephony was progressing rapidly in Europe as well as in the United States, and a number of circuits were established within a year or so of the first HF transmissions. In September 1929, a demonstration was held in conjunction with the British Post Office, in which Bell System officials in New York talked with Australia via London, a total circuit length of 15,000 miles. Considering that overseas telephony was only two years old, this was an historic event.

II. SHORTWAVE ANTENNAS

The theory of multielement antenna arrays had been worked out by G. A. Campbell in 1919,[5] but it was not until the HF work at meter wavelengths that highly directional arrays became practical. The basic element of the early tuned antennas was the half-wave dipole. The problem of exciting array elements was eased with the realization that an element slightly longer than a half wavelength acts as a reflector to incident energy, and a shorter element acts as a director. The early shortwave transmitting antennas consisted of simple vertical half-wave dipoles with center inductive loading or two dipoles in parallel in the form of an H, with or without reflectors. Four coplanar dipoles separated

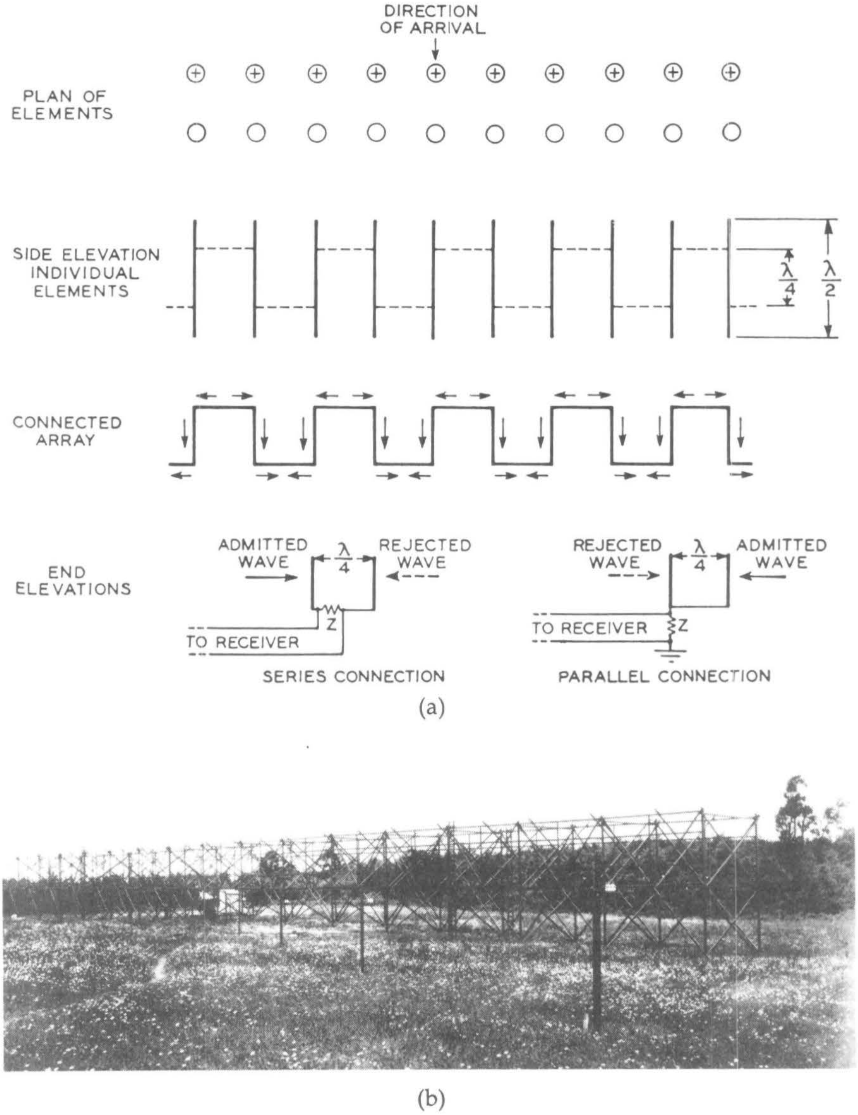


Fig. 2-10. High-frequency receiving antennas at Netcong. (a) Schematic of Bruce Array antenna. (b) Receiving antenna, with its supporting structure, 1929.

by a half wavelength, with reflectors—or, alternatively, a single dipole with three passive reflectors forming a rudimentary cylindrical parabolic reflector—were also used. Thermocouple ammeters were placed in the centers of the driven dipoles and were read from the ground using field glasses.

The antennas used at Netcong were of a form originally called the Grecian Key, but subsequently called the Bruce Array, after its inventor.[6] This was a broadside array of vertical half-wavelength elements, each connected to the adjacent element, spaced a quarter wavelength away, by bending the upper and lower eighth-wavelength sections horizontally [Fig. 2-10]. The currents generated in the vertical elements by a wave of the correct frequency arriving perpendicular to the array were all in phase and added in the receiver by connecting the elements alternately at the top and bottom. No reradiation occurred from the horizontal connecting elements, because currents were oppositely phased about their center point. A reflector of similar character, spaced a quarter wavelength behind the first array, made it unidirectional and increased the gain by 3 dB. The directional pattern of the antenna was measured by circling it with a small transmitter at a range of about 1000 feet [Fig. 2-11]. The gain of the Netcong antenna was 40 times that of a single vertical half-wave antenna (16 dB). K. G. Jansky, using a smaller, but steerable, antenna of this type, discovered radio-frequency interstellar radiation from the Milky Way in the early 1930s.[7,8]

Because of the mechanism of high-frequency propagation and its variability, sharp vertical directivity was not desirable in the receiving antenna. Less concern

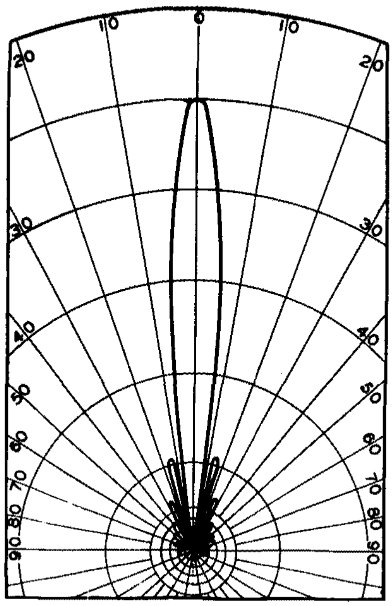


Fig. 2-11. Directional characteristic of Netcong antenna.

was felt about launching waves in a narrow vertical pattern, and in the transmitting antenna, vertical as well as horizontal stacking of elements was used.[9] The basic element was a so-called Sterba Array, consisting of one half-wave (1/2) and two quarter-wave (1/4) elements in a vertical lineup, paralleled by an identical, adjacent, three-element vertical line displaced horizontally a half wavelength away [Fig. 2-12]. The horizontal links acted as feeders for the vertical elements and radiated essentially no energy because of the oppositely phased currents in the closely adjacent pairs. These panels were replicated in a larger horizontal array to increase the horizontal directivity and were backed by a reflector to make them unidirectional [Fig. 2-13]. A gain of about 17 dB was realized. One virtue of the Sterba Array was that each panel or, indeed, several of them, could be formed from a continuous conductor. By the use of small capacitors (open circuits at 60 Hz, but short circuits at RF) and RF quarter-wave lines (short circuits at 60 Hz, but open at RF), it was possible to pass 150 kw of 60-Hz power through the No. 6 gauge conductors to melt sleet, even with the transmitter in operation.

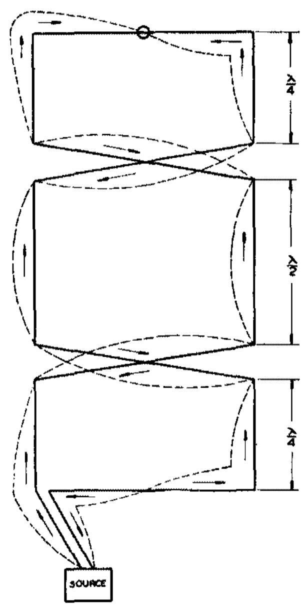


Fig. 2-12. Multielement transmitting antenna structure of the type used at Lawrenceville in 1928. A single conductor was bent to form one section of a directional antenna. Solid lines are conductors; dotted lines show current direction and amplitude.

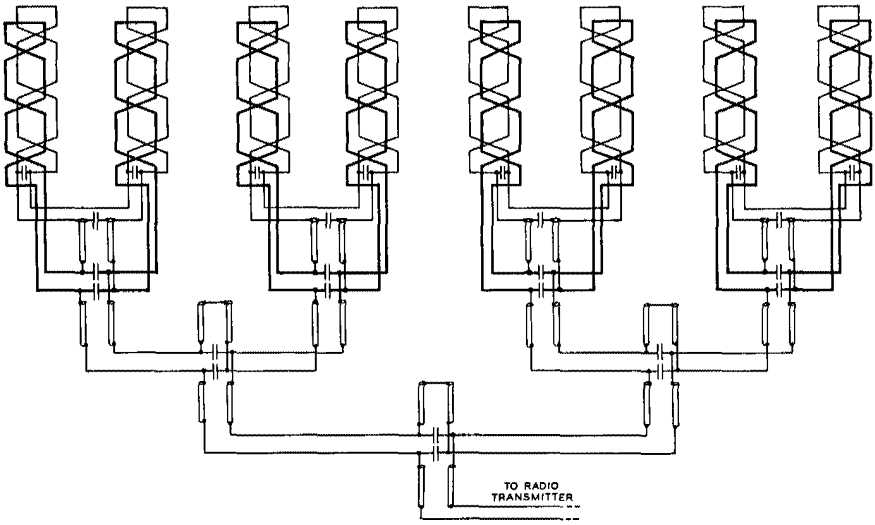


Fig. 2-13. A complete Sterba Array antenna with reflectors for one wavelength.

About the time the first antennas were placed in operation, E. Bruce had observed that a wire tilted toward the source of a wave could act as an effective antenna over a much broader range of frequencies than the rather sharply tuned simple vertical half-wave dipole. He then further observed that an even-broader-band (less-tuned) antenna could be achieved by connecting a second tilted wire to the first in an inverted V. Finally, working with H. T. Friis, he recognized that his vertically disposed, inverted V was completed to a double V, or diamond, by its reflected image in the ground.[10] It remained to turn the V on its side, and physically complete the four sides with actual conductors, to make a horizontal, broadband, HF, rhombic antenna [Fig. 2-14]. Unlike the

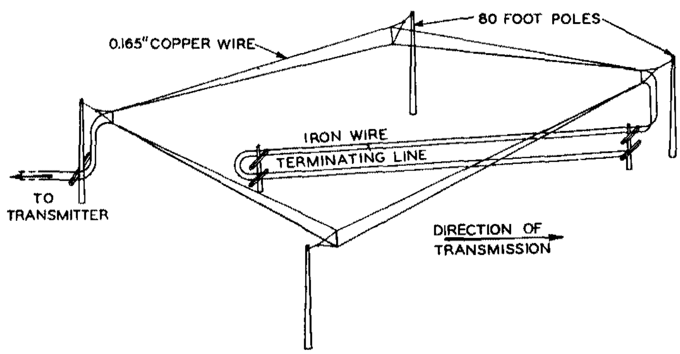


Fig. 2-14. Experimental rhombic antenna showing arrangement for iron-wire, high-power termination when used as a transmitter.

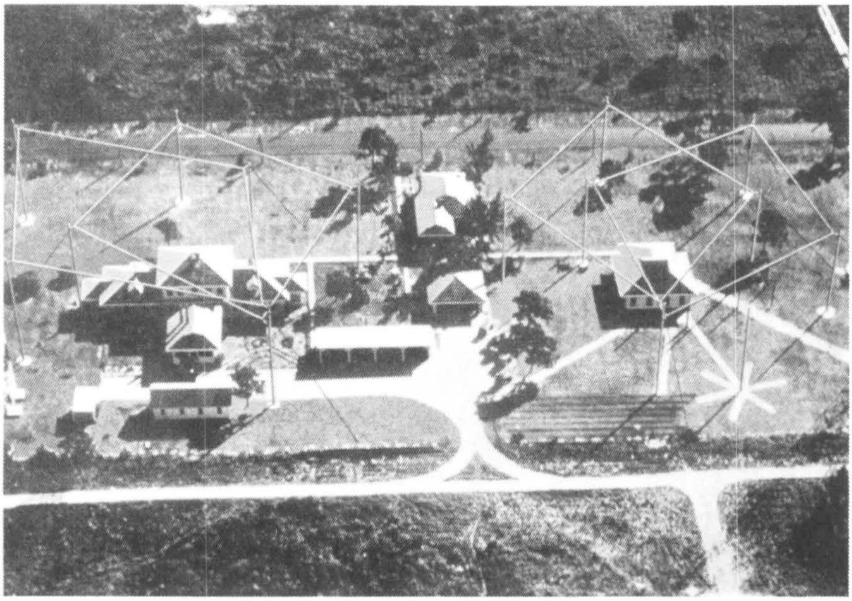


Fig. 2-15. Rhombic antennas near Miami for HF radio links to Colombia and Venezuela.

earlier arrays, the antenna was responsive to horizontally polarized waves. Working at 75 MHz so that models scaled to the shorter wavelength were easier to build and test, Bruce showed how to design such antennas for a wide range of horizontal and vertical directive patterns. He demonstrated gains of about 14 dB and achieved a response flat to ± 1 dB over a four-to-one frequency range.[11]

The rhombic antenna became immensely popular, both within and outside the Bell System, largely because of its simplicity and the fact that a single antenna could be used over nearly the entire HF band [Fig. 2-15]. Many papers were written on the subject, and although it was shown that the antenna had shortcomings and was frequently used under conditions where it should not be, it did not have a rival for HF use until the development of certain forms of the log-periodic antenna 25 years later. The rhombic antenna turned gigantic wire nets and many large towers into junk when it replaced the large arrays used earlier. (For a research perspective on shortwave antennas, see *Communications Sciences (1925–1980)*, Chapter 5, Section I.)

2.1 Multiple-Unit Steerable Antenna

The first receiving antennas were not made highly directional in the vertical plane because of the varying direction of arrival of the ionospherically confined waves. By the mid-1930s, considerable research had been done on the angle of arrival of HF radio waves in both the vertical and horizontal planes, and on the rate and range of variation.[12] The simultaneous arrival of waves at different vertical angles over paths of different lengths could cancel each other (interfere) and was the principal cause of fading at HF. It occurred to Friis that an antenna with sharp discrimination in the vertical plane would be very advantageous if it could be steered to track only one of the incoming waves [Fig. 2-16]. A single rhombic antenna was sensitive over a fairly broad vertical angle. With a linear array of antennas along a line toward the transmitting station and with the outputs combined in the correct phase relationship, a higher gain and more directive vertical pattern could be maintained. The array beam could be steered in the vertical plane by varying the phase of the signals received from each element of the array before combining them. In 1935, Friis built an array of six rhombic antennas at Holmdel and demonstrated an increase in gain of 6 or 7 dB over a single antenna and the ability to steer the narrow lobe by adjusting the relative phase shift between the units.[13] This design was named the multiple-unit steerable antenna (MUSA) [Fig. 2-17].

Something of an heroic apex in HF transatlantic technology was reached in 1939 with the design and construction of a commercial MUSA receiving station. This station at Manahawkin, New Jersey featured a two-mile-long array of 16 rhombic antennas deployed along the great-circle path toward the transmitter in Rugby, England. The highly directional and vertically steerable array was built in anticipation of a poor period of HF transmission due to a sunspot

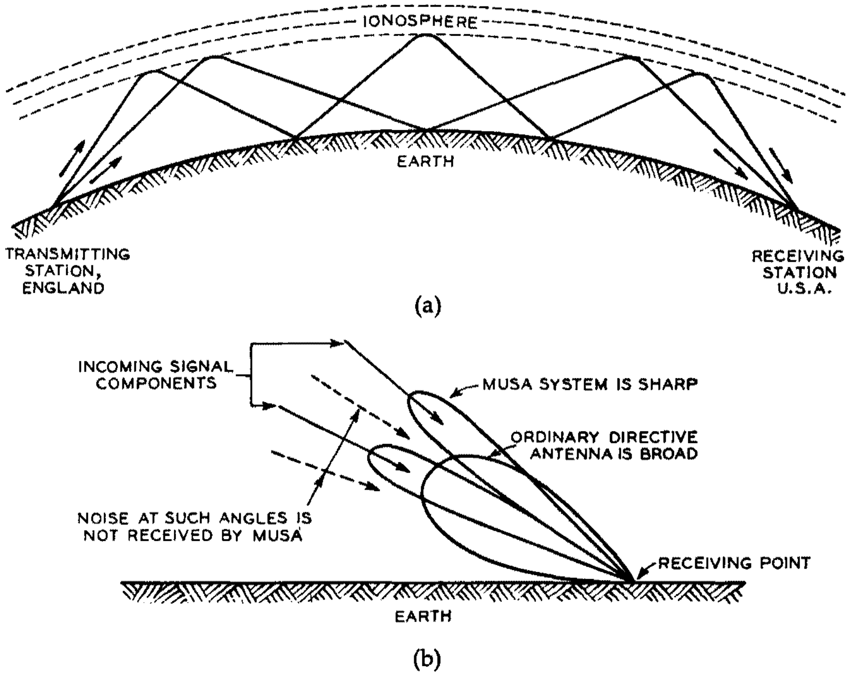


Fig. 2-16. Paths of shortwaves over long distances, determined by the angle of reflection from the ionosphere. (a) Idealized wave paths. (b) Vertical-plane directive patterns.

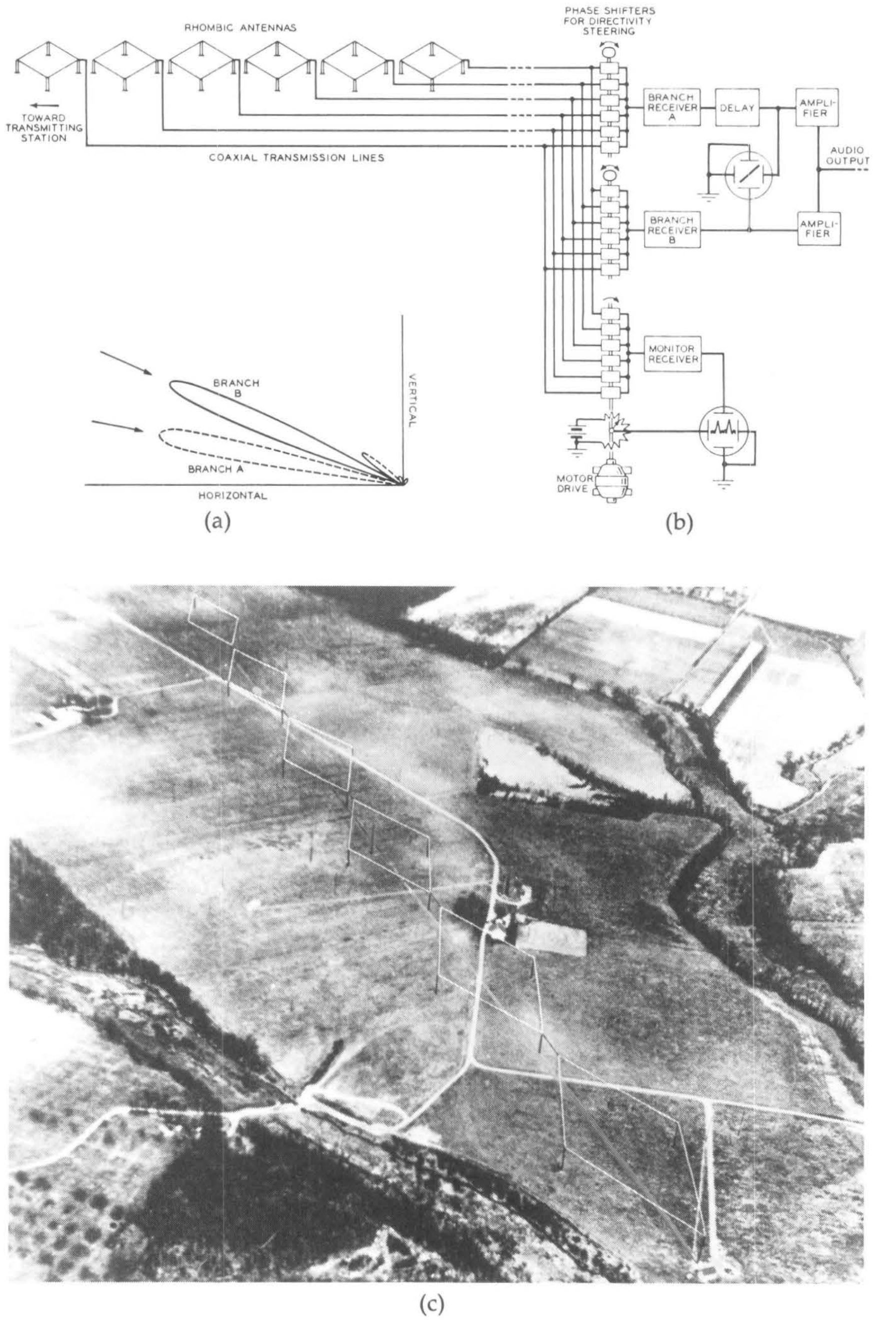


Fig. 2-17. Multiple-unit steerable antenna (MUSA) operation. (a) Antenna patterns. (b) Schematic. (c) Holmdel six-rhombic experimental MUSA.

maximum in the early 1940s.[14] The station building was located at the ninth antenna to reduce the total amount of interconnecting coaxial cable required. All the cables were buried to reduce the phase changes that would otherwise occur with changes in temperature. The Manahawkin site was a flat marshland with lots of room, ideal for the purpose.

In a further elaboration, three banks of adjustable phase shifters connected in parallel permitted training the array on as many as three incoming rays and combining the outputs, with suitable delay correction, to realize a form of angle-of-arrival diversity. An additional bank of phase shifters was continuously rotated to provide an output that scanned the vertical-angle range once a second, and furnished information on incoming-wave directions needed to adjust the other banks [Fig. 2-18]. It had been observed that, while fading due to interference from multipath transmission might be rapid, the direction of arrival of the various wave paths changed rather slowly. However, monitoring and manually adjusting for the constantly changing transmission paths would have been exceedingly tedious and susceptible to errors that could easily negate the advantages of the array. Automatic setting of the angles of reception and the delay adjustment of the diversity branches solved the problem.

Since the array was broadband and capable of receiving signals over the entire HF band at all times, it could be used for more than one receiver at a time. At Manahawkin, a second receiver at a different frequency, with its own

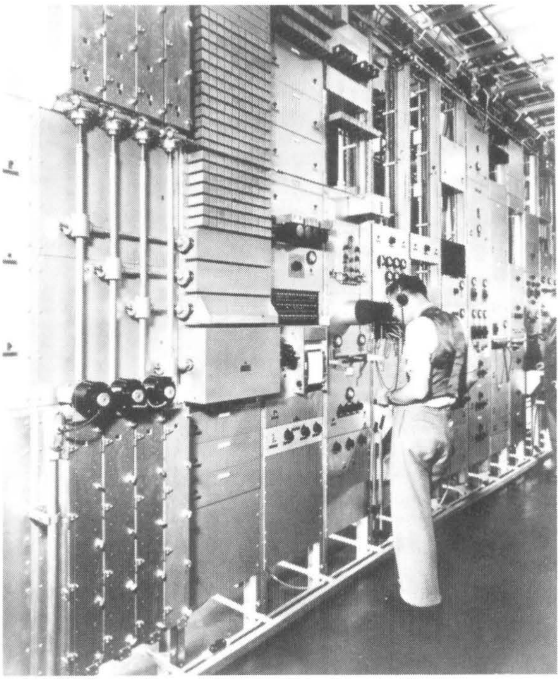


Fig. 2-18. Manahawkin MUSA receiver bays. *Left*, phase shifters for three beams.

banks of phase shifters, was connected and used simultaneously. (See *Communications Sciences (1925–1980)*, Chapter 5, Section 1.5.)

Performance tests indicated a substantial improvement in quality over a single receiver, and, during World War II, a number of channels of encrypted digital signals were successfully transmitted using the MUSA. But the improvement in the signal-to-noise ratio was appreciably less than the number of antennas would indicate, and performance was somewhat disappointing. It was observed that selecting the best received audio signal gave essentially the same performance as delay compensation and combining, with much less maintenance. A simple selection of the best branch was used most of the time.

III. RECEIVERS AND TRANSMITTERS

3.1 Double-Sideband Operation

The receivers at Netcong were of the double-detection or superheterodyne type; that is, they had tuned RF stages followed by a frequency conversion to an intermediate frequency, where most of the gain was supplied, followed by the final recovery of the audio signal (the second detection). The circuits were calibrated and controlled, so that quantitative measurements of signal characteristics could be made. A volume indicator and shelf for telegraphic order wire were also provided.

At long wavelengths, receiving sensitivity was limited by atmospheric disturbances caused by lightning, auto ignitions, and other sources of interference, so that noise internal to the receiver was not a problem. Above 10 MHz, the noise limiting reception was that generated in the receiver rather than in the medium. Consequently, considerable attention was given to obtaining a high signal-to-noise ratio in the receiver. These receivers are believed to be the first commercial equipment to be rated with respect to thermal noise, which later became the recognized standard of reference. As part of the effort to approach the thermal-noise limit, multiple input circuits were provided, one for each frequency to be used, which were permanently connected to their respective antennas. DC-heated filamentary triodes were used, except in the RF stages where screen-grid tubes, which were just coming into use, were used to avoid the troublesome grid-plate capacitance of the triodes. In this and other respects, the new receivers were better than anything available on the commercial market. Automatic gain control, which adjusted the gain of the receiver according to the strength of the incoming RF signal, was at first used only in the intermediate-frequency stages, but was later applied to the RF stages as well to reduce intelligible intermodulation from strong interfering signals. It was a continuing problem to obtain gain control of sufficient speed to track the fading without cutting into the lower-frequency speech components.

A number of double-sideband receivers were supplied to the International Telephone & Telegraph Company (ITT) as well as to AT&T. These receivers gave excellent service for more than 20 years and were removed from service only when single-sideband transmissions completely superseded double-sideband transmissions in the 1950s.

3.2 Single-Sideband Operation

All the early work on HF radio used double-sideband amplitude modulation. Single sideband promised a signal-to-noise advantage of 9 dB, 6 dB by using the entire power capacity of the transmitter for one sideband instead of the carrier plus both sidebands and 3 dB by halving the noise bandwidth. (See *Communications Sciences (1925–1980)*, Chapter 1, Section 2.1.1.) But the problem of frequency stabilization required for carrier suppression and recovery was beyond the capability of the equipment of that early date. Single-sideband equipment was built and tested as early as 1929, but the required manual frequency synchronization was not practical even for tests.

In 1931, a receiver was built and installed at Netcong that could be used for receiving double-sideband signals but that could also remove one sideband and keep, reject, or recondition the carrier to simulate single-sideband reception of incoming double-sideband transmissions. This receiver was used to determine, without having to build and test a complete single-sideband system, whether there was anything peculiar about single-sideband shortwave propagation that would prevent the successful exploitation of this means of transmission. The receiver consisted of seven 6-foot bays of equipment and, so far as is known, included the first crystal band filters to be used in any radio equipment. The filters were used at the 150-kHz intermediate frequency to obtain sharp discrimination against closely adjacent frequencies. Tests went on for several years (it was the Great Depression) and showed no insurmountable obstacle to single-sideband operation.

In 1935, designs went forward for commercial single-sideband operation. The equipment used crystal filters in both transmitter and receiver, and had two 6000-Hz passbands, one on each side of the carrier. A carrier 20 to 26 dB below normal was transmitted and used for automatic volume control and automatic frequency control at the receiver. The carrier was filtered out and amplified separately (reconditioned) and then reinserted at greater than normal amplitude in a balanced demodulator to give low distortion.

The first single-sideband equipment was manufactured at the Bell Laboratories model shop at West Street in Manhattan. Later, in the 1940s, large numbers were manufactured by Western Electric and sold to foreign customers of AT&T, government departments, and others. They set a standard of excellence, with performance far better than that of any previous HF radio equipment. They promoted the worldwide shift from double-sideband to single-sideband operation, with substantial savings in frequency spectrum, reduced interference, and increased transmission efficiency.

3.3 Multichannel Operation

With the advent of single-sideband operation, it was recognized that, if intermodulation could be kept low enough, it would be feasible to put an independent second channel on the opposite side of the carrier from the first channel. For economical operation of radio transmitters, however, it was necessary to drive them so hard that the distortion products could be too high for more than single-channel operation. In December 1936, tests at Lawrenceville showed it was possible to get acceptable two-channel operation by transmitting one band next to the carrier, and one on the other side of the carrier, but separated from it by one speech bandwidth (3 kHz). This avoided interference from the largest third-order intermodulation products, those that extend only one speech bandwidth on either side of the transmitted band. Two-channel operation was used extensively in the period before World War II.[15]

During the war, the Long Lines Department obtained another channel by using the frequencies adjacent to the carrier, previously left vacant, and combining the multichannel transmission with privacy equipment. The privacy equipment split the voice band into five subbands and transmitted them with the order scrambled and with some of the subband frequencies inverted. (In more elaborate arrangements, the scrambling-descrambling order was periodically changed synchronously at the transmitter and receiver.) In the three-channel transmission, the 15 privacy bands were intermixed on both sides of the carrier. At a still later date, with more linear transmitters having lower intermodulation, four 3-kHz channels were transmitted in a 12-kHz band.

When more highly directive and steerable antenna systems were proposed in the mid-1930s, it appeared that a much higher-capacity multichannel system could become possible. Studies were made of the possibility of obtaining bandwidth in the equipment capable of handling up to 48 voice channels, and experimental receivers were designed for such possible broadband use. In the late 1930s, investigation focused on the more modest objective of transmitting up to 12 channels simultaneously in the HF range. At Deal, a 20-kw amplifier was successfully tested in which sufficient feedback was applied directly at radio frequency to keep distortion within permissible limits for 12-channel operation, but further development was interrupted by the war. After the war, nationalistic considerations opposed transmission of a large block of channels through one nation for distribution to others, and for this reason, as well as a general withdrawal in Bell Laboratories from work on high-power apparatus, the project was not resumed.

3.4 Voice-Control Terminals

In 1932, a new transmitter was introduced that, to save power, suppressed the carrier when no speech was present. A new receiver was designed to work with this transmitter, using the new AC heater-type tubes then becoming available, and equipped with a CODAN (carrier-operated device, antinoise) to control the receiver when working with the intermittent carrier. Considerable trouble was experienced with the CODAN operation, and various parts of the system were rebuilt from time to time in an effort to improve operation. When the carrier was off for long periods during the time when no speech signals were being received, the automatic tuning control would drift so that the signal sometimes would not be received when the carrier reappeared. Finally, in the summer of 1936, it was agreed that future circuits to England would use single-sideband, reduced carrier operation, and the development of the CODAN and improved double-sideband equipment was dropped.

The CODAN was only one of an entire family of radio-terminal devices designed to cope with the difficult problem of associating the highly variable overseas radio link with the wire lines of the continental networks. The connecting land lines between the subscriber and the radio terminal had a wide range of loss and, as always, talkers spoke with a wide range of speech volume. In time, the terminals were elaborated with VOGADs (voice-operated, gain-adjusting devices), compandors (contraction for compressor-expandors) for syllabic-rate, speech-volume range compression and expansion, VODASs (voice-operated devices, antisinging), and privacy equipment [Fig. 2-19]. VO-

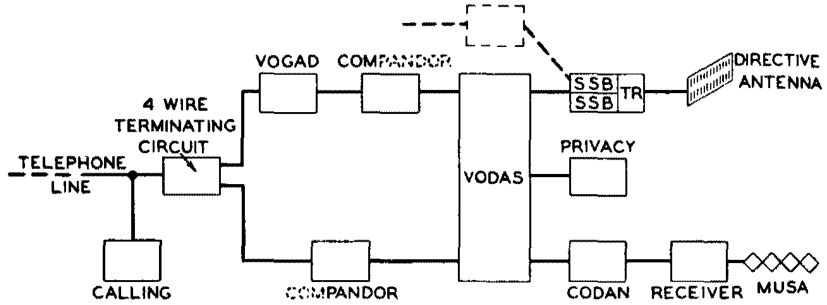


Fig. 2-19. Schematic of high-frequency radio terminal showing location of voice-operated devices (VOGAD, compandor, VODAS, and CODAN) and primary circuit.

GADs were relatively slow gain controls that pushed the transmitter to full modulation, regardless of the connecting line loss or basic talker volume. Compandors operated by compressing the speech-volume range at the transmitter and restoring the range at the receiver. Compression offered the advantage of reducing the level of high-volume talkers, making it possible to operate the transmitters closer to their maximum output for average talkers, but most of the signal-to-noise advantage was realized by raising the level of the weak signals for transmission over the noisy radio link. When the weak signals were reduced in level in the expandor, the noise was reduced with them. Compandors gave a substantial improvement in the perceived signal-to-noise ratio and found application in several later transmission systems.[16] VODASs opened the return circuit to prevent singing under some of the high-gain conditions. A number of privacy devices were developed, including one that periodically inverted the frequencies of the speech band, restoring them synchronously at the receiving terminal. Most of these gadgets were props, of course, to improve transmission that was, by current standards, not very good. However, there were no alternatives. It should be remembered that radio was the sole means of overseas voice communications for almost 30 years.

IV. THE SPREAD OF HIGH-FREQUENCY SERVICE

Shortly after the start of service to England, Bell Laboratories, in collaboration with ITT, engineered an HF link to a station near Buenos Aires. Service began in April 1930. Within the next few years, additional links were established to South and Central America via an AT&T station near Miami. West Coast stations in California at Dixon (transmitting) and Point Reyes (receiving) were built for circuits to Hawaii and beyond. By 1939, there was a substantial worldwide network [Fig. 2-20].

In addition to the overseas circuits, ship-to-shore HF service was also established, with a transmitter at Ocean Gate, New Jersey and a receiving station at Forked River, New Jersey. The transmitting station at Ocean Gate was like that at Lawrenceville in most respects. In fact, in the course of time, the Ocean Gate and Manahawkin stations were often used interchangeably for high-seas mobile and transoceanic point-to-point operation. The service to ships was considered experimental for several years, but the growth in traffic, except for

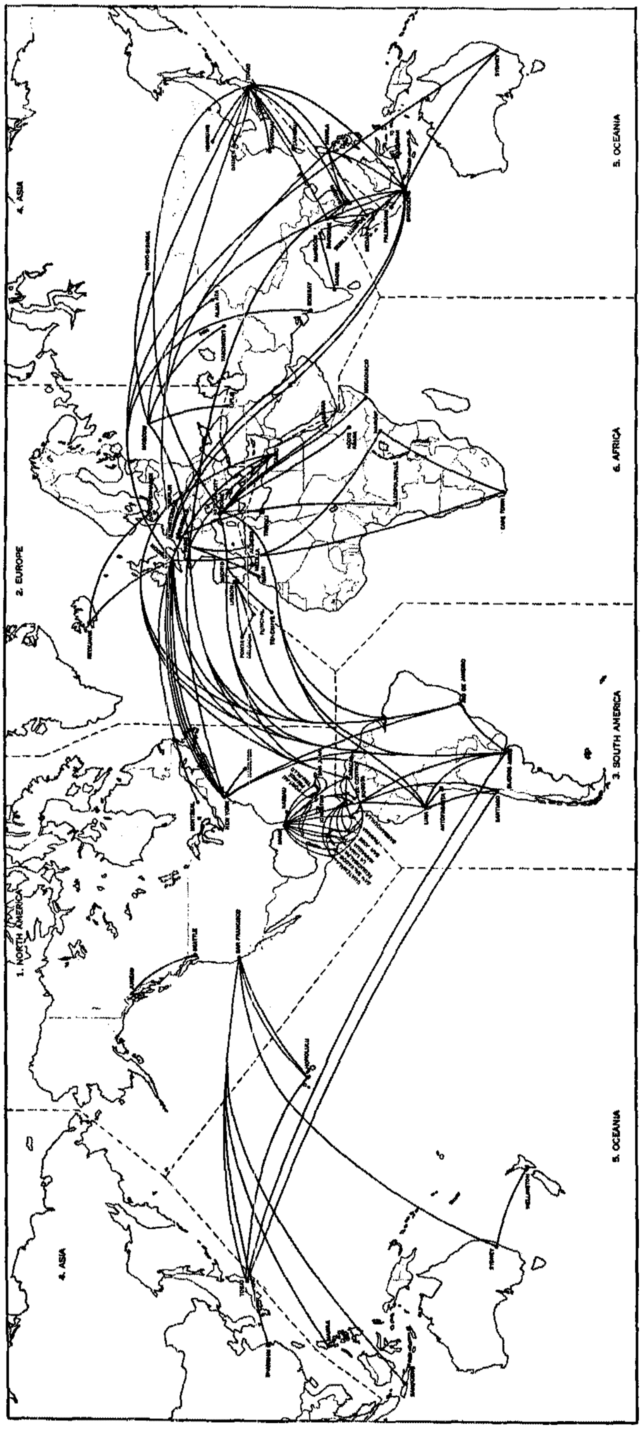


Fig. 2-20. Transoceanic radiotelephone circuits in operation as of January 1, 1939.

an interruption at the beginning of World War II, was rapid, reaching 200,000 messages per year by 1946.

During the war, no new HF equipment was designed, but following the war, a number of improvements were made. In 1946, a smaller receiver, coded LD-R-1, with much more convenient frequency-changing provisions was tested at Netcong. Substantial numbers were manufactured by Western Electric starting in 1948. In about 1950, a transmitter with sufficient linearity for four-channel operation went into service. Four-channel operation became widely accepted, and a number of manufacturers subsequently produced similar equipment.

Some other relatively minor developments continued, both in the Bell System and elsewhere, but the prospect of submarine-cable circuits and, later, of satellites, inhibited major changes. Shortwave radiotelephony was still used for communication to a few remote areas in the mid-1970s, but it had been completely replaced by high-capacity submarine cables across the Atlantic and Pacific Oceans and, even in remote areas, was rapidly being displaced with the ubiquitous satellite circuits. A sentimental reunion ceremony was held for the many engineers who contributed to the development of the shortwave art upon the occasion of closing the Lawrenceville station in February 1976. Most of the people had already retired, but history would have been kinder to them if it had taken a bit longer for the shortwave radiotelephony art to become obsolete.

REFERENCES

1. E. H. Armstrong, "The Spirit of Discovery—An Appreciation of the Work of Marconi," *Elec. Eng.* **72** (August 1953), pp. 670-676.
2. R. Bown, "Transatlantic Radio Telephone," *Bell Syst. Tech. J.* **6** (April 1927), pp. 248-257.
3. R. K. Potter, "Transmission Characteristics of a Short-Wave Telephone Circuit," *Proc. IRE* **18** (April 1930), pp. 581-648.
4. J. C. Schelleng, "Long Waves or Short," *Bell Lab. Rec.* **4** (June 1927), pp. 349-354.
5. G. A. Campbell, U.S. Patent No. 1,738,522; filed September 30, 1919; issued December 10, 1929.
6. E. Bruce, "Short-Wave Receiving Antennas," *Bell Lab. Rec.* **7** (August 1929), pp. 514-518.
7. K. G. Jansky, "Directional Studies of Atmospherics at High Frequencies," *Proc. IRE* **20** (December 1932), pp. 1920-1932.
8. K. G. Jansky, "Electrical Disturbances Apparently of Extraterrestrial Origin," *Proc. IRE* **21** (October 1933), pp. 1387-1398.
9. E. J. Sterba, "Short-Wave Transmitting Antennas," *Bell Lab. Rec.* **7** (August 1929), pp. 502-509.
10. E. Bruce, "Developments in Short-Wave Directive Antennas," *Bell Syst. Tech. J.* **10** (October 1931), pp. 656-683.
11. E. Bruce, A. C. Beck, and L. R. Lowry, "Horizontal Rhombic Antennas," *Proc. IRE* **23** (January 1935), pp. 24-46.
12. H. T. Friis, C. B. Feldman, and W. M. Sharpless, "The Determination of the Direction of Arrival of Short Radio Waves," *Proc. IRE* **22** (January 1934), pp. 47-78.
13. H. T. Friis and C. B. Feldman, "A Multiple Unit Steerable Antenna for Short-Wave Reception," *Proc. IRE* **25** (July 1937), pp. 841-917.
14. A. A. Oswald, "The Manahawkin Musa," *Bell Lab. Rec.* **18** (January 1940), pp. 130-134.
15. J. C. Gabriel, "Single-Sideband Short-Wave Receiver," *Bell Lab. Rec.* **18** (November 1939), pp. 84-87.
16. N. C. Norman, "The Voice-Operated Compandor," *Bell Lab. Rec.* **13** (December 1934), pp. 98-103.

Chapter 3

Broadcast Radio

I. THE SITUATION IN THE LATE 1920'S

The Bell System entered broadcasting as a natural extension and application of the technology it was developing for point-to-point radiotelephone communications. At one time or another, AT&T was engaged in almost every aspect of the broadcasting business, including station operation, program development, network interconnection, equipment manufacture, and the underlying research and development in propagation and measurements. (The Bell System's earliest activities in broadcast radio are covered in more detail in another volume of this series, *The Early Years (1875–1925)*.) AT&T withdrew from the station operation and programming business with the sale of its New York station, WEAF, to the Radio Corporation of America (RCA) in 1927. It continued as a major equipment supplier for 20 more years and provided program network interconnection on an ongoing basis.

In the equipment area, Western Electric concentrated on transmitters and studio equipment; little work was done on home receivers beyond the earliest years of broadcasting. Many basic improvements were made in Western Electric's transmitters; Station WEAF was crystal controlled at its assigned frequency of 610 kHz as early as June 1924. In the mid-1920s, however, home-receiver selectivity was improved, permitting discrimination between high-power stations on different frequencies. As field-strength surveys showed the need for more transmitter power to provide adequate received-signal levels in large urban areas, higher power became a major objective of the transmitter designers [Fig. 3-1]. The power of Station WEAF was increased from the initial 500 w in several increments to 5 kw by September 1925.

The early transmitters used glass-envelope radiation-cooled tubes for power output up to several hundred watts. By using the copper-to-glass seal developed by W. G. Houskeeper in Bell Laboratories (Western Electric, prior to 1925) and water cooling, designers were able to advance from 500 w to 5 kw in 1925, and to 50 kw a few years later. The first of the 50-kw designs, the No. 7A radio transmitter, was placed in service in 1929 at Station WLW, the Crossley station near Cincinnati. In the 7A, as in all the later high-power Western Electric transmitters, the design featured a low-level modulator and a linear power amplifier, in contrast to the radio-frequency (RF) output stage plate modulation used in most other transmitters at the time. Low-level modulation was accomplished in tubes and components of moderate size and power, with many

Fig. 3-1. Portable field-strength measuring equipment in use in the 1920s.

consequent advantages. A frequency response flat to ±1 dB from 30 to 10,000 Hz was provided with audio distortion less than 5 percent, and with RF harmonics only 0.03 percent of the carrier. Finally, with low-level modulation, the output stages could be driven to 100-percent modulation, that is, the RF envelope was varied from twice its unmodulated amplitude to zero, with little distortion. Since in early plate-modulated transmitters, distortion became excessive above about 50-percent modulation, this benefit alone was as effective as a four-to-one increase in carrier power. For all these advantages to be realized, it was, of course, necessary to provide high-power output stages of sufficient linearity to satisfy the exacting requirements of radio broadcasting with its broader band, greater dynamic amplitude range, and more critical signal content (e.g., symphonic orchestra music) than single-sideband telephony. However, the intensive research and development program on power amplifiers for overseas telephone transmission also provided the understanding that permitted the designers to meet the needs of broadcast radio.[1] The knowledge gained from single-sideband, suppressed-carrier transmitters carrying telephony of limited-volume range was successfully applied to broadcast radio's double-sideband, carrier-transmitted AM signal with its wide-volume range.

From the 1920s until after World War II, Western Electric also manufactured a wide range of studio audio and studio-to-transmitter equipment. This was important, not only to the radio-broadcast business, but, in the 1920s, was also linked to the pioneering work that led to talking pictures and, a little later, to the first high-fidelity stereo disc recordings. Equipment manufacture of radio transmitters and in related areas continued until 1948, when Western Electric withdrew from the business, and the designs and manufacturing facilities were sold to other suppliers.

Work in radio propagation, at least on terrestrial paths, has always been a somewhat arcane art compared to engineering in guided transmission. Transmission on wire pairs, especially after the most extreme effects of weather were excluded by the widespread use of sheathed cables, became a highly quantitative and precise technology. In the early years, radio propagation, in contrast, remained in many ways a mystery. There were large variations in loss over paths where none were expected. Even more baffling, reception was reported over paths where the accepted view held transmission to be impossible (e.g., the amateur reception of transatlantic shortwaves in 1921). In addition, arc lamps, automobile ignition, and nature itself contributed static from a variety of sources that varied in little-understood ways. Given the limited knowledge of the nature of the atmosphere as a radio-transmission medium, and that the existence, complex structure, and variations of the ionosphere were just beginning to be dimly perceived, it is little wonder that radio propagation remained more of an art than a science through the 1920s.

Bell Laboratories and its predecessor organizations in Western Electric led the field in quantitative radio measuring techniques and in their application in radio field-strength surveys. Work on improved radio field-strength measuring gear was begun in 1920 at the Cliffwood, New Jersey laboratory by C. R. Englund and H. T. Friis.[2] A portable loop antenna that could be transported in a car or light truck and readily calibrated in the field with an accompanying oscillator and attenuator made detailed measurements of absolute field strength possible. Field-strength measurements were first carried out at the low frequencies used for long-wave overseas radio, but they were later extended to the frequencies of the AM broadcast band and to much higher frequencies [Fig. 3-1]. Field surveys in the AM band in the New York City area showed that groups of large steel-frame buildings caused interference fading at night in the shadow zones of such clusters. This work led to an understanding of the need for higher transmitter power to provide adequate field-strength levels over a large metropolitan area. It also revealed the need for tight frequency control to prevent a fluttering type of fading in the interference patterns of the shadow zones. The exacting work on quantitative measurements in radio became characteristic of Bell Laboratories research in the field. The objective, as stated at the time by Friis, was "to demystify radio."

Despite the many problems inevitable in a new field, broadcast radio had advanced a long way in the 1920s and, by the end of the decade, both its technical nature and its business nature were much better understood than they had been in the earliest years. The functions needed for broadcasting were recognized by all parties, and organizational roles had been sorted out. The general view around 1930 in broadcast engineering was that there was little opportunity for progress.[3] Frequency response in radio transmitters was substantially uniform over the range of 30 to 10,000 Hz, even though commercial receivers were not responsive over this range. As far as could be determined from available measuring techniques (harmonic analyses of traces from string oscillographs), nonlinear distortion of the modulated wave did not exceed a few percent. Noise was about 60 dB below a fully modulated signal. Harmonic radiation at RF had been well suppressed by suitable filtering in the coupling networks leading to the antenna system. Temperature-controlled, quartz-crystal oscillators had satisfied all needs for carrier frequency stability.[4,5] Few thought that powers greater than 50 kw would be used. Business was headed into the Great Depression, but the depressed economy of the 1930s did not quench the enthusiasm of the budding broadcast industry. New technical advances made the decade a historic period in the growth of radio broadcasting.

II. THE QUEST FOR HIGHER POWER

Whatever radio engineers did or did not understand about radio propagation in the early days, they understood one thing very well—the more power transmitted, the higher the signal level received. In radiotelephony, higher power was used to overcome the large path losses of the overseas links and to override static. In broadcasting, designers knew that the received signal diminished roughly inversely as the square of the distance from the transmitter. To double the range and quadruple the area over which an adequate signal was received meant quadrupling the power. Interference between stations, of course, was another matter, but that problem was at least temporarily solved when the Federal Radio Commission increased and stabilized the band assigned to AM broadcasting. When clear channels were assigned, free of interference from nearby stations on the same frequency, broadcasters eagerly embraced the higher-power approach to increasing their coverage.

In overseas radio, the success of the first transatlantic radio links had greatly stimulated traffic, and most of the increase was served of necessity with the newly established shortwave high-frequency (HF) links. But shortwave transmission was complex and uncertain. In the early stages, circuit availability was not better than about 80 percent, and during periods of magnetic storms transmission might be interrupted for days at a time. The low-frequency transatlantic circuits on the other hand, although subject to severe static in the summer, were unaffected by the solar-induced magnetic storms. As a result, in 1929, it was judged advisable to provide additional long-wave circuits to supplement the existing HF circuits. In 1930, tubes with a peak power output in excess of 100 kw at a plate potential of 18,000 v became available. Work on a new long-wave transmitter was started with a plan for six such tubes in the final stage, having a combined peak power of 600 kw, the highest ever realized. The tubes were to be used in push-pull pairs. In this arrangement, each tube of the pair was biased at cutoff, with one tube polarized to handle positive signal excursions and the other negative swings. In the absence of a driving signal, little or no output current flowed, making the stage highly efficient. Three such push-pull pairs in parallel were to make up the output stage. It was planned that two such transmitters, separated in frequency by a few kilohertz in the 60-kHz range, would operate into a common antenna. At the same time, it was contemplated that two such tubes might compose the final stage for a new design of a 50-kw broadcast transmitter with a peak envelope power of 200 kw, although no such broadcast transmitter was then scheduled for development.

The radiated-power limits of all the high-power tubes were governed by the ability to dissipate the power that was not radiated. Efficiency, besides reducing power utility bills, was therefore as important as any other characteristic in achieving high radiated power. Exhaustive experiments at various

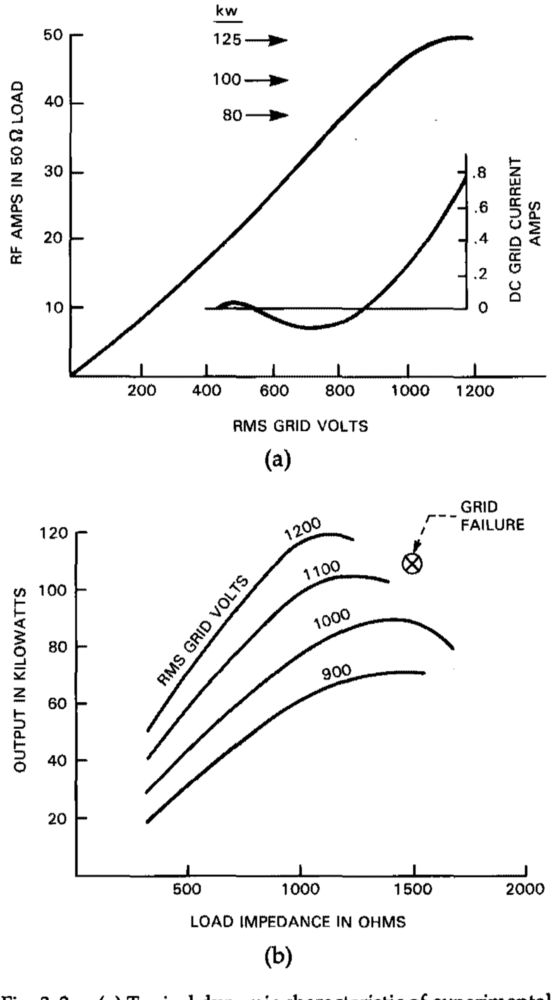


Fig. 3-2. (a) Typical dynamic characteristic of experimental 100-kw (peak-power) tube at a plate voltage (E_B) of 18,000 v and a grid bias (E_C) of −550 v. (b) Composite power diagram for high-power stage under different load-impedance and input-drive conditions. Optimum load impedance is approximately 1200 Ω. Observed grid failure was for sustained output and would not occur with a modulated wave.

grid biases and load impedances were carried out in the mid- and late 1920s to determine the best operating conditions for the output stages. Characteristics for grid bias and load impedance were selected for good linearity as well as high efficiency [Fig. 3-2]. The important load impedance for the output stage was the impedance of the antenna at the fundamental radio frequency, since the impedance at RF harmonic frequencies was made very low by the shunt capacitances of the tuned output circuit. The DC grid current into the output stage was important to the circuit designer, since it determined the required driving power from the previous stage. The nonlinear variations in the output-stage input impedance were made negligible by a shunt-resistance load for the driving stage low enough to avoid distortion of the input envelope. This ordinarily meant that the driving stage required an output power capacity five to ten percent of that of the final stage. Further improvements in HF radio and the onset of the Great Depression ended the work on the long-wave transmitter, but enough had been learned to ensure that the new high-power tubes, under appropriate operating conditions, would perform admirably in any new generation of broadcast equipment.

III. NEGATIVE FEEDBACK AND DIRECTIONAL ANTENNAS—STATION WOR

At this time, transmitter designers became beneficiaries of a bonanza deriving from multiplex-wire telephony, the invention of negative feedback. Conceived by H. S. Black in 1927 to reduce nonlinear effects in multichannel repeaters, the negative-feedback principle came to the radio broadcast field at an auspicious time. (See Chapter 4.) E. B. Ferrell, of the Radio and Vacuum Tube Research Department at the Deal laboratory, first proposed inclusion of a radio transmitter in a feedback loop, a rectified sample of the final RF output being fed back in opposition to the audio input [Fig. 3-3].[6]

Ferrell showed that by including the phase shift due to the limited band of the RF interstages and output stage in the same way as the phase of the audio stages, a stable design could be achieved. The negative-feedback principle was first applied to broadcast transmission in the new 50-kw transmitter built for Station WOR at Carteret, New Jersey.[7] The new equipment, placed in service early in 1935, though basically similar to other 50-kw transmitters built earlier, ushered in a new era in high fidelity and embodied other innovations. The radio circuit had crystal-controlled carrier generation, a low-level modulation stage, a push-pull parallel power-output stage, and a rectifier for feedback to the audio input [Fig. 3-4].

Located about 16 miles southwest of New York City, the new Station WOR was planned to give strong coverage along the northeast-southwest Philadelphia-New York axis and relatively low signal strength toward the more sparsely populated areas of northwestern New Jersey and toward the ocean. Thus, an hourglass-shaped radiation pattern was desired.[8] The antenna array chosen consisted of three radiators, each a quarter wavelength high (tower height was limited by proximity to established airways) and approximately a quarter wavelength apart, on a line at right angles to the major axis of the desired

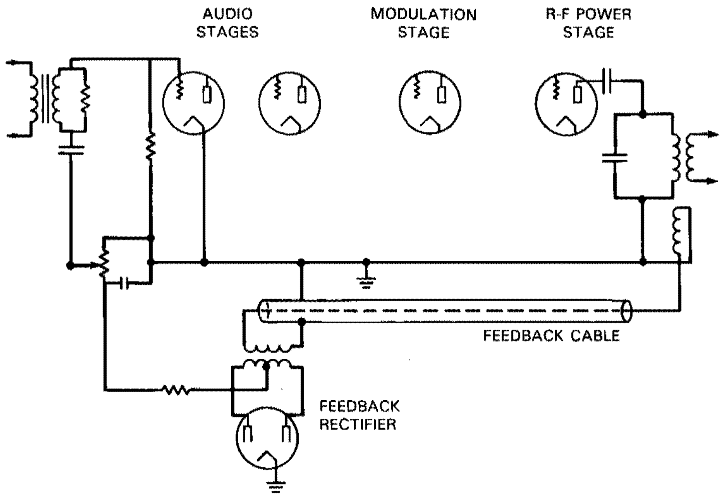


Fig. 3-3. Envelope feedback, as applied to a radio broadcast transmitter.

pattern [Fig. 3-5]. Each radiator was fed by coaxial transmission lines and driven in phase with approximately equal currents [Fig. 3-6].

Station WOR also embodied pioneering ideas in antenna line-branching transformers, ground system design, phase and amplitude adjustment and phase-monitoring arrangements.[9] A method of protecting against damage to antenna coupling circuits from high-power RF follow-up surges caused by lightning hits on the towers was used for the first time in this station. The

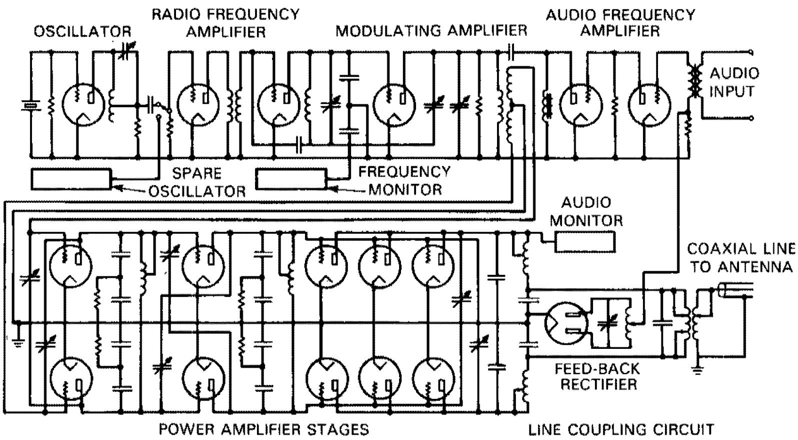


Fig. 3-4. Circuits for the 50-kw transmitter used in Station WOR in 1935. [Poppele, Cunningham, and Kispaugh, *Proc. IRE* **24** (1936): 1072.]

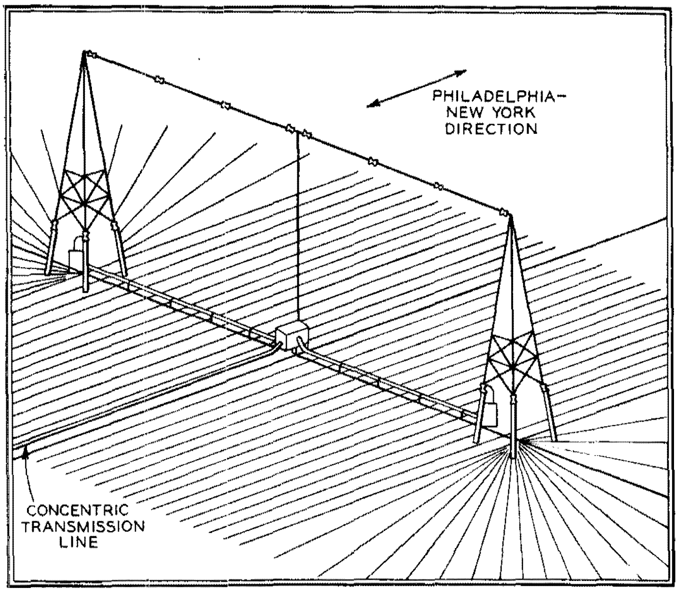


Fig. 3-5. Directional antenna and ground-system arrangement for Station WOR's 50-kw broadcasting station (1935). The central element or 'tower' was a copper cable suspended from a steel messenger supported by the two other radiators. The ground plane consisted of a buried grid of No. 10 copper wires.

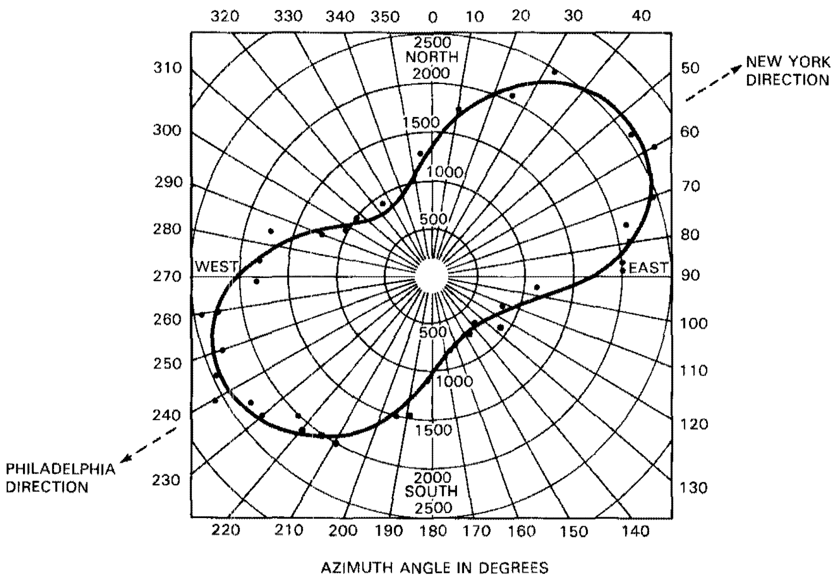


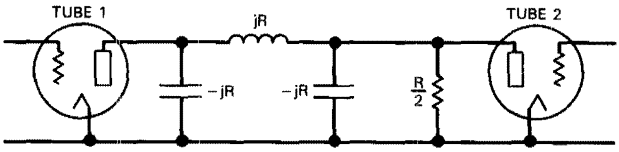
Fig. 3-6. Observed distribution of radiation from the WOR directional array. [Poppele, Cunningham, and Kispaugh, *Proc. IRE* **24** (1936): 1081.]

method, devised during the long-wave transatlantic development, involved a continuous monitoring of the impedance seen from the plates of the power tubes.[10] The carrier supply was momentarily interrupted, halting the RF drive, to allow clearing of the arc any time an impedance change due to flashover at any of the towers was detected. The method proved invaluable, especially with multielement arrays, and was later incorporated in all high-power transmitters. Station WOR was also one of the early transmitters to employ mercury-vapor tubes (developed by General Electric) in the high-voltage rectifier in place of the water-cooled, high-vacuum rectifiers previously used. The relatively low drop—about 15 v instead of 1000 v or more—contributed appreciably to operating economy.

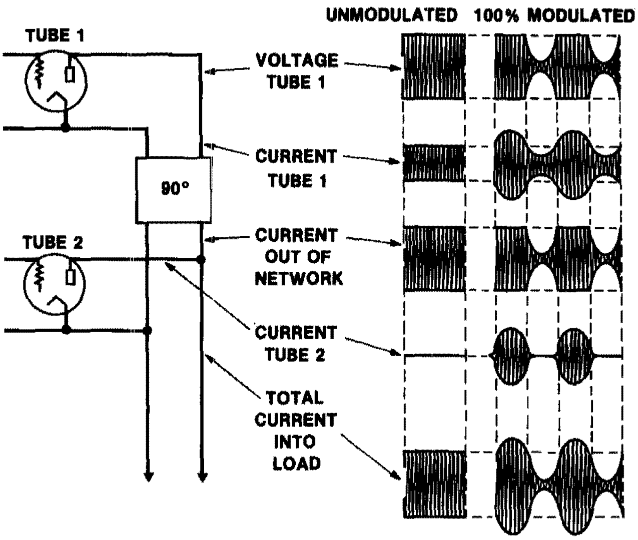
IV. SUPER POWER AND THE QUEST FOR EFFICIENCY

As these improvements were evolving in the early 1930s, interest continued in the use of higher power. In May 1934, Station WLW in Cincinnati started operation with an authorized transmitter power of 500 kw. Such power levels placed an even higher premium on more efficient ways of operating the final amplifier stage. The 500-kw transmitter at Station WLW, built by RCA, used a highly efficient Class C (biased below cutoff in the absence of RF grid drive) RF output stage. This was plate modulated by a Class B (biased at cutoff) audio amplifier, which drew little power except in the presence of a strong audio signal. With this high-level Class B system, particularly at 500 kw, an extremely large audio transformer and choke were necessary so that the balanced Class B stage could work into the single-sided load presented by the RF stage. To a generation to whom high power in an audio amplifier means 50 or 100 w, and where device dimensions are often measured in micrometers, the numbers are astounding. The high-level audio transformers were 11 feet high, weighed 19 tons each, and were immersed in oil. Two were required! The choke, needed to bypass the DC plate current from the audio transformers, weighed 12 tons. The high capacity to ground of such massive components made both RF and broadband audio designs difficult. The initial expense and the cost of the power consumed were very high. These difficulties and other problems led Bell Laboratories engineers to look for other ways to achieve high power and high efficiency. A most ingenious solution was discovered late in 1933 by W. H. Doherty at the Whippany radio laboratory, which had become the center for high-power work.[11,12]

The day-long efficiency of an AM transmitter is essentially that realized with little or no audio signal modulating the carrier. The trick was to make this idling efficiency very high, while still retaining the capability of delivering the much higher peak power required at the maximum audio level (100 percent modulation). Doherty's amplifier achieved high efficiency by obtaining the unmodulated carrier from one of two output stages operating into a load impedance twice as high as the optimum for a fully modulated carrier. Power above the unmodulated level—that is, on the positive swings of the audio signal—was then obtained by bringing into play on a dynamic basis the second stage, which, in the absence of audio, was biased well beyond cutoff. Had the two tubes been simply connected in parallel or push-pull, this supplementary action would not have been possible, since the contribution of power by the second tube would only have raised the apparent impedance into which the first tube was working and thus reduced its maximum possible output power by an equal amount. Doherty's stratagem was to employ a quarter-wave impedance-inverting network to reverse this effect, *reducing* the impedance seen by the first tube. While still operating close to its maximum possible plate-voltage swing, the first tube could deliver more current, and hence more power, to the load. At the instantaneous peak of modulation, the two tubes shared the load equally,



(a)



(b)

Fig. 3-7. The Doherty amplifier. (a) At the instantaneous peak of a fully modulated wave, each tube faces its optimum impedance R and, at its maximum voltage swing E, delivers its maximum power E^2/R as in a conventional linear power amplifier. At lower instantaneous outputs, the contribution of tube 2 is reduced, and at unmodulated carrier output (one-fourth of the peak power), the impedance-inverting network finds itself terminated in $R/2\ \Omega$ and thus presents to tube 1 an impedance of $2R$, which enables tube 1 to deliver the carrier power $E^2/2R$ at its full voltage E, and consequently full efficiency. [Doherty, *Proc. IRE* **24** (1936): 1173.] (b) Envelope waveforms in the Doherty amplifier at unmodulated carrier output and with a fully modulated wave.

each working into its optimum impedance [Fig. 3-7]. Since the quarter-wave impedance-inverting network shifted the phase 90 degrees, another network was inserted in the grid circuit so that the grid excitations on the two tubes would also be 90 degrees apart. In other respects, the system operated as a conventional linear power amplifier. It was applicable to single-sideband telephony, as well as to AM broadcast transmitters, and lent itself to the use of negative feedback for achievement of high fidelity.

The Doherty principle was immediately adopted for a new line of transmitters beginning with a 50-kw AM broadcast radio transmitter, the No. 407A-1, employing two of the new 100-kw tubes in the output stage [Fig. 3-8]. The power stage achieved the expected 60-percent efficiency compared to 33 percent in conventional linear transmitters. With 30 dB of negative feedback, it was possible for the first time to reduce AC filament hum in the high-power tubes to a level where separate DC generators for filament and grid-bias supply were no longer necessary. Two transmitters made by the Northern Electric Company from the Western Electric design were installed in Canada in late 1937 and early 1938. Western Electric's first production set was placed on the air a few months later at Station WHAS in Louisville.[13] Many others went into service in the next few years. Meanwhile, the principle was applied in new 5-kw and

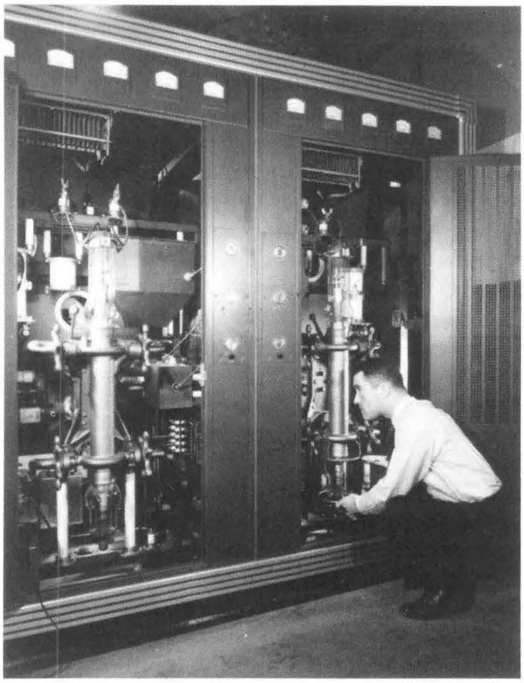


Fig. 3-8. The two-tube output stage of a Doherty amplifier for a 50-kw AM broadcast transmitter. The inventor is at the right.

1-kw transmitters, and was adopted by several other transmitter manufacturers under Bell System license. For the 1-kw transmitter, of which large numbers were sold, a modification of the high-efficiency scheme was devised, in which the two tubes were subjected to a constant RF drive and their grid biases varied at the audio frequency, reducing the number of stages in the feedback loop.[14]

During this time, interest in super power (500 kw or more) continued. Much of the preliminary design work on a 500-kw Doherty amplifier was carried out at the Whippany laboratory, including testing of early models of a new tube having a peak power capacity of over 250 kw. This was the most powerful sealed-off vacuum tube that had ever been made, with a three-phase tungsten filament consuming 17 kw. Several of these tubes, coded the No. 320A vacuum tube, were successfully made at the Bell Laboratories West Street laboratory and eight were installed in a 500-kw transmitter using the Doherty circuit under license. This transmitter, built in Mexico by the Continental Corporation, operated for several years with no tube failures. The market for the No. 320A vacuum tube disappeared, however, when the FCC decided to limit broadcast power in the United States to 50 kw, requiring Station WLW in Cincinnati to cut back to that level and ending the quest of the clear-channel stations to augment their coverage by this means.

V. FREQUENCY-MODULATION RADIO

The late 1930s and early 1940s saw another radio-broadcasting development that was to prove of major importance—the advent of wideband FM. Interest in FM was almost as old as radiotelephony. J. R. Carson of AT&T had shown mathematically in 1922 that modulation of the frequency of a carrier wave, even by very small amounts, could not *conserve* frequency space, an important objective in his time. What was not then foreseen was the possibility, or even the desirability, of improving transmission through the *expenditure* of frequency space. This was the great and imaginative contribution of the inventor E. H. Armstrong. Armstrong deduced and demonstrated that by employing wide frequency swings, easily accommodated in the relatively unused very-high-frequency (VHF) band above 30 MHz, interference could be suppressed to a high degree.[15] The key to Armstrong's discovery was the fact that, as long as a disturbing signal is weaker than the desired signal, the phase of their resultant can scarcely be disturbed by more than about one radian, while the phase shift due to modulation can be many radians. Thus, if an audio signal of 1 kHz is allowed to swing an RF carrier ±75 kHz—the swing that was finally standardized upon as a maximum—then the deviation ratio and consequent suppression of phase disturbances are on the order of 75 to 1. Amplitude disturbance can be removed by an instantaneous limiter and the original modulating signal recovered by a discriminator. A discriminator is a simple network in which the output signal amplitude varies linearly with the input frequency within the deviation range, thus furnishing, when rectified, the original amplitude and frequency variations of the audio input.

Armstrong's revelations, supported by many tests by himself and others, were something of a sensation in the broadcast field, offering a method of greatly improving the quality of local broadcast coverage with relatively low power. There were many advantages. The short wavelengths in the VHF band permitted high antenna directivity in the vertical plane with small structures and with consequent enhancement of effective radiated power. The absence of sky-wave transmission (reflections from the ionosphere) made it possible to duplicate channel assignments at relatively short distances. Perhaps most exciting to a generation just becoming aware of the possibilities of high fidelity, was the prospect of low-noise, wider-audio-band programs, since the RF spectrum occupancy required was determined primarily by the deviation of the radio-frequency carrier rather than by the highest audio frequency.

Bell Laboratories engineers quickly recognized the potential advantages of wide-swing FM for broadcasting, and Western Electric promptly negotiated with Armstrong for rights to build FM equipment for the 40-MHz range. Experience with shortwave transmitters for transoceanic service had already provided some of the necessary tube and circuit background. (See Chapter 2.) FM, however, presented a new problem in stabilization of the mean carrier frequency, since it was not possible to vary the frequency of a crystal oscillator over the range required for modulation. Armstrong had handled this problem by phase modulating the output of an oscillator, then heterodyning down to low frequency (shifting the mean frequency, but not the frequency swing) to increase the phase deviations. Multiplying up again (that is, multiplying both the mean frequency *and* the deviation) increased the consequent total phase multiplication thousands of times. Another approach involved direct modulation of a high-frequency oscillator and frequency determination and control by a precise, temperature-controlled discriminator with an auxiliary crystal oscillator as a reference.

The Bell Laboratories approach was to separate completely the functions of modulation and of carrier frequency control. A circuit to achieve this was invented by J. F. Morrison in the late 1930s [Fig. 3-9].[16] In this scheme, a vacuum-tube oscillator operating at one-eighth of the final carrier frequency was frequency modulated by a reactance-control tube. (In a reactance-control tube, the capacitive reactance of the tube varied in proportion to the magnitude of an input signal.) The oscillator frequency was made to follow the signal am-

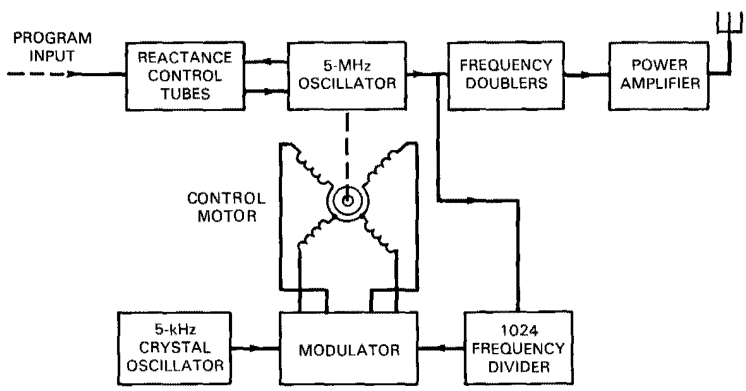


Fig. 3-9. System block diagram for synchronized FM frequency control.

plitude by making this reactance part of the tuned circuit. The circuits were carefully designed to give extremely low distortion and wide frequency response. A sample of the oscillator output was then divided down by a factor of 2^{10} (= 1024) in a series of simple frequency dividers.[17] This large division transformed the character of the wave spectrum to one having, in effect, almost a single frequency in the 5-kHz range with only small phase excursions. This frequency was then compared with that of a crystal oscillator, and their difference, if any, was fed to a small two-phase synchronous motor actuating a variable capacitor in the primary frequency-modulated oscillator circuit. The small phase excursions were suppressed by the inertia of the mechanical system and a deadbeat synchronism achieved, in contrast to the inevitable small frequency error characteristic of other control systems. Morrison's system, named Synchronized Frequency Modulation by Western Electric, was incorporated in a 1-kw FM transmitter, the No. 503A-1, 15 of which went into broadcast service before World War II. All 15 were destined for conversion, by means of an added frequency-doubler stage, to the new FM band (88 to 108 MHz) after the war. Before full-scale hostilities broke out, diverting the energies of most broadcast engineers to urgent military work, a 10-kw amplifier was designed for addition to this set.

VI. THE POSTWAR YEARS

The wavelengths of the FM band, about 3 m long, made it possible to design directional antenna arrays of relatively small dimensions. What was desired was an antenna with a uniform pattern in azimuth, but high concentration in the vertical plane. However, even the moderate size of a multielement FM band array made these structures inconveniently large for laboratory work. The utility of higher-frequency, physically scaled-down models had been demonstrated in the earlier testing of the rhombic antenna for shortwave overseas transmission. (See Chapter 2, Section 2.1.) By taking advantage of skills acquired during the war in a new frequency domain at thousands of megahertz (the microwave range), P. H. Smith constructed and measured models of an antenna he invented for the new FM band at 88 to 108 MHz on a scale of one-tenth.[18,19] Earlier forms of radiating systems for giving azimuthally uniform patterns of horizontally polarized waves had been based either on the "turnstile" or rotating-field principle involving quadrature-fed dipoles, or on attempts to establish rings of uniform current as elements in a vertical array. These approaches were electrically and mechanically complex, a complexity which Smith eliminated with the cloverleaf antenna, constructed of galvanized steel throughout and employing no insulators [Fig. 3-10]. Designated No. 54A, the antenna used a one-foot-square lattice tower as the outer conductor of a skeleton coaxial transmission line, the inner conductor being a three-inch pipe. The latter was rigidly supported at half-wavelength intervals by clusters of radiators in the form of a four-leaf clover. Each leaf consisted of a two-inch-diameter curved pipe securely attached at one end to the central pipe, and at the other end to one of the corners of the tower. These pipes behaved somewhat like quarter-waves lines terminated only by the radiation resistance of the outermost portion of the leaf. Thus, the structure was extremely rugged and lent itself to easy erection.

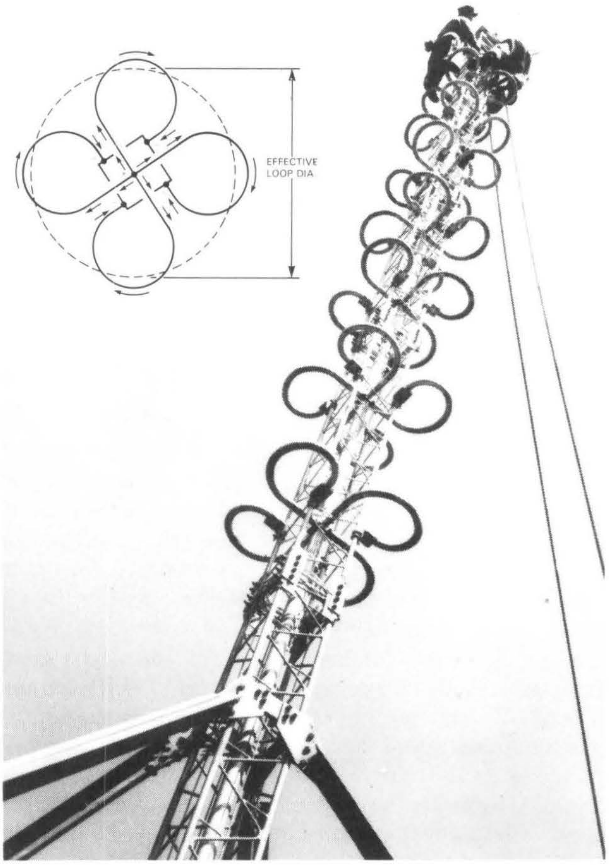


Fig. 3-10. The No. 54A antenna for FM broadcasting, shown in the process of installation. Insert indicates current flow in the four 'leaves' of each radiating element or cluster. [Smith, *Proc. IRE* **35** (1947): 1557.]

In accordance with transmission-line theory, the half-wavelength spacing of the clusters assured equal potentials at all the connecting points. The phase was also identical, except for the phase reversals, which were taken care of by reversing the leaf direction at successive clusters. The currents flowing in the outermost or radiating portions of the clusters were then all identical, irrespective of mutual impedances between clusters, since the current at the terminating end of a quarter-wave line is a function of the driving-end potential only. As a result, the power gain was a linear function of the number of elements. In azimuth, the field strength was uniform within ± 0.2 dB [Fig. 3-11]. The ruggedness and simplicity of design and the freedom from insulators resulted in a structure that would withstand high wind velocities and heavy icing loads, with full protection against lightning surges. Sleet melting was easily accom-

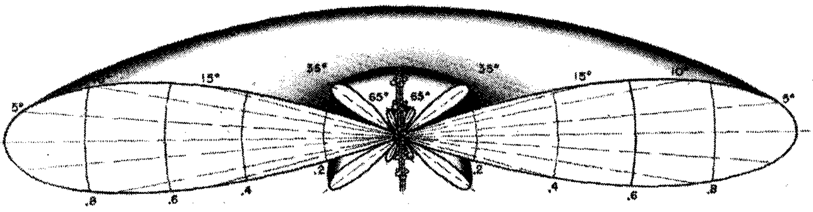


Fig. 3-11. Vertical radiation pattern of a five-element cloverleaf antenna, as measured on a precise one-tenth scale model. The five-element array provided three times the gain of a single element.

plished with heating cables inside the tubular radiating elements, requiring no filters. The No. 54A antenna, with deliveries beginning in mid-1946, proved a popular choice in the rapidly expanding market for FM broadcast equipment.

Smith had originally been a member of the Radio and Vacuum Tube Research Department, where he took part in the early development of transmission lines and directional antennas for the shortwave overseas circuits. In 1931, he had devised a transmission-line calculator that gave the input impedance of a transmission line of any length for any terminating impedance. In a more advanced form, this became universally used and known as the Smith Chart [Fig. 3-12].[20]

A high-priority project in the post–World War II program was a new 10-kw transmitter for the new FM band at 88 to 108 MHz. Coded the No. 506A-2 transmitter, it made its appearance in mid-1946. This transmitter used forced air instead of water cooling of the anode of its single output tube, a thoriated-filament triode operating at 8500 v. In a conventional circuit, the large capacitance to ground of the anode fin structure of a high-power, air-cooled tube entailed high circulating currents with high losses and some bandwidth limitation. These effects were reduced in the new circuit in which the anode was at RF ground, with the output taken from cathode terminals.[21]

In conventional circuitry, if the cathode of a vacuum-tube amplifier was made the output terminal, the stage was known as a cathode follower and had no voltage gain because of the feedback action. In the new circuit, this feedback action was circumvented by running the grid drive up through the inside of the tubing of the cathode output tuning inductance, which in the range from 88 to 108 MHz was the center conductor of a short-circuited coaxial line [Fig. 3-13]. Thus the circuit behaved as a normal triode amplifier with full gain. As an additional advantage, since the anode was at RF ground, the inductance to neutralize the grid-plate capacitance also consisted simply of a grounded coaxial stub. The No. 506A-2 transmitter had an audio-frequency response flat to less than 0.5 percent and an FM noise level 65 dB below the carrier.

As part of the FM transmitter work, Bell Laboratories engineers developed a method for accurate power and impedance monitoring for use by FM broadcast station operating personnel. An assembly of two directional couplers sampled the current and voltage in the antenna-feed coaxial line and provided a direct reading of the power flow in the forward direction. A similar pair, poled to measure the reflected power, gave a simple indication of the impedance match to the antenna. The power and impedance monitor filled a serious need and

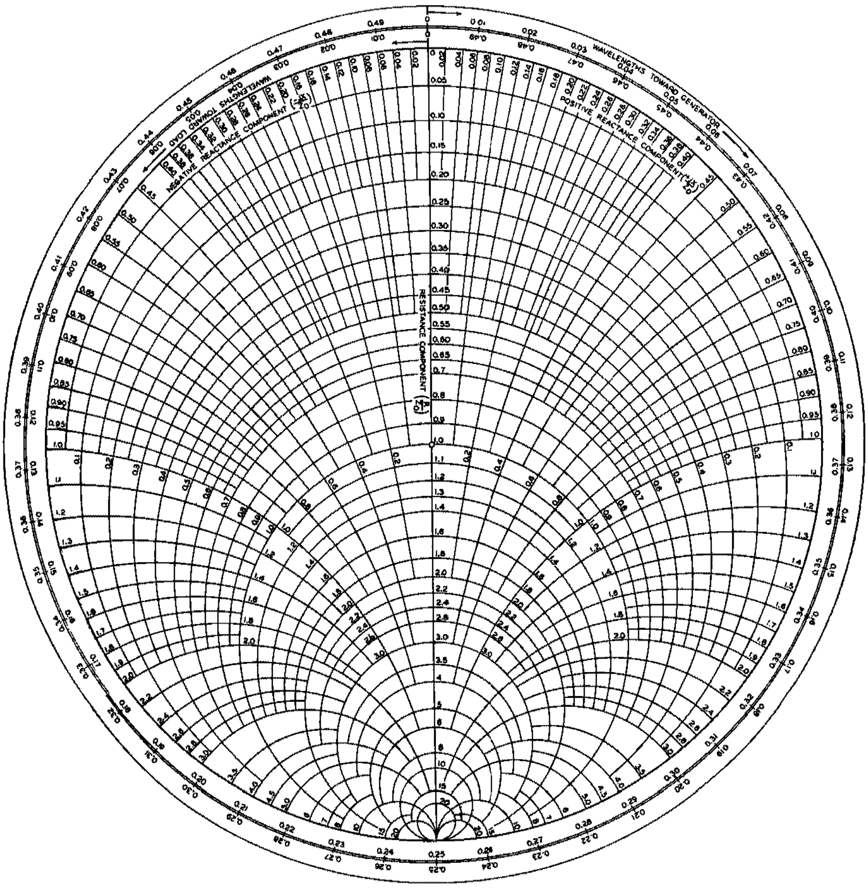


Fig. 3-12. The Smith chart transmission-line calculator. By means of the chart, the input impedance and standing-wave ratio may be determined for a transmission line of any length terminated in any impedance. [Smith, *Electronics* **12** (1939): 31.]

was made available to the field generally, as well as in association with the 10-kw transmitter.[22,23]

The RF grounded-anode, air-cooled approach was also used for a new AM transmitter in the 5- to 10-kw range. This transmitter also had built-in means to cope with a long-standing problem. The impedance of a broadcast antenna, especially an array, often varied widely over the transmission band, adversely affecting the performance of the transmitter. The new "5-10" transmitter included a variable phase shifter or "line stretcher" in the output circuit to permit reorientation of the impedance-frequency characteristic of any sharply tuned load.[24] From a Smith chart plot of the measured impedance into the coaxial antenna feed, the station could be furnished with the data for setting the optimum phase shift, even under power [Fig. 3-14]. This method of coping with

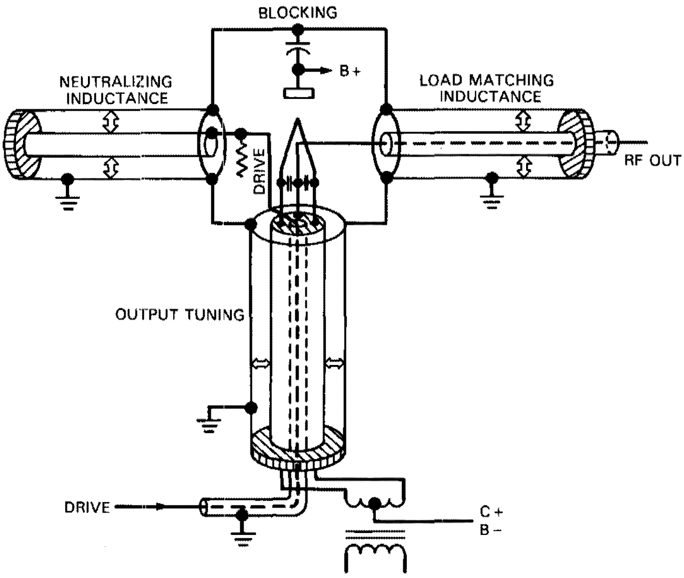


Fig. 3-13. The grounded-plate amplifier, invented for the original 40-MHz FM band. It became especially simple for the band from 88 to 108 MHz, where all tuning was done with coaxial stubs. The grid drive was given a 'free ride' up to the high-RF output potential of the cathode by running it up through the center tubing of the output tuner.

the antenna impedance problem, the last major Bell contribution to the technology of broadcast transmission, was to serve AM transmitter designers and antenna system designers for many years.

When the years 1947 and 1948 brought a buildup instead of a decline in military work, it became necessary to re-examine the Bell System's position as a supplier of commercial products, of which broadcast transmitters and broadcast studio equipment were the prime examples. Urgent developments, such as microwave radio and the deep-sea telephone cable, were also in need of experienced engineers. It was determined that Bell Laboratories should henceforth concentrate its research and development resources on Bell System work and government-sponsored projects whose nature pointed to Bell Laboratories and Western Electric as especially qualified participants. A factor in the decision was the expected replacement of the Bell System's high-power HF radio by multichannel, deep-sea, repeatered cables as the primary medium of transoceanic telephony.

A small group of engineers was selected to carry on with Western Electric's broadcast installation and servicing responsibilities as members of Western Electric's Field Engineering Division. The prototype models of the new "5-10" AM transmitters, with parts already purchased for a quantity of these, were sold to the Continental Corporation, which was to become the world's leading manufacturer of high-power radio transmitters. Western Electric's high-power-

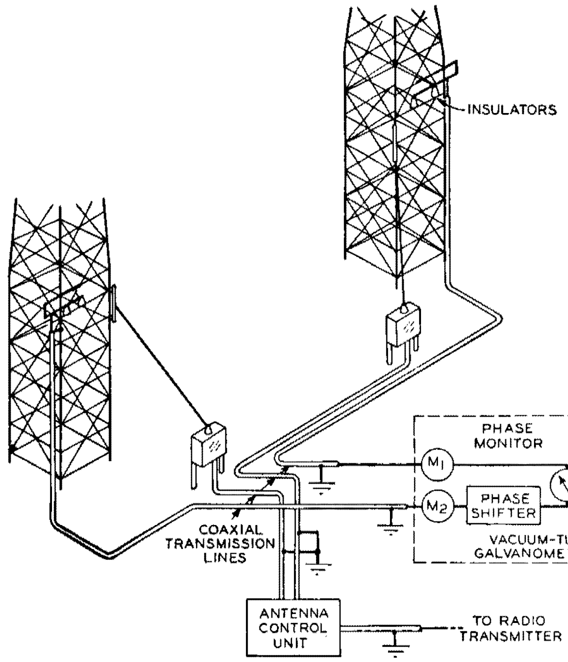


Fig. 3-14. Block schematic of method of directly measuring antenna currents by means of No. 2A phase monitor.

tube designs were turned over to Machlett Laboratories; microphone, loudspeakers, and disc-reproducing equipment to the Altec Co.; and studio control equipment to Daven. Thus ended an era of more than 25 years of the Bell System's vigorous promotion of high-power broadcast radio.

REFERENCES

1. A. A. Oswald and J. C. Schelleng, "Power Amplifiers in Trans-Atlantic Radio Telephony," *Proc. IRE* **13** (June 1925), pp. 313-361.
2. R. Bown, C. R. Englund, and H. T. Friis, "Radio Transmission Measurements," *Proc. IRE* **11** (April 1923), pp. 115-152.
3. E. L. Nelson, "Radio Broadcasting Transmitters and Related Transmission Phenomena," *Proc. IRE* **17** (November 1929), pp. 1949-1968.
4. O. M. Hovgaard, "A New Oscillator for Broadcast Frequencies," *Bell Lab. Rec.* **10** (December 1931), pp. 106-110.
5. R. E. Coram, "A Frequency Monitoring Unit for Broadcast Stations," *Bell Lab. Rec.* **11** (December 1932), pp. 113-116.
6. E. B. Ferrell, U.S. Patent No. 2,159,020; filed June 9, 1934; issued May 23, 1939.
7. J. R. Poppele, F. W. Cunningham, and A. W. Kishpaugh, "Design and Equipment of a Fifty-Kilowatt Broadcast Station for WOR," *Proc. IRE* **24** (August 1936), pp. 1063-1081.
8. J. F. Morrison, "Controlled Radiation for Broadcasting," *Bell Lab. Rec.* **13** (April 1935), pp. 232-237.

9. J. F. Morrison, "Simple Method for Observing Current Amplitude and Phase Relations in Antenna Arrays," *Proc. IRE* **25** (October 1937), pp. 1310-1326.
10. W. H. Doherty, U.S. Patent No. 2,066,522; filed December 19, 1934; issued June 5, 1937.
11. W. H. Doherty, "A New High-Efficiency Power Amplifier for Modulated Waves," *Proc. IRE* **24** (September 1936), pp. 1163-1182.
12. W. H. Doherty, U.S. Patent No. 2,210,028; filed April 1, 1936; issued August 6, 1940.
13. W. H. Doherty and O. W. Towner, "A 50-Kilowatt Broadcast Station Utilizing the Doherty Amplifier and Designed for Expansion to 500 Kilowatts," *Proc. IRE* **27** (September 1939), pp. 531-534.
14. F. E. Terman and J. R. Woodyard, "A High-Efficiency Grid-Modulated Amplifier," *Proc. IRE* **26** (August 1938), pp. 929-945.
15. E. H. Armstrong, "A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation," *Proc. IRE* **24** (May 1936), pp. 689-740.
16. J. F. Morrison, "New Broadcast-Transmitter Circuit Design for Frequency Modulation," *Proc. IRE* **28** (October 1940), pp. 444-449.
17. R. L. Miller, "Fractional-Frequency Generators Utilizing Regenerative Modulation," *Proc. IRE* **27** (July 1939), pp. 446-457.
18. P. H. Smith, "Cloverleaf' Antenna for FM Broadcasting," *Proc. IRE* **35** (December 1947), pp. 1556-1563.
19. P. H. Smith, U.S. Patent No. 2,521,550; filed February 28, 1946; issued September 5, 1950.
20. P. H. Smith, "Transmission Line Calculator," *Electronics* **12** (January 1939), pp. 29-31.
21. W. H. Doherty, U.S. Patent No. 2,314,132; filed June 21, 1941; issued March 16, 1943.
22. J. F. Morrison and E. L. Younker, "A Method of Determining and Monitoring Power and Impedance at High Frequencies," *Proc. IRE* **36**, pp. 212-216.
23. W. H. Doherty, J. F. Morrison, and E. L. Younker, U.S. Patent No. 2,575,799; filed May 27, 1948; issued November 20, 1951.
24. W. H. Doherty, "Operation of AM Broadcast Transmitters into Sharply Tuned Antenna Systems," *Proc. IRE* **37** (July 1949), pp. 729-734.

Chapter 4

Negative Feedback

I. THE INVENTION

"Although many of Harold Black's inventions have made great impact, that of the negative feedback amplifier is indeed the most outstanding. It easily ranks coordinate with [L.] de Forest's invention of the audion as one of the two inventions of broadest scope and significance in electronics and communications of the past 50 years. . . . Without the stable, distortionless amplification achieved through Black's invention, modern multichannel transcontinental and transoceanic communications systems would not be possible." The quotation is from a 1957 speech by M. J. Kelly, President of Bell Laboratories, on the occasion of the presentation of the American Institute of Electrical Engineers (AIEE) Lamme Gold Medal to H. S. Black [Fig. 4-1].

This epochal invention was not "a solution looking for a problem." It was the outcome of years of effort by an inventive engineer, first to understand the problems of long-distance transmission and then to find ways to solve them. In 1921, Black, newly employed in the Engineering Department of Western Electric, was looking beyond the three-channel carrier telephone systems then being designed for use on open-wire lines. He foresaw future systems carrying many more channels over greater distances with many more repeaters in tandem.

One characteristic of vacuum-tube amplifiers was looming as a major problem. The plot of output plate current versus input grid voltage was not a straight line, but was curved—that is, the tubes were inherently nonlinear [Fig. 4-2]. As a result, the output waveform was not an exact amplified replica of the input. With a single frequency input, it contained harmonics of the input frequency, as well as the amplified original. With a speech signal input, there would be harmonics and many other intermodulation products of all its many components. With a single voice signal, if the signal was not too large and was confined to the more linear portion of the curve, these unwanted products were not too troublesome. When several voice signals were applied simultaneously in a common amplifier, the nonlinearity produced a great many products that fell into all the channels to varying degrees. Some of these were noiselike, but some appeared as transfers of completely intelligible voice signals from one channel to another. All products, noiselike or intelligible, had to be held to very low levels to meet the standards of high-quality transmission.

Clearly, the more voice channels that were applied, the more products pro-

Fig. 4-1. H.S. Black holding one of the condensers from a negative-feedback amplifier being tested in 1930.

duced. Also, as more channels were carried, higher frequencies were needed with higher line losses and hence more closely spaced repeaters [Table 4-1].

The implication is clear. If a system can tolerate only a certain total accu-

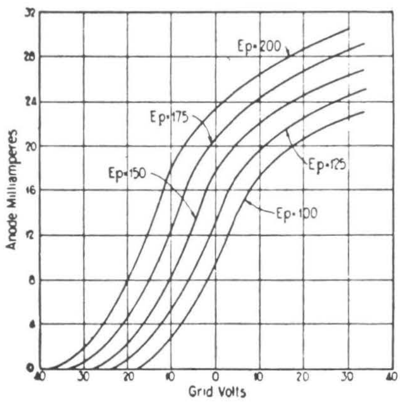


Fig. 4-2. Vacuum-tube characteristic, 1923.

TABLE 4-1: Toll Repeater Development				
System	**Date**	**Channels per Pair**	**Loss in dB (3000 Miles)**	**Repeaters (3000 Miles)**
1st Transcontinental	1914	1	60	3–6
2nd Transcontinental	1923	1–4	150–400	6–20
Open-wire Carrier	1938	16	1,000	40
Cable Carrier	1936	12	12,000	200
First Coaxial	1941	480	30,000	600

[Bode, *Proc. Symp. Active Networks and Feedback Systems* **10** (1960): 5.]

mulation of undesired modulation products, then each repeater can contribute only its pro rata share. As system channel capacity and loss increase, the repeaters must become more and more linear.

In 1921, Black wrote a memorandum evaluating the effect of both more channels and more amplifiers. Although only three-channel carrier systems had been used by that date, he carried his analysis to systems with thousands of channels and transcontinental lengths. Intrigued by the problem of achieving ultralinearity, he asked to be allowed to work on improved amplifiers. His supervision was understandably somewhat bemused by the scale of his thinking, given the difficulties they were having making a system with only 3 channels and 10 or 12 amplifiers work, but he was granted permission to work on improved amplifiers, providing it did not interfere with his other assigned work.

The problem was even tougher than Black first thought. In his analysis, he had assumed that the distortion products from successive amplifiers would add in random phase. If this were so, he would get a root-mean-square addition of the distortion products. Thus, if there were 1000 amplifiers in tandem with random phase, each could contribute 1/30 the total permitted distortion. His associate, R. V. L. Hartley, quickly pointed out that a very important class of products, those arising from third-order distortion of the amplifier's transfer characteristic, would add in phase. Under this condition, each repeater could be permitted to contribute only 1/1000 the allowable total. If, under his original assumption, he was seeking a 30-dB reduction in the third-order distortion, he would now need a 60-dB improvement.

Of course, designers at the time were not thinking in terms of systems with 1000 amplifiers. The amplifiers available were barely adequate for the systems in use, requiring only 10 or 20 amplifiers, and attempts to improve them by straightforward means were not proving fruitful. By the late 1920s, the proposal for a multichannel system on paired cable, with its high loss and consequent large number of repeaters, confronted them with the need for an improvement by a factor of 100 (20 dB). This was far beyond any prospect in view with the technologies of 1925.

1.1 Feedforward and Feedback

Black later described events in a private letter to A. C. Dickieson, his associate in the 1920s:

. . . The job as I well knew required an amplifier vastly superior to any then existing. Many other researchers were aware of this need. Like the others my first approach was to try to linearize the tube characteristic. Dr. Kelly cooperated with me closely in an endeavor to provide an adequately linear tube characteristic but all to no avail.

In the course of this work I attended a lecture by C. P. Steinmetz at an AIEE meeting and was impressed by the Steinmetz way of getting down to the fundamentals of a problem. As a result I restated my assignment as being that of removing distortion products from the amplifier output. I immediately observed that by reducing the output

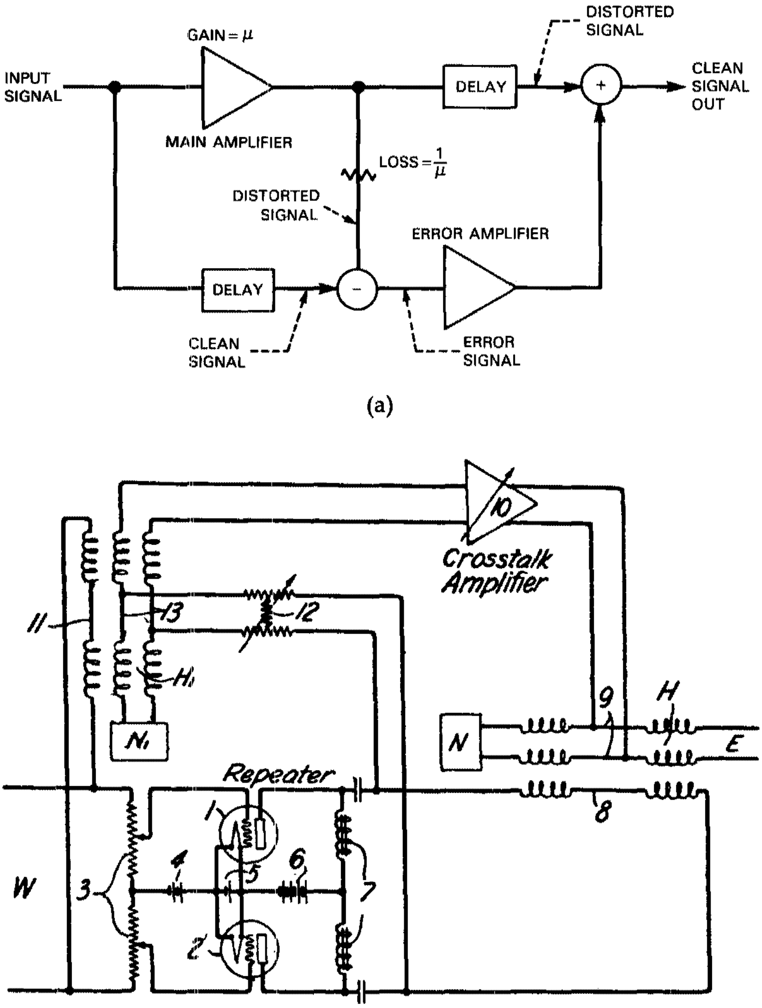


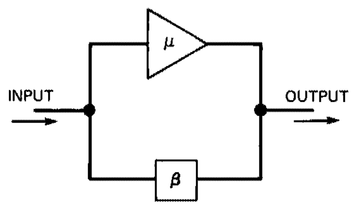
Fig. 4-3. The feedforward amplifier. (a) Schematic of the circuit. (b) Circuit as disclosed in the H.S. Black patent application.

to the same amplitude as the input and subtracting one from the other, the distortion products only would remain which could then be amplified in a separate amplifier and used to cancel out the distortion products in the original amplifier output. . . Thus, the Feedforward Amplifier came into being.[1]

The feedforward amplifier was reduced to laboratory practice in March 1923, and worked as expected [Fig. 4-3]. However, it called for precise balances and subtractions that were hard to achieve and even harder to maintain, especially over a wide band of frequencies. The feedforward amplifier reenters this history in a most extraordinary way in the development of single-sideband microwave radio in the 1970s, but in the 1920s, it was only a step on the way, not the solution. (See Chapter 11, Section IV.)

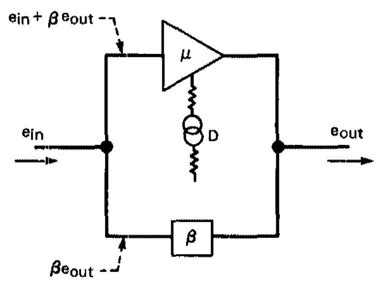
Black continued to wrestle with the problem for several years with little success. Finally, on August 2, 1927, while on the Lackawanna Ferry crossing the Hudson River on the way to work, he had a flash of insight, famous in the annals of Bell Laboratories. He suddenly realized that by employing *"negative feedback,"* that is, by inserting part of the output signal into the input in reversed phase, virtually any desired reduction in distortion could be obtained by a sacrifice in amplification. He sketched the diagram and scribbled the basic equations on a page of the *New York Times* he was carrying. He had it witnessed upon arrival at his office; negative feedback had been invented [Fig. 4-4].

The circuit invented was deceptively simple, with some extremely valuable properties. All that Black had time to sketch on the ferryboat was the circuit and basic equations. An amplifier with a forward gain of μ has the output fed back to the input through a network with the loss β. The gain is shown to be reduced by the factor $1/(1 - \mu\beta)$. By itself, the equation was not very illuminating, but in subsequent analysis, Black showed it was the key to the whole matter [Fig. 4-5]. The gain with feedback becomes $\mu/(1 - \mu\beta)$. This gain reduction is the price paid for two extremely important advantages. If $\mu\beta$ is large compared with 1 (and for 40-dB feedback, $\mu\beta = 100$), the amplification factor reduces to $1/-\beta$. That is, the net gain is dependent only on the β network. If this is passive and stable, as it usually is, the variation in the active amplifier gain caused by power-supply voltage variations or by aging in vacuum tubes is virtually eliminated. Also, the overall gain-versus-frequency characteristic can be shaped by the designer in the β network by the techniques of passive



$$\frac{\text{OUTPUT}}{\text{INPUT}} = A_F = \frac{\mu}{1-\mu\beta} = \frac{1}{-\beta}\left[1 - \frac{1}{1-\mu\beta}\right]$$

Fig. 4-4. Negative feedback—the Lackawanna ferryboat invention.



(1) $e_{out} = \mu(e_{in} + \beta e_{out}) + D$

(2) $e_{out}(1 - \mu\beta) = \mu e_{in} + D$

(3) $e_{out} = e_{in} \times \frac{\mu}{1 - \mu\beta} + \frac{D}{1 - \mu\beta}$

where: e_{in} = input signal
e_{out} = output signal
μ = voltage gain of amplifier
β = voltage loss of feedback network
D = distortion produced in output stage.
Output signal adjusted to the same amplitude with and without feedback.

Fig. 4-5. Feedback relations and the reduction of distortion.

network design to match, for example, the loss characteristics of the transmission line. The second term in Equation (3) shows that, at a given output signal level, the distortion introduced by the amplifying element is also reduced by the factor $1/(1 - \mu\beta)$. This was the key to ultralinear amplification, the prize sought by Black.*

The soundness of his insight was quickly demonstrated, and in the next few months he and his coworkers were able to demonstrate reductions in distortion on the order of 50 dB (100,000 to 1). The big advantage of feedback over feedforward was that it was self-correcting. Any change in the μ amplifier was automatically corrected by a corresponding change in the feedback signal. Black recognized that the invention had far-reaching implications, extending much beyond telephone amplifiers. An elaborate patent application, consisting of 52 pages with 126 claims and hundreds of figures, was drawn up and submitted in 1928. There were long, drawn-out dealings with the United States Patent Office because the principles seemed diametrically opposed to much of the accepted theory of the time. Black persisted, however, refusing to back down on a single claim, and the patent was granted in December 1937.[2]

---

* Strictly speaking, nonlinear circuits cannot be analyzed precisely by the dodge of introducing "distortion generators," such as *D*, and carrying out an otherwise linear analysis. Feeding back the distortion products, along with the amplified input signal, generates additional distortion products in the nonlinear elements. The feedback is slightly less effective than the linear analysis indicates.

II. REALIZATION

That was the inspiration; then came the perspiration. The path from the insight and initial demonstrations to the realization of practical amplifiers was a long and arduous one. For one thing, the conclusions drawn from Black's discoveries assume a stable system, that is, one that would not oscillate. But experience showed that when the output of an amplifier was connected back to the input, the system usually broke into violent oscillation. The conditions controlling stability, or nonoscillation, were of primary importance and were not fully understood in 1927. Designers did understand that both μ and β are vector quantities, that is, they have a phase value as well as amplitude. What they wanted was a large negative $\mu\beta$ over the useful band, that is, feedback in such phase that $1 - \mu\beta$ would be a large positive quantity. As long as this was true, the circuit was useful and stable. If an amplifier with 30 dB of net gain was needed, with distortion 20 dB better than a nonfeedback circuit, it was necessary to design a μ circuit with 50 dB of gain and a β circuit with 30 dB of loss. Under these conditions, the $\mu\beta$ loop gain (the amount of feedback) was 20 dB. But a large gain around the $\mu\beta$ loop could not be maintained over an infinite band of frequencies. Inevitably, because of the physical limitations of the components, the gain dropped, and the phase shifted at higher frequencies. In practical multistage amplifiers, the phase shift ultimately reached and exceeded 180 degrees. That is, the so-called negative feedback became positive. If at such frequency the magnitude of $\mu\beta = +1$, then $1 - \mu\beta = 0$ and the gain became infinite—obviously an unstable condition.

Designers of oscillators knew from experience that the transmission around the loop did not have to be exactly unity ($\mu\beta = +1$) for sustained oscillations. Ordinarily, if at the frequency of a 180-degree phase shift there was loop gain ($\mu\beta > 1$), oscillations would start and grow until the system started to overload, that is, until the tube gains were reduced. At some amplitude, this would reduce the loop gain to unity, producing a stable and sustained oscillation. Black postulated that, if the gain around the loop could be reduced to a loss ($\mu\beta < 1$) before the phase shift reached 180 degrees, the system would not oscillate.

It was not difficult to meet this condition in a simple single-stage amplifier, but much more gain was required than was available in a single stage, since a large amount of gain had to be sacrificed to improve the linearity by feedback. When circuits were designed to reduce the large gain above the useful band, large phase shifts resulted, with instability as a frequent consequence. Black and his early coworkers did not have the knowledge of network theory relating the changes in phase and gain of the $\mu\beta$ loop with frequency. That was to await H. W. Bode's work in the last half of the 1930s. It was clear, however, that, if the cutoff was too rapid, the phase quickly exceeded 180 degrees; the reduction in gain had to be gradual, over an extended frequency range. Feedback-amplifier designers would have to control transmission around the feedback loop to well above the useful passband—perhaps to frequencies 10 or 15 times higher than the top of that band. In the search for this design, they encountered a host of problems.

For one thing, the device, component, and assembly techniques of that time were ill suited to such broadband applications. The vacuum tubes in common telephone-repeater use were filamentary triodes with very low gain per stage and with large grid-plate capacitance, which introduced an unwanted local feedback loop in addition to the one desired. Designers experimented with 4-, 5-, and 6-stage amplifiers, trying to achieve enough gain so they could sacrifice the desired amount to feedback. They were continually defeated by the inevitable large phase shifts associated with so many stages, plus a multiplicity of unwanted couplings caused by grounding and shielding problems.

Operating the filaments of four vacuum tubes in series on the 24-v office battery added one more very annoying and unwanted coupling. The telephone-transmission designers pressed the vacuum-tube designers for more gain per stage and better high-frequency performance. In this they were joining hands with other amplifier designers pushing for vacuum-tube applications in radio systems, such as home receivers in the range from 500 to 1500 kHz, and high-frequency radio in the range from 3 to 25 MHz. Improvement came first with the indirectly heated cathode—an absolute requirement for home radio receivers to permit AC operation and to get rid of storage batteries to heat filaments—followed by the invention and development of a great variety of tetrodes and pentodes to reduce plate-grid feedback. The passive components (resistors, capacitors, inductors) in use in 1927 were mostly carryovers from the mechanical switching technology—large, cumbersome, and suited only to DC and low frequencies. Grounding and shielding were primitive. Components were mounted on equipment panels in an orderly array; wiring was in formed cables laid along the panels and racks. Much of this had been worked out as an advance over the hodgepodge assemblies of 1918 through 1920, but it was deadly for high-frequency design. Feedback paths were often yards long, with the inevitable accompanying phase shift.

III. DESIGN THEORY—NYQUIST'S CRITERION AND BODE'S SYNTHESIS

During this struggle on the hardware side, the theorists were advancing the understanding of negative feedback. (See another volume in this series, *Communications Sciences (1925–1980)*, Chapter 1, Section IV.) In a classic 1932 paper called "Regeneration Theory," H. Nyquist, using the concepts of complex-function theory, made a landmark contribution to feedback theory by establishing the conditions under which a transient would grow (instability and oscillation) or decay (stability and nonoscillation).[3] Nyquist's criterion is typically used by designers in the form of a vector plot of $\mu\beta$ [Fig. 4-6]. If the plot encloses the point *1,0*, the circuit is unstable and will oscillate; if it does not, the circuit is stable. (It is interesting to note that there is a class of circuits in which the gain may be greater than unity and the phase zero, that are nevertheless stable [Fig. 4-7]. These conditionally stable circuits actually offer the possibility of more feedback than the unconditionally stable class but are tricky to design and have found little application in telephony.) With Nyquist's criterion in hand, designers knew at last exactly what they were trying to do, and had an ironclad analytic test to determine whether they had succeeded or not. This knowledge was most helpful, but still left them with no straightforward method to synthesize circuits to specified conditions, getting the most feedback obtainable with prescribed margins against instability. The great advance on that front was the contribution of Bode.

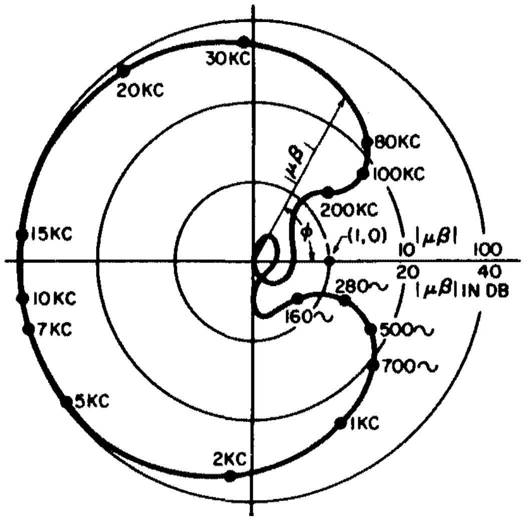


Fig. 4-6. The Nyquist criterion—a vector plot of $\mu\beta$ for a stable loop. [Bode, *Proc. Symp. Active Networks and Feedback Systems* **10** (1960): 8.]

During the early 1930s, work was started on the first carrier telephone system on coaxial cable involving a passband of 1 MHz or more, with the prospect of several hundred repeaters in tandem. (See Chapter 6.) This was by far the most ambitious feedback-amplifier design of its time. In 1934, as part of this project, Bode was asked to design a variable equalizer to compensate over a considerable range for the effect of temperature variations in the coaxial line. This was a fairly difficult problem in itself, but it was also proposed that the equalizer be inserted in the feedback path (β circuit) of a feedback amplifier, which was otherwise already designed, without causing instability. Bode wrestled with this difficult problem for a long time, concentrating on his equalizer, without

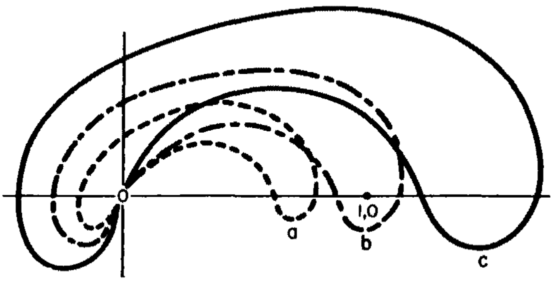


Fig. 4-7. The Nyquist criterion—*a*, stable; *b*, unstable; *c*, conditionally stable. [Bode, *Proc. Symp. Active Networks and Feedback Systems* **10** (1960): 9.]

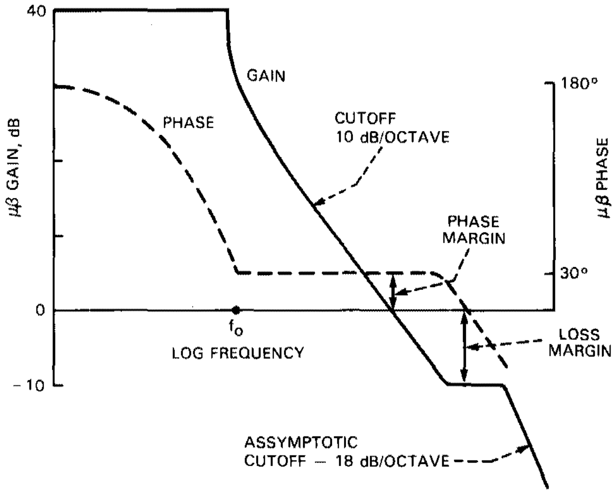
success. He then began to tinker with the input and output circuits and interstages of the amplifier. Finally, he found a solution, but only by redesigning the complete feedback loop. There was no other way to meet the equalization and stability requirements simultaneously.

As a result, Bode addressed his attentions to "several simple relations between the gain around an amplifier loop, and the phase change around the loop, which impose limits to what can and cannot be done in a feedback design."[4] Bode not only defined these limits, but showed how the powerful tools of network theory can be used in the synthesis of stable amplifiers designed to meet desired objectives. In the final analysis, the very-high-frequency gain of a vacuum-tube amplifier is determined by the shunting capacitance of its elements to ground. There is a high frequency above which the amplifier will have less than unity gain (0 dB) working into its own parasitic capacitances. Working back from this asymptote, Bode showed exactly how to design to meet specified margins in gain and phase to prevent oscillation and, given those margins, how much feedback could be realized for any bandwidth.[5] Designers were at last on solid ground.

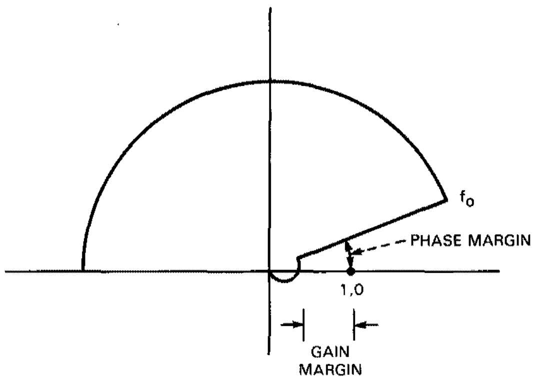
Bode's analysis confirmed and quantified the earlier realization concerning the wide frequency band that must be controlled. If the phase shift may not exceed 150 degrees (30-degree phase margin) until the loop has the desired loss margin, then the rate of cutoff cannot exceed 10 dB per octave. If the desired loss margin is 10 dB and the in-band feedback desired is 40 dB, then the cutoff must extend over five octaves—a factor of 32 [Fig. 4-8]. The designers of a so-called 1-MHz amplifier found themselves required to design and control performance from very low frequencies to about 30 MHz, no easy task with the instruments and components then available, but inescapable with the constraints their hard-won knowledge placed on the art.

Negative feedback has become ubiquitous in all fields of communications, and its use in other areas, such as industry, military, and consumer electronics, far exceeds the use in telephone communications. Black recognized early on other fields of applications. He foresaw that the understanding developed in connection with electronic amplifiers could be carried over and applied to mechanical or acoustical systems as well. After all, feedback systems were not new to the world—there were examples all around: the sailor steering a ship by compass, the whirling balls controlling a steam engine. It was because of this perception that he insisted that the claims in his patent be written very broadly, and why he worked so tenaciously with the United States Patent Office to keep the broad applicability as granted.

Since the invention, applications have grown in a spectacular way. Modern control systems, essential to automation in an enormous variety of fields, would not be possible without the application and understanding of feedback. Within Bell Laboratories, the principles have been extended and applied to fields as diverse as military gun control and rocket guidance. All these systems must be carefully designed to prevent them from becoming dynamically unstable. The necessary theory and design techniques were developed mainly at Bell Laboratories and were made freely available to the electronics and control industries. The invention has endured as a major technical milestone for almost 60 years, and seems certain to rank as one of the major inventions of the 20th century.



(a)



(b)

Fig. 4-8. The $\mu\beta$ feedback loop design for 10-dB gain and 30-degree phase margin against oscillation. (a) Gain-frequency plot. (b) Polar plot.

REFERENCES

1. H. S. Black, U.S. Patent No. 1,686,792; filed February 3, 1925; issued October 9, 1928.
2. H. S. Black, U.S. Patent No. 2,102,671; filed April 22, 1932; issued December 21, 1937.
3. H. Nyquist, "Regeneration Theory," *Bell Syst. Tech. J.* **11** (January 1932), pp. 126-147.
4. H. W. Bode, "Relations Between Attenuation and Phase in Feedback Amplifier Design," *Bell Syst. Tech. J.* **19** (July 1940), p. 422.
5. See reference 4, pp. 421-454.

Chapter 5

Carrier on Cable

I. INTRODUCTION

The invention of negative feedback could not have come at a more auspicious time. During the last half of the 1920s, system growth was extraordinarily rapid; the major growth facility in these years was voice-frequency, loaded, repeatered cable. By the end of the 1920s, circuits in cable accounted for over 60 percent of the network voice-circuit mileage. This growth in capacity was accomplished partly by extension of the cable network to new areas, but also by enhancements in the capacity of existing routes. The increase in traffic along the busiest routes constantly pressed on the capacity of the facilities. Just as the increase in traffic had outgrown the capacity of the earlier open-wire lines, it now threatened to outgrow the capacity of the recently installed cables. Since carrier had proven so useful in increasing the capacity of open wire, it was only natural that system designers should consider carrier as a means to increase the capacity of cable. As early as November 1927, Bell Laboratories and AT&T engineers analyzed the possibility of adding one or two carrier channels over voice on the side circuits in a cable quad (four wires, two twisted pairs twisted together). The difficulties were formidable. Not only did amplifiers of the required stability and linearity not exist (this was before the general availability of feedback amplifiers), but the designers, fresh from their struggle to deal with cable crosstalk at voice frequency, recognized that crosstalk, perhaps more than any other factor, would set a limit to the frequencies available. They concluded, however, that if it were technically feasible, carrier did indeed promise savings over the simple addition of wires to a route.

The studies continued. Early in 1928, designers considered the possibility of deriving two voice circuits from the 8-kHz program channel for cable, then in development. In May 1928, a report was issued on the prospect for carrier on nonloaded or lightly loaded cable pairs. No. 16 gauge pairs were considered necessary to offset the higher loss that would be met without the usual loading. The loss of unloaded No. 16 gauge cable pairs was about 1.5 dB per mile at 20 kHz. There appeared to be no prospect of coping with the loss over the normal 50-mile spacing between voice-frequency repeaters. Intermediate repeaters would be required at least at the 25-mile midpoint. On May 19, 1928, a key meeting took place: H. Nyquist and others conferred with H. S. Black on the properties of his improved amplifier with the objective of applying it to the cable carrier problem. It indeed proved to be the key. In November 1928, a specific proposal was made for a field trial of carrier on nonloaded No. 16 gauge pairs in cable. One thousand miles of two-way circuit was to be derived by looping pairs in 25 miles of cable. Repeaters would be spaced 25 miles apart, and the system was initially planned for eight carrier channels in the band from 4 to 36 kHz. Acceptance was prompt. Early in 1929, a complete development program was outlined by A. B. Clark of Bell Laboratories that included extensive work on cable as well as electronics. As recorded in the notes of a conference at AT&T headquarters in April, it was agreed that the development should proceed as soon as possible "on a class A basis."

Many problems had to be considered, especially about the cable itself. The wonderful new amplifier had done nothing to alleviate the problem of crosstalk. Ideally, it would have been best to accomplish two-way transmission within single existing cables, but at an early stage, it was recognized that this was not possible with the cables and technology of that time. Near-end crosstalk, the coupling of the high-level, high-frequency output signals of an amplifier in one direction into the low-level input circuits of the oppositely directed amplifier, within the close confines of a single cable, ruled this out. Even if it were physically possible on a few specially prepared and located pairs, it would prove unduly restrictive on every other system parameter. The useful band, the number of channels, the number of pairs that could be used, and repeater spacing would all be severely restricted. It was necessary to use separate cables with the shielding provided by metallic sheaths for the opposite directions of transmission.

But even with the decision to use separate cables, many issues remained. Could existing installed cables be used, or were new improved designs required? Were No. 16 gauge wires required, or was No. 19 gauge satisfactory? (A gauge increase of three doubled the amount of copper and always raised the simple alternative of running more pairs.) Should cables of mixed gauges be planned? Was there any role for loading? Should different structures be considered? The so-called spiral-four quad (four wires, not paired, but twisted symmetrically) was widely used in Europe and had 15 percent lower capacity per pair. What role in the plant expansion would the newly proposed coaxial cable play? All these, and a host of other problems, were studied and debated at length. The urgency for speed remained strong, however, and, in February 1930, a decision was reached to order and place 25 miles of conventional, quadded, paired cable, including a substantial number of No. 16 gauge pairs sandwiched among No. 19 gauge pairs for shielding.

Even after that decision, the study of, and experiments with, cable alternatives continued for a year or two. As business conditions deteriorated in the early 1930s with the onset of the Great Depression, plant growth slowed abruptly. This brought about renewed interest in the application of carrier to cables already in place to defer the necessity of expensive new line construction, as even the slow growth exhausted the capacity of existing lines. The slowdown effectively ended the speculation about novel cable designs and larger-capacity systems (up to 23 channels had been considered on carefully manufactured cables of otherwise conventional design). Most of the experimental special cables were not made in any appreciable quantity and ultimately the design objective settled on 12 carrier channels, with application primarily to existing cables. A factor in this final position was the growing expectation that coaxial cable would be equally economical, or even less expensive, for a much larger number of channels, plus its prospect for carrying a wideband television signal.

Another issue that occupied much attention and was widely debated was the need for and practicality of unattended repeaters. There was much effort to avoid the necessity for them, and much of the continued interest in special loading and novel cables lay in the prospect that unattended electronic equipment might be eliminated, or at least reduced in quantity. Part of the concern was economic and objective. The repeater stations would be fairly complex and expensive. They would depend on none-too-reliable local 60-Hz power and require expensive converters and backup power sources. But part of the concern also lay in the attitude toward the still relatively new vacuum-tube technology. By 1930, vacuum tubes had been made quite reliable, but memory of the first struggling days was still fresh. Tubes were fussy compared to passive apparatus; they failed and had to be replaced; amplifiers required frequent adjustment. In system studies as early as March 1929, Nyquist had pointed out the advantages of two, rather than one, unattended intermediate repeaters. But in late 1930, alternatives involving continuous loading and very-high-power repeaters were still being studied in an effort to eliminate unattended auxiliary repeaters. No doubt the depressed economy was a factor in this as well, but the success of the trial, which included a test of unattended equipment, finally settled the issue. Unattended repeaters at about 17-mile spacings, two between each of the existing attended voice repeater stations spaced at 50 miles, became the accepted design.

The files on the development of cable carrier do not deal directly with the collapsing economy of the time, but they bear sadly eloquent witness to the impact of that disaster on the urgency of telephone developments. In early 1929, the need for speed was emphasized repeatedly. Decisions were promptly made on cable structures, repeater equipment, and the trial site. In the summer and fall of 1930, there was a renewed emphasis on total costs and savings, and studies of light routes and slow growth reappear in the record. Early in 1931, estimates of growth were further reduced. In May 1931, a conference to consider the site for an initial commercial installation following the trial ended with the appointment of a committee to study the question. In June, Bell Laboratories notified AT&T that the development of a nine-channel system could be completed within the original cost estimate, but recommended further development before standardization. The recommendation was accepted. An analysis of the value of early completion showed little profit in having a commercial design earlier than 1934. In the same month, a memo considering the final configuration of the system suggested that an executive conference to consider the subject be held in September! Clearly the pressure was considerably reduced from the buoyant days of early 1929. Nevertheless, the development work continued.

II. THE MORRISTOWN TRIAL

The site chosen for the first field experiment was Morristown, New Jersey, located about 25 miles west of New York City on the existing New York-

Chicago cable route.[1] Twenty-five miles of cable was installed in existing underground ducts. The cable was looped, that is, both ends were in Morristown with the two 12-1/2 mile sections laid out to the west, to a doubling-back point [Fig. 5-1]. The cable, made specially for the test, contained 68 No. 16 gauge pairs in quads, in addition to a complement of No. 19 gauge quads. The 34 No. 16 gauge quads were connected in series to form an 850-mile, two-way, four-wire circuit on which the carrier circuits were to be transmitted. Nine carrier channels were derived in the band from 4 to 40 kHz. The modulators and demodulators (modems) for the frequency translations used techniques similar to those used in the three-channel carrier system for open wire [Fig. 5-2]. (See another volume in this series, *The Early Years (1875-1925)*, Chapter 4, Section 4.2.5.) All the repeaters, as well as the multiplexing terminal equipment, were at Morristown.

Although the circuits were true two-way circuits, all transmission through the cable was in one direction. The two 850-mile one-way pairs of the four-wire circuit were established by appropriate interconnections (patching) at Morristown. Connection to a cable end was therefore either high or low level, but not mixed. Despite this, extreme care was necessary to keep crosstalk within acceptable limits. Within the repeater station, the high-level transmitted signals were at voltages and currents 300 times greater (50 dB, a factor of 100,000 in power) than the low-level received signals. Careful segregation and shielding were required to keep the in-station crosstalk paths within limits. A great deal of emphasis and attention was devoted to the crosstalk issue, since, with the feedback amplifier available, crosstalk, rather than stability or linearity, loomed as the largest problem.

The experience with repeatered voice-frequency cable led the designers to consider capacitive coupling between pairs as the principal crosstalk linkage. To reduce coupling, they sandwiched the No. 16 gauge pairs between No. 19 gauge pairs, which acted as electrostatic shields. Far-end crosstalk was not aggravated by the high gain of the repeaters, since all signals were treated alike

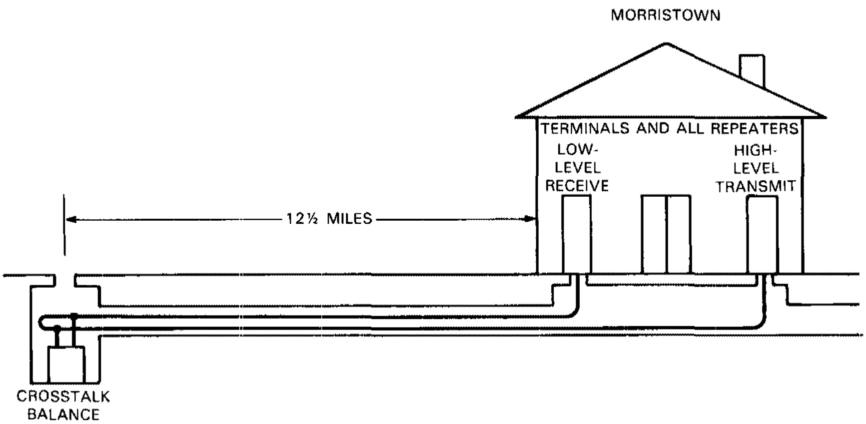


Fig. 5-1. Diagram of the Morristown cable-carrier trial.

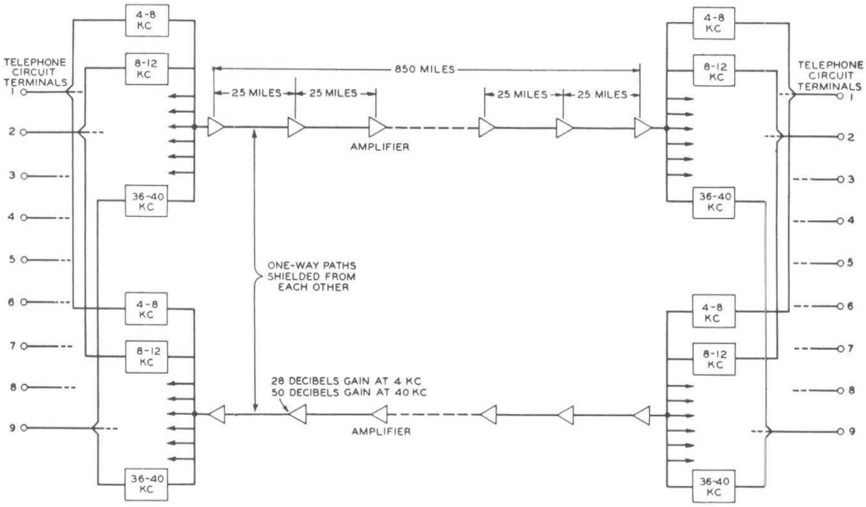


Fig. 5-2. Morristown trial layout.

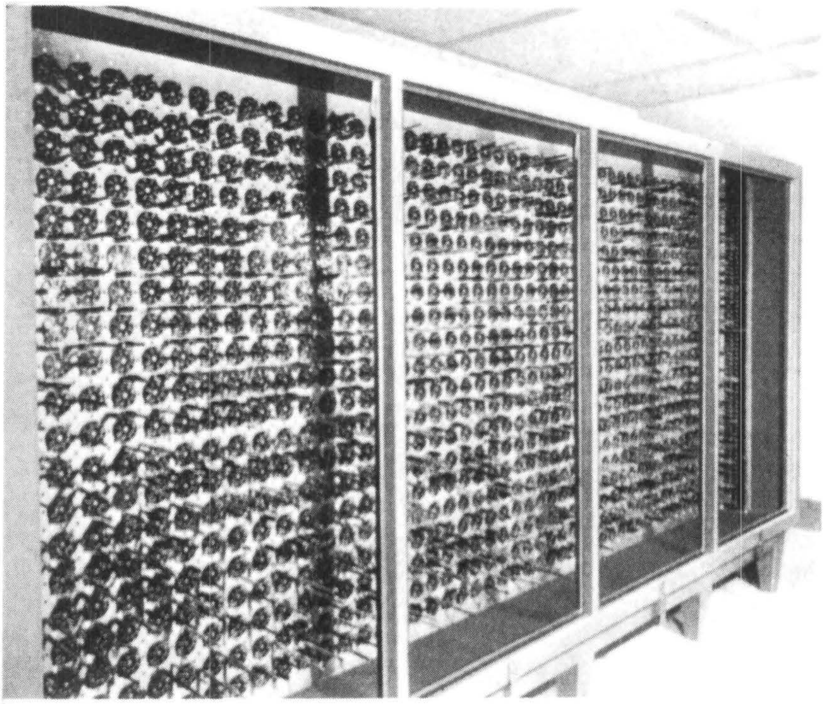


Fig. 5-3. Special crosstalk balancing panel for the Morristown trial.

and remained approximately alike in level at all points along the cable. But even with the most carefully constructed new cable, and with the shielding from the intervening pairs, far-end crosstalk did not meet objectives at the high carrier frequencies. However, it was known that with unloaded pairs, the crosstalk components coupled along the line into the disturbed circuit and the disturbing signals all traveled at the same velocity, reaching the same points at the same time. This made it possible to reduce the net crosstalk by balancing, i.e., introducing reverse-phase coupling at a single point. In principle, this could have been done at any place along the 25-mile section. In the field test, the balancing was done at the looping site, the midpoint of the repeater section. Balancing was accomplished by connecting small capacitors between every pair and every other pair used for carrier. Balancing required a large array of capacitors, but proved to be effective, to the extent that the intervening pairs were found to be unnecessary [Fig. 5-3]. It was concluded that balancing could be made effective over even wider bands than were used on the trial.

2.1 Transmission and Temperature Regulation

The 850-mile circuit had a loss of about 1300 dB in the middle of the band. This unprecedented loss had to be compensated by the gain of the amplifiers to within 1 or 2 dB over the entire band. By rigorous measures to exclude all noise from external sources, the noise on the line was brought close to the limit imposed by the thermal noise of the conductors and the somewhat greater tube noise. To provide amplifiers with a power capacity much above 1 w would have been very expensive and may not have been technically feasible, given the other performance qualities needed. This was one of the major factors in the decision to use intermediate repeaters. With the repeater spacing at Mor-

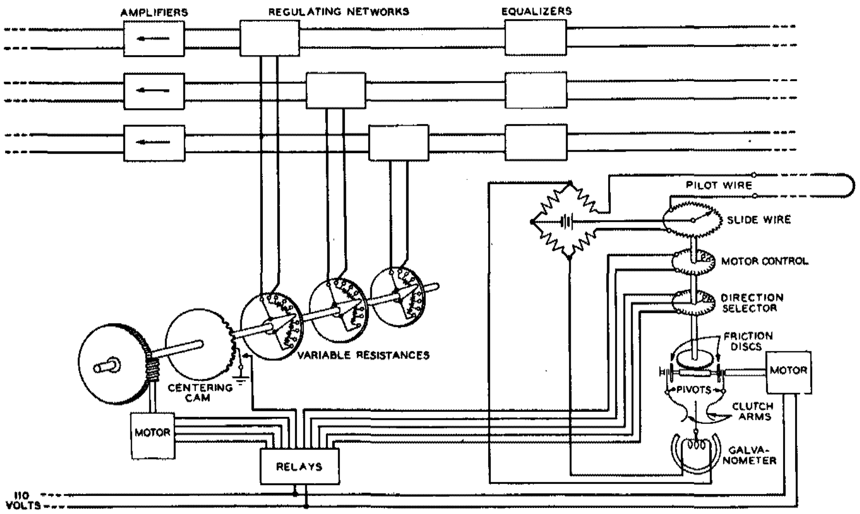


Fig. 5-4. Pilot-wire automatic transmission-regulating system for the Morristown trial.

ristown, it was possible, with triode amplifiers of moderate power, to keep the signal levels high enough at the repeater input so that noise was not a significant factor.

A major problem in the cable carrier system was the change in loss caused by the temperature changes in the cable. For underground cables in the temperate part of the United States, temperature varies by about ±18 degrees F around a mean of 55 degrees F. For a circuit of 1000 miles, this would result in a loss change of about 100 dB; clearly, some form of transmission regulation against temperature change was required. To cope with this at Morristown, the designers included a pilot wire-regulation system, similar in principle to systems previously used on repeatered voice-frequency cables [Fig. 5-4]. To a fairly good approximation, transmission loss could be related to the DC resistance of a pair of wires in the same cable as the transmission pairs. This pilot wire was made one arm of a self-balancing Wheatstone bridge. The regulating networks were designed so that the change in a single resistance caused the network transmission loss to vary with frequency in the same way as the line loss varied with temperature. The relay and motor-drive system was arranged to rotate the shaft controlling the regulating resistances so that it followed the shaft, adjusting the self-balancing bridge.

2.2 Results

Nine separate channels were transmitted successfully over the trial link without difficulty due to intermodulation or excessive crosstalk. With amplitude equalizers and the temperature regulation, fairly flat transmission over the frequency band was achieved and maintained, although it was noted that the equalization and regulation "presented intricate and difficult problems of design."[2] Terminal multiplexing was accomplished by translating each voice signal to the assigned line frequency in a single step by means of balanced vacuum-tube modulators and the appropriate modulating carrier frequency. As in the open-wire carrier systems, the carrier was suppressed by the balance of the modulator, and only one sideband, selected by a channel filter, was transmitted. Single-sideband, suppressed-carrier transmission kept the signal power and the total frequency band required as low as possible. The translated voice signal at line frequency was confined to its assigned slot by conventional coil and capacitor filters.

The objective of the trial was to demonstrate satisfactory transmission over circuits up to 4000 miles long, with as many as five links (i.e., lines and terminals) in tandem. Transmission over transcontinental distances was envisioned from the outset, and conclusions as to satisfactory or unsatisfactory performance were made with respect to that objective. Tests of transmission over very long systems, including repeated modulation and demodulation in the terminals, were possible by connecting the 850-mile, carrier-derived channels end to end. By such a linking, a voice band of 2500 Hz, from 250 to 2750 Hz, was realized through five links in tandem. Finally, all nine carrier links were connected in tandem for 7650 miles of two-way circuit, and the quality was found to be satisfactory. As a last measure, all the lengths were connected to form a 15,300-mile, one-way circuit, through which transmission was "not greatly impaired."[3]

The line loss of this ultimate tandem link without amplifiers would have been about 24,000 dB. (A popular occupation in those days was to note that such ratios transcended other huge ratios, such as the size of the total universe compared to the size of the smallest known particle of matter.)

The Morristown trials were eminently successful and demonstrated conclusively that carrier on cable was indeed practical. In a sad epilogue, the article describing the trial concludes, "Under the present economic conditions there is no immediate demand for the installation of systems of this type," followed, however, by the brighter note that development work was being continued in anticipation of ultimate commercial use.[4]

III. COMMERCIAL CABLE CARRIER—TYPE K CARRIER

Development did indeed continue toward a commercial system but with significant changes in objectives from the trial system.[5] Transmission was to be primarily over existing installed cable. The system was expanded from the 9 channels of the Morristown experiment to 12 channels, and the frequency band changed from the 4 to 40 kHz of the trial to 12 to 60 kHz [Fig. 5-5]. The band choice was an economic judgment, as there was no sharp technical limitation to the frequencies that could be used at either end. Higher frequencies increased the crosstalk problem and entailed higher loss, which made repeaters more difficult to realize. A lower bottom frequency would increase the ratio of bottom-to-top frequency and make the circuits more difficult to design. In particular, it would increase the problem of accurate compensation for temperature changes. As in the experiment, the carrier channels were to be transmitted by single sideband with suppressed carrier, at what was by then the standard interval of 4 kHz. No voice circuits were applied below the carrier. These changes intensified some of the old problems and brought some new ones into prominence.

In the late 1930s, there was about 15,000 miles of cable installed, commonly made up of No. 19 gauge wires, with repeaters at attended stations at about

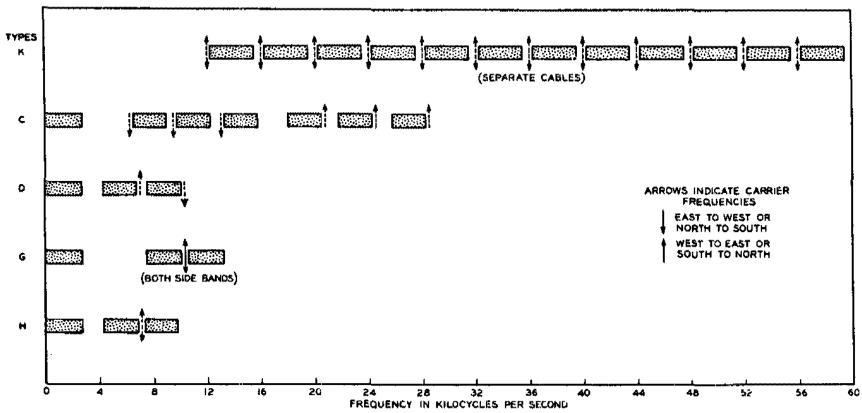


Fig. 5-5. Frequency allocation of Type K carrier and open-wire carrier telephone systems.

50-mile intervals. With the loss of No. 19 gauge pairs at the carrier frequencies, the existing amplifier art limited the spacing to a maximum of 19 miles. To coordinate with the existing stations, two intermediate repeaters about 17 miles apart, became the usual layout. Furthermore, about two-thirds of the existing mileage was aerial cable, which was exposed to a temperature range (and corresponding change in loss) about three times greater than that for underground cable [Fig. 5-6]. And, while the temperature of underground cable changed only slowly with the seasons, the daily variation of aerial cable could be as much as one-half of its much greater annual change. Regulation to maintain constant transmission loss, therefore, had to cope with a much larger range and with a rate of change several hundred times as rapid as was experienced in the underground cable of the trial.

The overall objective remained—to realize high-quality transmission over distances up to 4000 miles, and with as many as five links in tandem. Added emphasis was placed on reducing the cost of the multiplexing terminals, as the minimum distance at which the carrier system would prove economical was largely determined by the terminal cost, which was, of course, a constant and independent of the cost of the intervening carrier line. Low-cost terminals would not only bring the economy of the carrier system to a broader field of application, but would also increase the volume of production and help to reduce costs further.[6,7] Work continued through the 1930s with the new technical and cost objectives in mind. In 1937, a commercial version was placed in service, on a trial basis, on 150 miles of cable between Toledo, Ohio and South Bend, Indiana.

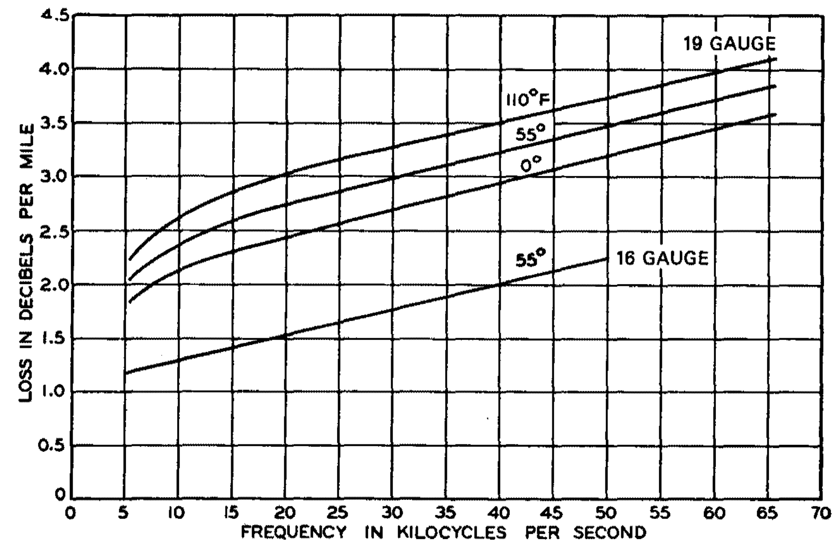


Fig. 5-6. Attenuation of Nos. 16 and 19 gauge nonloaded cable pairs as a function of frequency. Change in loss for No. 19 gauge pairs is shown for a typical range of aerial-cable temperature.

3.1 Crosstalk and Noise

Thermal and tube noise set an absolute lower limit to the level to which signals could be allowed to fall. At the repeater output, the high linearity of the feedback amplifiers kept intermodulation between channels satisfactorily low, even at levels close to overload. The system was roughly in balance with respect to the major parameters of bandwidth, noise, intermodulation, and crosstalk, but the early state of system-design technology, and the relatively uncharted territory of the existing cable plant in the new frequency range prevented a truly quantitative optimization. Crosstalk, in particular, continued to be a major concern. It required a great deal of effort and some special measures to keep it within tolerable bounds.[8] Careful shielding at repeater stations, and special filters where branch cables or open-wire lines connected to the main cable were always required. As in the experiment, the primary defense against crosstalk was the use of separate cables for the opposite directions of transmission. At each repeater, the transmission path was transposed to the opposite cable [Fig. 5-7]. By this means, each cable was either low level or high level at the repeater, but not both. This prevented the otherwise high crosstalk coupling that would have resulted via the metallic voice pairs that passed directly through the repeater station in the same cable with the carrier channels.

With the use of existing cables and higher frequencies, crosstalk balancing became more necessary and more difficult. The approach was basically the same as that used in the earlier trial; the crosstalk due to the cable structure was balanced out by introducing a small inverse coupling from every carrier pair to every other carrier pair in the same cable. In Type K carrier, this was done at every repeater and by small mutual inductances, rather than with capacitors as used on the Morristown trial [Fig. 5-8]. With the coupling units in a matrix array, it was possible to associate every pair with every other pair through simple wiring paths. The shift to balancing by inductors was a fruitful one. At voice frequency, capacitive coupling between cable pairs was the principal crosstalk path, and much care had been taken in cable design to minimize

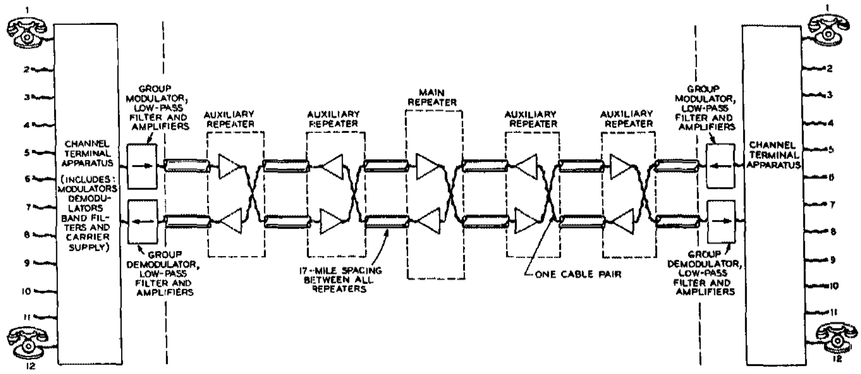
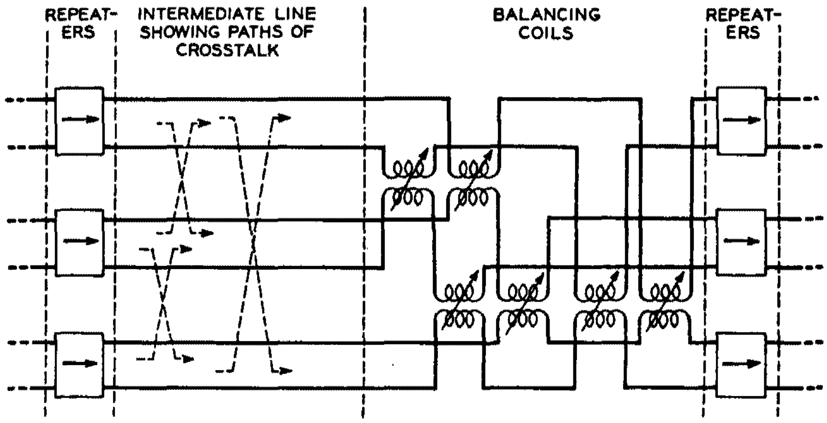
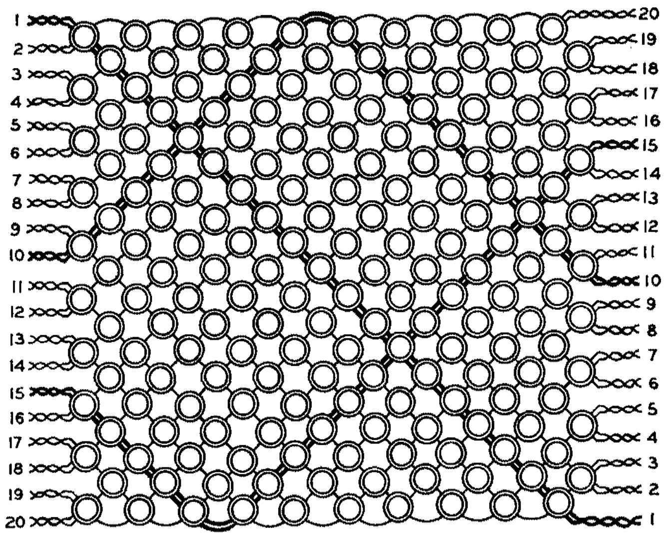


Fig. 5-7. Schematic of Type K carrier system. Main repeaters and terminals were at attended sites. Transmission was transposed between cables of each repeater.



(a)

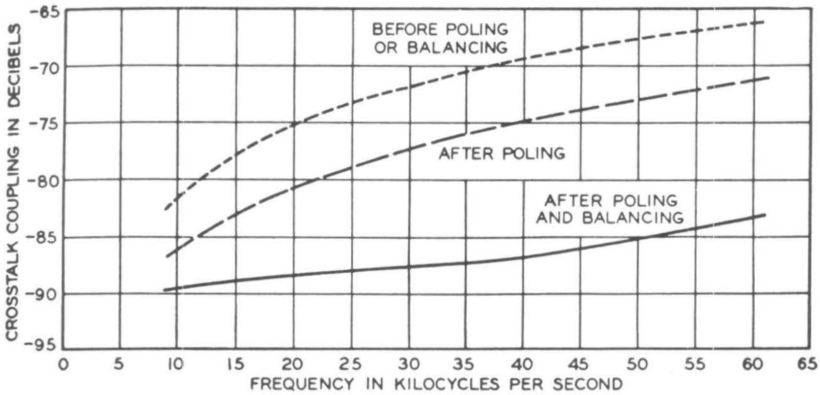


(b)

Fig. 5-8. Crosstalk balancing in Type K carrier. (a) Balancing with mutual inductance between pairs. (b) Schematic arrangement of coupling units in the complement for 20 circuits.

capacitive imbalance. When carrier was applied to cable, both capacitive and inductive coupling increased directly with frequency, but unloading the pairs lowered the line impedance from about 800 to 135 Ω. Capacitive coupling decreased in direct proportion to the lower line impedance, while inductive coupling increased by the same ratio. Compared to a voice-frequency line, the combination of impedance change and higher frequency for Type K carrier

(a)



(b)

Fig. 5-9. Crosstalk poling. (a) Measurements for crosstalk poling as actually carried out in the field. (b) Far-end crosstalk per repeater section from measurements on three repeater sections of cable.

increased capacitive coupling by a factor of 2.5, while inductive coupling was increased by a factor of 90. As a result, the inductive linkage, instead of being the minor component, was about three times as significant for crosstalk as capacitive coupling. The use of small mutual inductors gave 5 dB more reduction than would have been possible with capacitors.

But balancing alone was not sufficient, and several other measures were necessary to keep crosstalk within acceptable limits. An extensive program of testing and upgrading the cables was launched. If the many pairs in a cable were spliced in a systematic pattern, as was often the case, pairs that were poor from the standpoint of crosstalk coupling, become proportionally worse for longer cable lengths. In preparation for carrier application, the cable pairs were respliced at all the old loading points (about every 6000 feet) in an attempt to approach the effect of random splicing. Because of their physical proximity, crosstalk between the two side pairs of the same quad was especially difficult to control. This was reduced by *poling*: the side-to-side coupling was measured and the quads respliced at the midpoint of a repeater section, so that the coupling in one half section neutralized, as far as possible, that in the other half section. Even with poling, side-to-side crosstalk was greater than that for other pair-to-pair couplings. In a further measure, the quads were split at every repeater; pairs would not be associated in the same succeeding sections, diluting the side-to-side crosstalk in a long system. By all these means, the crosstalk was successfully reduced to a level suitable for 4000-mile transmission [Fig. 5-9].

3.2 Repeaters and Regulation

Three-stage amplifiers were used in the repeaters, with feedback from the output-stage plate through a feedback circuit to the input grid [Fig. 5-10]. The feedback circuit included a line equalizer shaped to offset the basic loss-frequency characteristic of the line. It also incorporated the variable elements to adjust the gain of the amplifier and to offset line-loss changes due to temperature

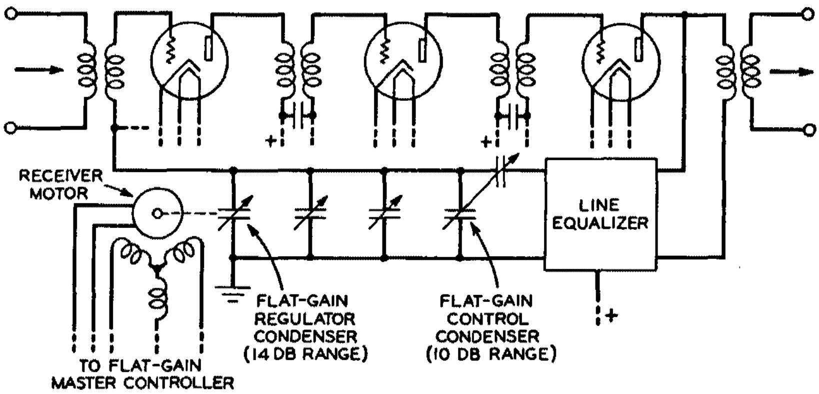
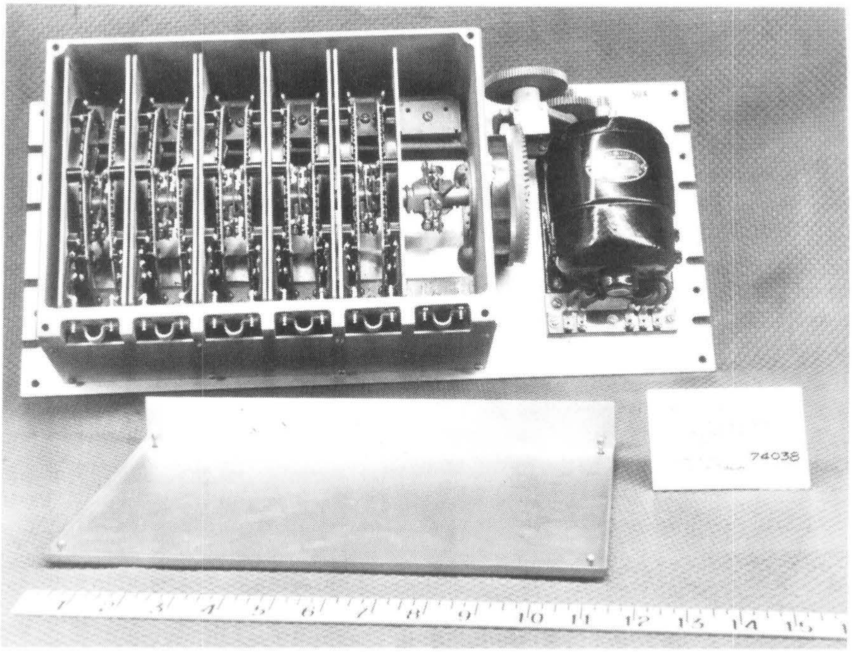
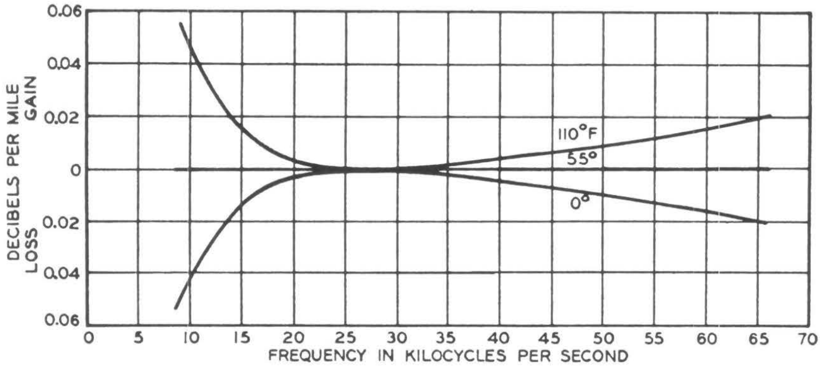


Fig. 5-10. Schematic of the Type K carrier line amplifier.



(a)



(b)

Fig. 5-11. Twist equalization. (a) Resistors by which adjustment in nonflat gain is accomplished. (b) Twist characteristics of No. 19 gauge nonloaded cable pair.

variation. Considerable progress in vacuum-tube technology had eased the life of amplifier designers. Pentodes with indirectly heated cathodes were used, the output stage being designed to provide more power than the first two "gain" stages. About 40 dB of feedback provided more-than-adequate gain stability and sufficient linearity so that intermodulation was not a problem. At

60 kHz (the maximum loss frequency), the individual channel signals at the repeater input were attenuated to about 60 dB below the voice level at the transmitting switchboard. They were then amplified at the repeater output to a level about 10 dB above the switchboard level.

Regulation for changes in line loss due to temperature changes was controlled by means similar in principle to the pilot-wire regulator used in the Morristown trial. A control motor in each repeater was synchronized to the master motor of a self-balancing bridge, in which a pilot wire of the cable formed one arm. The repeater motors tracked the master motor in rotation, adjusting variable condensers in the feedback circuits to make changes in the repeater gain to offset changes in line loss. This resulted in a gain change, essentially flat with frequency, at each repeater. Since the loss change of the cable with temperature was not precisely flat, an additional adjustable network, called a twist equalizer, controlled in the same fashion, was required after every six repeater sections (about 100 miles) [Fig. 5-11]. In addition, the repeater gain characteristic did not precisely match the desired nominal shape, and there were two different types of cable. Fixed deviation equalizers were required every tenth repeater span for the repeater difference and every 300 to 400 miles for the cable.

3.3 Terminals

The terminals introduced several major changes compared to the simple, direct-modulation scheme used in the earlier experiment. A two-step modulation

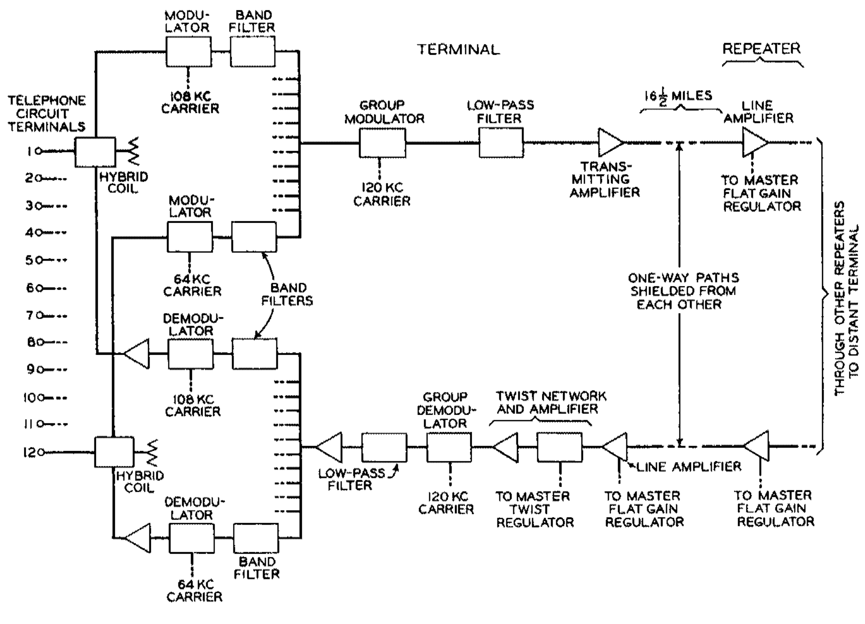


Fig. 5-12. Terminal arrangements for Type K cable-carrier system. Modulation to line frequency was in two steps. Each voice channel was limited to 4 kHz by quartz-crystal filters operating in the 60- to 108-kHz band.

sequence was used to shift the voice channels to their assignment in the 12- to 60-kHz line [Fig. 5-12]. This approach had several advantages; most notably, it made possible the use of quartz-crystal channel-band filters. The very low damping (high Q) of the quartz piezoelectric element permitted a very sharp cutoff and hence efficient use of the 4-kHz frequency band between the carriers [Fig. 5-13]. The filters provided a 3100-Hz band, essentially flat, from 200 to 3300 Hz, and developed very high loss to unwanted frequencies just outside the band. As an additional bonus, they occupied only one-eighth of the space of the coil and condenser filters used on the trial.[9]

Practical crystal filters, however, could not be built at frequencies as low as 12 kHz. It was first necessary to modulate each voice channel to a slot between 60 and 108 kHz to form a 12-channel basic group. The entire group was then shifted as a unit, in a second stage of modulation, to the line-frequency band from 12 to 60 kHz [Fig. 5-14]. This necessity turned out to have considerable virtue. During the cable-carrier development, it was planned that the multiplexing terminals would serve, as far as possible, as a common element for other broadband systems. Systems for coaxial cable and open wire were already in development during this period. The 60- to 108-kHz basic group became a universal standard for interconnecting a group of 12 channels between systems on coaxial and radio systems as well as on pairs. It was much easier to handle the less than two-to-one range of frequencies of the basic group than the much greater ratio of the minimum to maximum of the cable-line frequencies. The

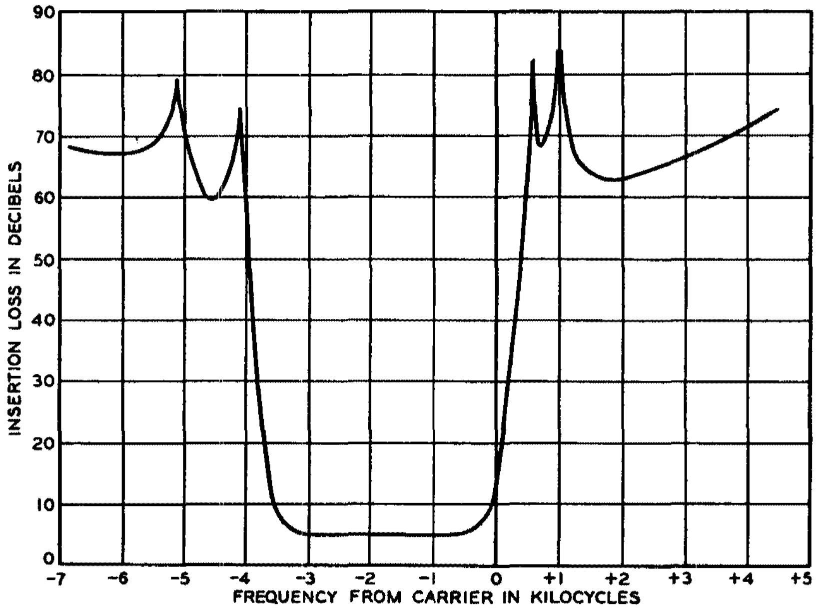


Fig. 5-13. Crystal channel filter transmission. When plotted in cycles removed from the carrier frequency, the insertion-loss versus frequency characteristics of each of the 12 crystal channel filters are for all practical purposes identical.

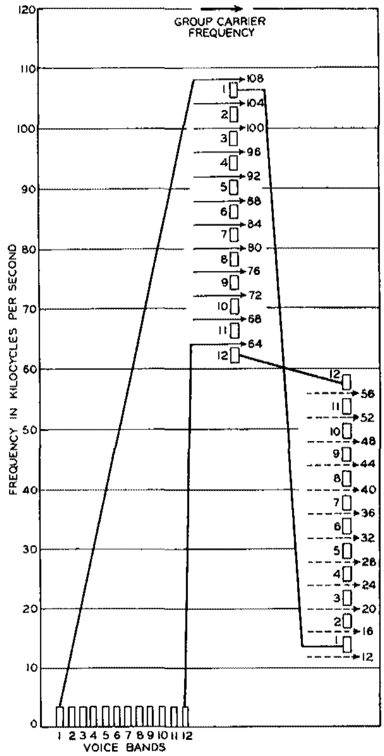


Fig. 5-14. In Type K carrier, 12 voice-frequency channels were first raised to occupy the band from 60 to 108 kHz, and then lowered to the band from 12 to 60 kHz for transmission over the cable.

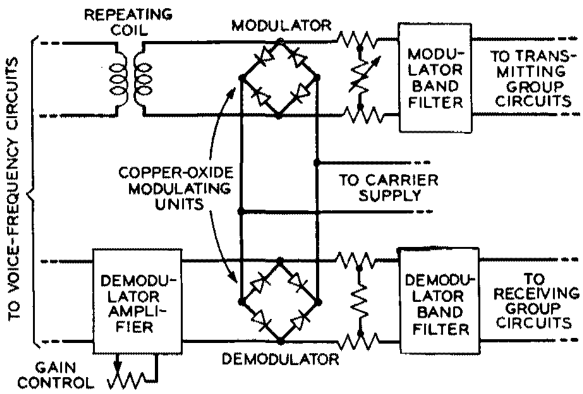


Fig. 5-15. Schematic of Type K carrier channel terminal equipment showing copper-oxide modulators.

60- to 108-kHz group-frequency block is one of the few worldwide communication standards.

The new terminals embodied many other features to improve performance and ease maintenance while reducing costs. Modulation was by copper oxide bridge circuits in place of the earlier balanced vacuum-tube circuits [Fig. 5-15]. These modulators provided good carrier suppression and life and stability approaching those of passive elements. In any nonlinear modulator, such as the copper oxide bridge, the amplitude of any modulation product is approximately proportional to the product of the amplitudes of the generating signals. In the group modulator, the use of a very high carrier drive (25 mw, +14 dBm) with very low channel signals (−46 dBm) accomplished the frequency shift with negligible modulation among the individual channels. The double balanced bridge-type circuit balanced out not only the carrier, but the 60- to 108-kHz components as well, simplifying the following line-frequency filter design.

3.4 Carrier and Pilot Supply

Since the carrier is not transmitted in a single-sideband system, it is necessary to supply demodulation tones at the receiving terminal substantially identical in frequency to those used at the transmitting modulators. Any difference between the frequencies will appear as a corresponding shift in the frequencies of the recovered voice signal. Experience with earlier systems and laboratory tests had demonstrated that a shift of only a few hertz was tolerable. In the Type K carrier terminals, the required 12 carrier frequencies were derived from a 4-kHz oscillator controlled by a tuning fork [Fig. 5-16]. A fork made of an alloy with a low temperature coefficient provided frequency stability of about one part per million per degree F, which was adequate for systems in this frequency range with terminals at main stations where temperatures were not likely to differ greatly. The output of the oscillator was amplified and used to drive a highly nonlinear (saturable) magnetic coil. The coil generated an alternately positive and negative spiky signal, rich in odd harmonics. These harmonics, isolated by appropriate filters and amplified, were used directly for the

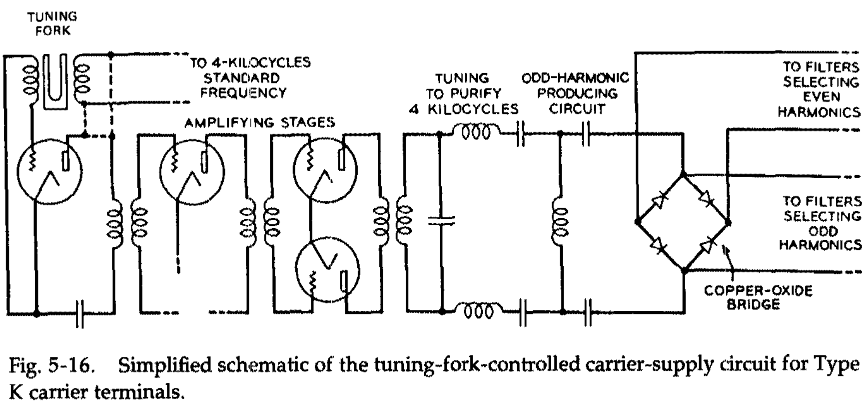


Fig. 5-16. Simplified schematic of the tuning-fork-controlled carrier-supply circuit for Type K carrier terminals.

carriers that were odd harmonics of 4 kHz. In addition, a portion of the coil signal was rectified, producing a signal of only one polarity, rich in even harmonics of 4 kHz. The rectifying bridge structure also balanced out all odd harmonics at the even-harmonic output. A pure carrier was needed at each channel modem so that the shifted channel would occupy only its assigned frequencies in the 60- to 108-kHz band. With separation of odd and even harmonics, the requirements on the following carrier supply filters was much eased. The array of 12 carriers was then distributed to the channel modulators by buses common to ten systems. The 30th even harmonic (120 kHz) was selected by a filter, amplified, and supplied by bus as the modulating source to the group modems. This carrier was used for the group translation between the 60- to 108-kHz band and the line frequencies from 12 to 60 kHz. Since the carrier supply was common to as many as ten systems, a failure would be catastrophic, so for protection it was completely duplicated. The group-carrier bus was continuously monitored, and, in the event of a failure of the 120-kHz carrier, the standby was automatically switched into service.

The Type K carrier terminals also addressed a line-operation and maintenance problem. It was essential to be able to determine that the carrier line was in good transmission order without taking it out of service or requiring a good deal of human cooperation between widely separated sites. To accomplish this, pilot tones were generated and transmitted over the line. The pilots were derived from a 3.9-kHz oscillator similar in design to that used for the carrier generator. This signal was modulated against carriers at 68, 96, and 108 kHz; the resulting products at 64.1, 92.1, and 104.1 kHz were introduced into the 60- to 108-kHz band. After the subsequent group-modem translation, the pilots appeared on the line near the bottom, middle, and high end of the band at 15.9, 27.9, and 55.9 kHz, to be transmitted along with the voice signals. Pilot measurements were made with high-impedance test sets that could be bridged across the line without interfering with through transmission. Offsetting the pilots from the carriers by only 100 Hz kept them out of the recovered voice bands, but made it possible when measuring to discriminate between the pilot tones and carrier leaks by a simple filter.

3.5 Equipment and Operation

Signaling over Type K carrier was on a ring-down basis, using a 1000-Hz tone interrupted 20 times per second, the same method that had been used on the earlier open-wire Type C carrier system. No special feature in the carrier system was required for signaling, as the interrupted tone was transmitted from switchboard to switchboard in the same way as the voice signals.

At terminals, equipment was grouped in functional assemblies, that is, all the channel modems for as many systems as were equipped were in the same bay or in adjacent bays. Group modems were similarly collocated [Fig. 5-17]. The carrier supply, as previously noted, was common to as many as ten systems. This equipment approach reflected not only the developer's confidence that the systems would be installed in large numbers, but also made the design, access, and testing arrangements more consistent. It was much more straightforward to design, use, and maintain a high-frequency bay or a voice-frequency

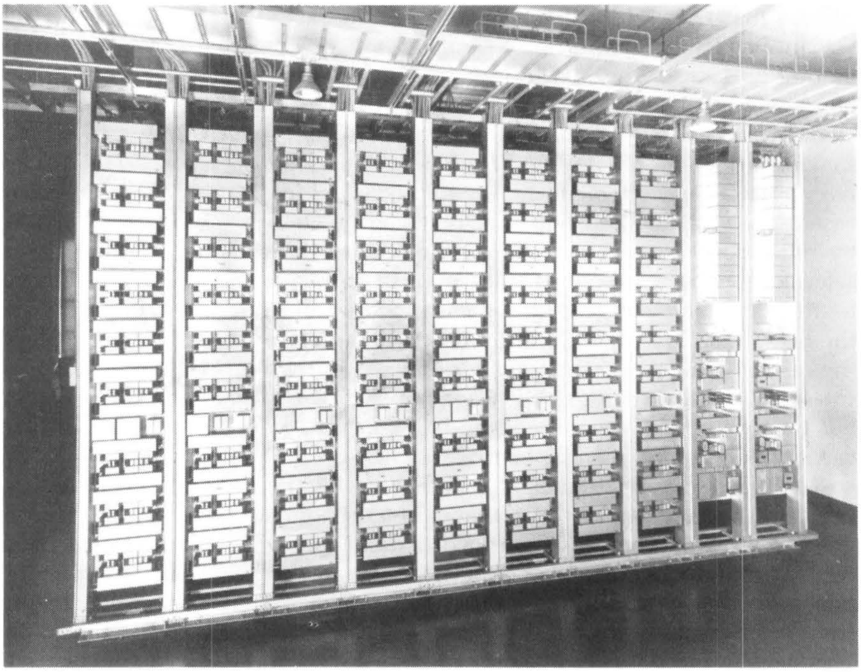


Fig. 5-17. Rear view of the channel-terminal bays for Type K carrier telephone systems together with their two carrier supply bays, *far right,* as they appeared in the Long Lines building at 32 Sixth Avenue, New York City, about 1940.

bay than a mixed assembly. Power was furnished by 150-v and lower-voltage batteries floated across the local AC supply. Standby 60-Hz power was available from motor generators that could be fired up if commercial power failed.

The unattended auxiliary stations were substantial huts of about 600 square feet [Fig. 5-18]. This provided sufficient room for as many as 100 systems (1200 channels, again reflecting that confidence in the future) [Fig. 5-19]. Power at auxiliary stations was also furnished by a 150-v battery floated across rectifiers on the local 60-Hz commercial power. In the event of an AC-power failure, the battery provided a reserve until AC power was restored. In addition to the carrier repeaters and necessary associated equipment, an "order-wire" voice-frequency talking circuit was provided between the auxiliary stations and the adjacent attended stations. Abnormal conditions, such as low voltage or current, or the absence of a signal, were made to generate an alarm by a relay closure. These indications were relayed to the nearest attended site to alert and inform maintenance personnel of the location and nature of the problem.

Between main stations, a complete standby high-frequency line was provided that could be patched and switched into service in place of any working line without interruption to service beyond the relay transit interval of a millisecond or so. This click would not interfere with voice or telegraph signals (high-speed data was not yet born). In a similar manner, at all equipment locations, an individual repeater could be replaced without interrupting service by patching

Fig. 5-18. Hut for Type K carrier unattended auxiliary repeater.

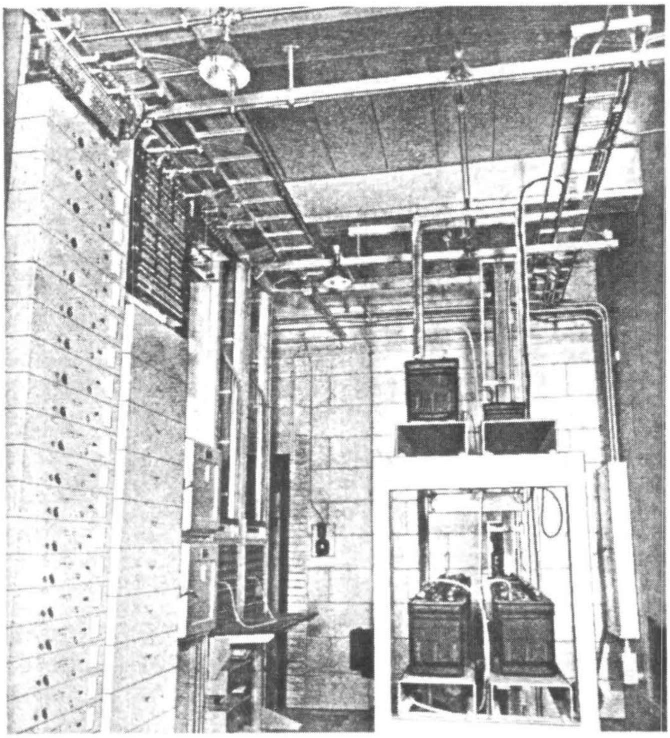


Fig. 5-19. Interior of Type K carrier auxiliary repeater station. *Left*, repeater bays and *right*, in racks, are 150-v DC batteries floating on rectified commercial 60-Hz AC.

and rapid line transfers. This concern with the continuity of service and the uninterrupted flow of all traffic became a hallmark of the design of all transmission facilities. A great deal of ingenuity and considerable investment were devoted to means for coping with failures, maintenance, changes, and growth without affecting service. This approach, moreover, was used at every level, from individual equipment items to multichannel high-frequency lines, and even to entire multisystem facilities such as major cable or radio routes. (If only the planners and operators of our highways and transportation systems would serve their traffic so well!)

IV. TWELVE-CHANNEL CARRIER ON OPEN WIRE—TYPE J CARRIER

In the mid-1930s, as the success of the efforts to develop cable carrier became evident, work was started on a 12-channel carrier system for open wire lines. The motivation was to extend the new techniques to open wire, where, in some areas, even the slow growth of traffic would exceed capacity. It was also recognized that, as 12-channel carrier on cable became widespread, it was going to be desirable to extend the transmission of entire groups over open wire.[10]

Conditions for carrier on open wire were radically different from those in cable, and even though carrier had been first developed on open wire, they were in many respects more difficult. Crosstalk, as always, was a major problem. The shielding provided by separate cables for opposite directions of transmission was not available. As in the early three-channel Type C carrier, equivalent four-wire transmission, in which the opposite directions were transmitted over the same pair at different frequencies, was necessary to control near-end crosstalk. (Equivalent four-wire transmission was not just an unwelcome expedient

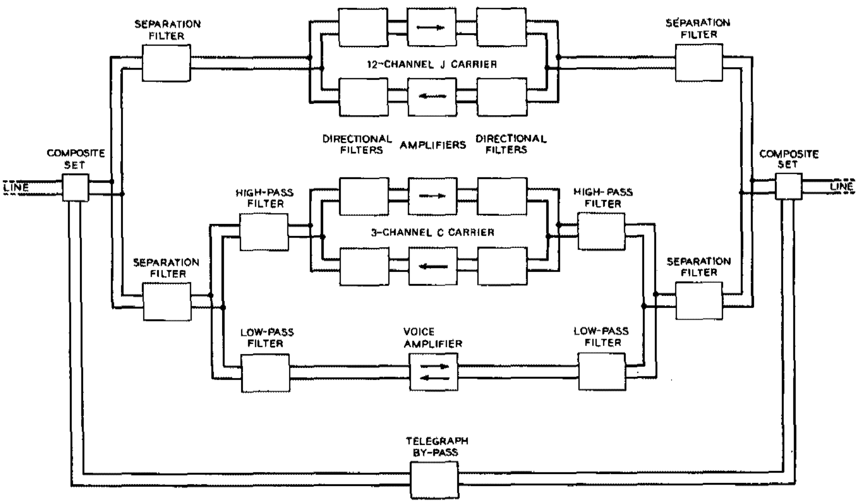


Fig. 5-20. Block schematic of a repeater station for 12-channel Type J carrier, a voice-frequency circuit and telegraph on the same open-wire pair.

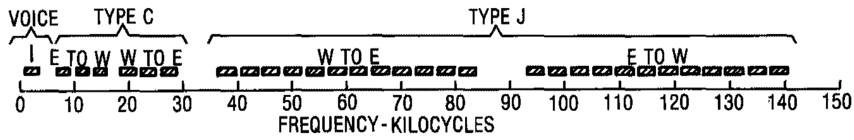


Fig. 5-21. Type J carrier frequency allocations.

to control crosstalk. Pairs on open wire were not as abundant as in cable, and getting as many circuits as possible on a single pair was highly desirable.) Separate bands on the same pair meant higher frequencies, and crosstalk coupling increased with frequency. In addition, there were 60,000 miles of open-wire line already equipped with C carrier. It was not economical to sacrifice the existing carrier channels on those lines. The new carrier system would have to be superimposed at frequencies above voice and C carrier (and telegraph), and coexist with them [Fig. 5-20]. The combination of placing the new system above C carrier and the wide band required for equivalent four-wire transmission resulted in a top frequency of 140 kHz, more than twice the 60 kHz of Type K carrier [Fig. 5-21]. Incidental cables, always necessary and always a problem on open-wire lines, introduced sections of high loss and, unless special measures were taken, presented severe impedance discontinuities, further aggravating the crosstalk problem. Finally, although open wire is a very-low-loss medium under most weather conditions, its loss increases enormously when ice forms on the lines. The loss increase was much greater for the new higher frequencies than at voice frequencies or in the C carrier band [Fig. 5-22].

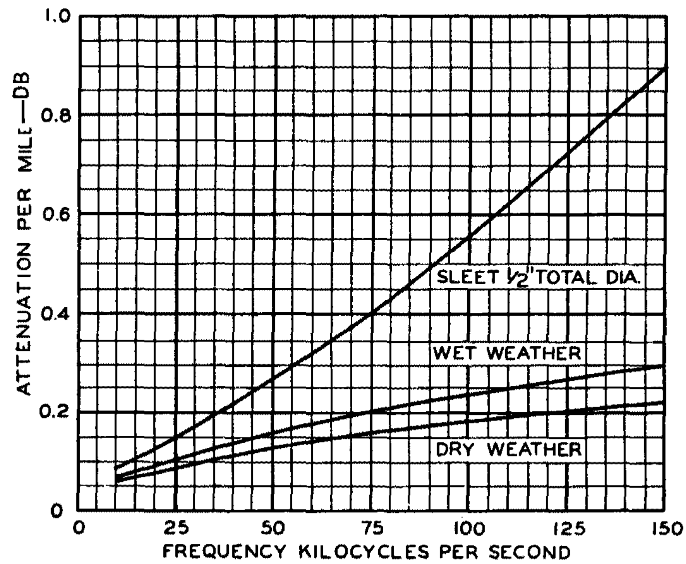
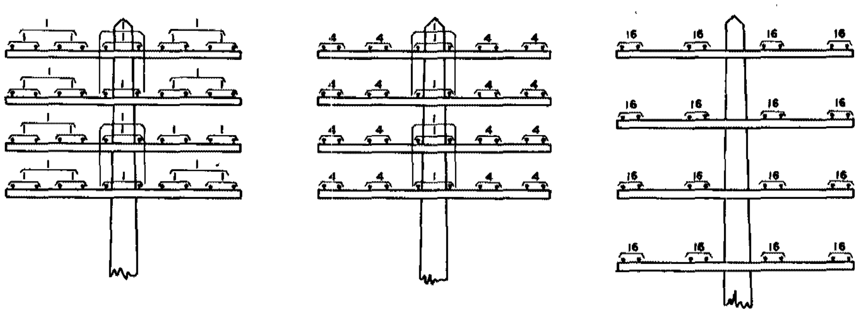


Fig. 5-22. Attenuation frequency characteristics of open-wire lines.

4.1 Crosstalk

Most Type J carrier installations were on lines already transposed for C carrier. On these lines, with 8-inch spacing between the pair wires, and only four pairs per crossarm, it was possible to equip 10 of the 16 pairs with the new system. Special care was necessary, however, even for lines satisfactory for C carrier. Transpositions had to be accurately spaced and line sags carefully equalized, but where pair spacings were already 8 inches, no major structural changes were needed. Where it was desired to apply J carrier to pole lines with 12-inch spacing, it was necessary to reduce some of the spacings to 8 inches, and even to 6 inches, for the carrier pair. On new lines, with better transpositions, 8-inch spacings, 4 pairs per crossarm, and with the spacing between crossarms increased to 36 inches, all 16 pairs could be equipped [Fig. 5-23].

As with other carrier systems, a smooth impedance on the lines was most important. Reflections at impedance discontinuities upset the crosstalk balances and created far-end crosstalk that could not be balanced out. The two principal sources of impedance irregularities were the couplings to other pairs and the inevitable incidental cables at repeaters, entrance links to urban terminals, and such places as river crossings. The transpositions used to reduce crosstalk were adequate to hold the line coupling to a satisfactorily low level, but special cables were needed to meet the second problem. One of these was a disk-insulated spiral-four structure of No. 16 gauge conductors within a shield of copper and steel tapes [Fig. 5-24]. This cable had a loss of 1.5 dB per mile at 140 kHz, and could be loaded to match the impedance of open-wire lines. It was used over short distances to transmit the J carrier band as well as the C carrier and voice signals. Other special cables, and even short lengths of a few unloaded pairs in existing conventional cables, could also be used. In these cases, the J carrier signals were separated from the lower frequencies by system



(a) (b) (c)

Fig. 5-23. Growth in line-carrying capacity. (a) With voice frequency only, capacity was 20 side circuits and 10 phantoms—a total of 30 circuits. (b) With eight-inch spacing and no phantoms on nonpole pairs, and with Type C carrier on all eight-inch-spaced pairs, capacity was 22 voice circuits (including two phantoms) and 48 carrier circuits—a total of 70 circuits. (c) With eight-inch spacing, no pole pairs, and crossarms 36 inches apart (rather than 24 inches), and with voice, Type C, and Type J carrier on all pairs, capacity was 16 voice circuits (no phantoms) and 240 carrier circuits—a total of 256 circuits.

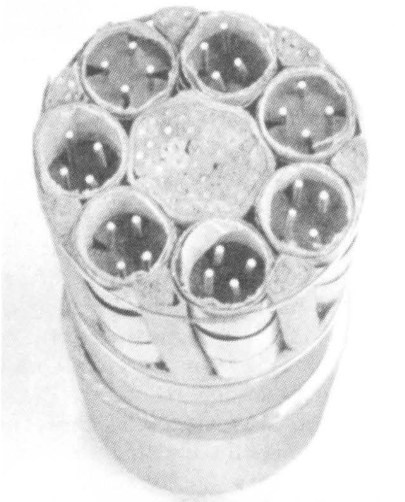


Fig. 5-24. Disk-insulated spiral-four shielded unit. These units were stranded with a number of paper-insulated pairs to form Type J carrier system entrance cables.

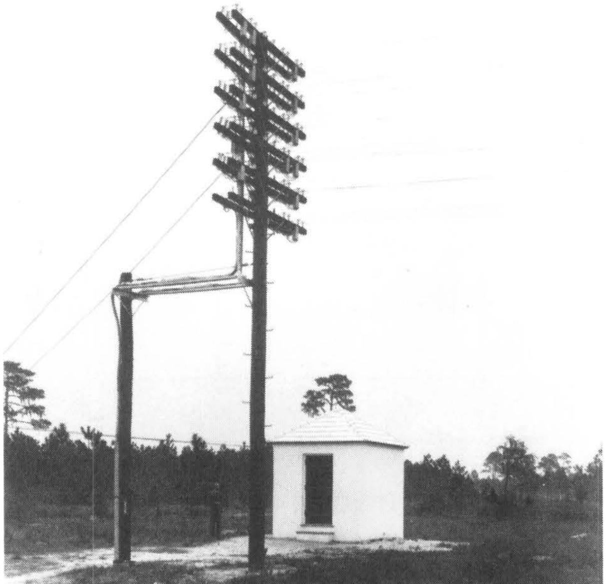


Fig. 5-25. A filter hut used at a Type J carrier system junction of open wire to cable.

separation filters at the junction of the open wire and the cable. The low- and high-frequency systems were then carried on separate cable pairs. The hut used to house the system separation filters was also used for transformers that matched the impedance for open wire and cable, and for crosstalk-balancing arrays similar to those used in K carrier [Fig. 5-25].

4.2 Line Loss and Repeater Gain

The shapes of the attenuation versus frequency of the commonly used open-wire pairs (104, 128, and 165 mils) were very nearly the same. That is, a smaller-gauge line merely looked like a longer length of the larger-gauge circuits. As a result, only one general transmission shape for all gauges was necessary. The repeater was designed to offset and compensate for the loss of the longest section in wet weather. Shorter sections were built out with artificial line networks to equal the loss of the nominal section.

The main line amplifier consisted of three stages. The first two stages were single pentodes. In order to meet the unusual power requirements of the line under adverse weather conditions, the output stage consisted of four power pentodes in parallel [Fig. 5-26]. Full advantage was taken of the newly developed feedback-design methods. (See Chapter 4, Section III.) The feedback circuit of an outer feedback loop provided the basic gain shape to offset the nominal line loss. This loop also included the input and output (hybrid) transformers, whose characteristics thus did not affect the external gain and whose linearity was improved by feedback, as was that of the vacuum tubes. An inner feedback path controlled the out-of-band cutoff and was designed to keep the amplifier stable. About 40 dB of feedback was provided across the entire carrier band.

At the top frequency of J carrier (140 kHz), the normal wet-weather loss

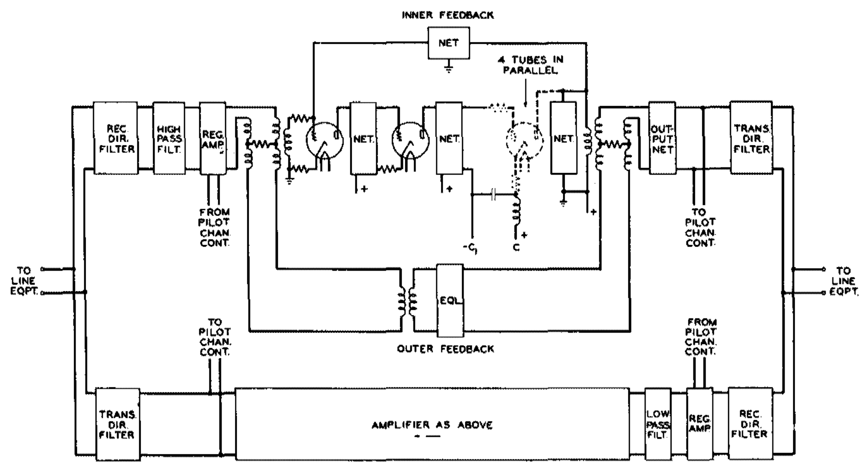


Fig. 5-26. Schematic of Type J carrier line repeater circuit. Four tubes in parallel were used for the high power needed under adverse weather conditions. An outer feedback loop controlled in-band characteristics. The inner feedback path was designed to maintain stability.

was 0.28 dB per mile, about twice that at the top C carrier frequency. This loss would have required repeaters at about one-half of the usual C carrier repeater spacing, that is, every 75 to 100 miles. But with ice, frost, or snow, the wet-weather loss could be greatly exceeded. Even a moderate coating of ice increased the loss of an open-wire line by a factor of three to four; under extreme conditions, with one or two inches of ice, losses of up to 5 dB per mile were observed. Needless to say, in such conditions, transmission (and very likely a good deal more) was lost. Where ice and sleet were common, repeater spacings of 50 miles, or even less, were used. In the initial design, the repeaters had a maximum gain of about 45 dB; later this was increased to 75 dB.

Since the normal dry-weather gain required was only 10 to 25 dB, automatic gain over such a wide range presented an extremely difficult design problem. Gain control was by means of a regulating network and amplifier preceding the main-repeater amplifier. The loss of the regulating network could be varied over a wide range by means of adjustable capacitors, whose position was controlled by a motor drive. A pilot frequency signal was transmitted over the line between the channels in the middle of the band for each direction. This signal was picked off at the repeater output and used to control a motor linked to the gain-control capacitors in the regulating network in such a way as to maintain the pilot at a predetermined reference level (another feedback loop).

Automatic regulation was essential in J carrier because of the rapid changes in line loss. Pilot tones instead of pilot wires were equally essential, as DC resistance did not come close to providing a measure of open-wire line loss under varying weather conditions. Pilot tones for regulation had many other advantages. They provided a direct measure of transmission and could be used for monitoring and alarms as well. The pilot-tone method was used, not only in J carrier, but was later incorporated into C and K carriers as well. In the improved K carrier system (Type K2), in addition to pilot-tone regulation, it was possible to replace the motor-driven variable elements with an all-electronic regulator. The pilot-actuated regulator varied the current that heated a thermally sensitive resistor (thermistor), which in turn controlled the regulating network loss.[11] The open-wire systems, however, could not dispense with the motor drives. In contrast to Type K2 carrier, the later version of open-wire carrier (Type J2) was forced by the problems of ice on the lines to adopt an even more elaborate motor-driven control system.[12] In the Type J2 carrier, additional pilots were transmitted at the edges of the bands in each direction. Those at 84 and 92 kHz were used, via motor-controlled networks, to make a flat gain adjustment in each direction. Pilots at 36 and 143 kHz were used to control a significant tilt (nonflat change in loss) across each band, by a similar mechanism.

4.3 Terminals

In J carrier, K carrier terminal equipment was used to form the basic 60- to 108-kHz group in the normal way. Since the group frequencies on the line, 36 to 84 kHz for the west-to-east groups, and 92 to 140 kHz for the east-to-west groups, both overlapped the basic group band, translation to the line frequencies required two additional stages of modulation [Fig. 5-27]. The group was first

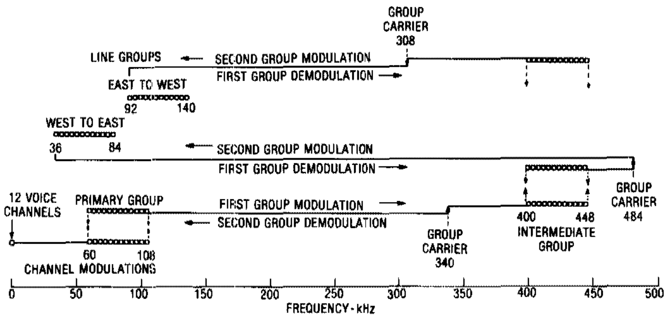


Fig. 5-27. Type J carrier frequency translations. An intermediate group band was required because the line frequencies overlapped the 60- to 108-kHz band of the basic group.

translated to 400 to 448 kHz by a group carrier at 340 kHz. This band was then translated to 36 to 84 kHz for the west-to-east group by a carrier at 484 kHz (the 121st harmonic of the 4-kHz supply) and to 92 to 140 kHz by a 308-kHz carrier for the east-to-west band group. Frequencies as high as 308, 340, and 484 kHz were chosen so that the carriers to shift the groups and undesired modulation products would be well separated from desired products and eliminated by simple filter structures. Amplifiers were inserted at several points in the terminal to change signal levels from those appropriate for modulation to those needed for the next stage of processing. The terminal output transmitting amplifiers for each direction raised the signals to the high levels needed for application to the line.

An entire family of new filters was required for J carrier, in addition to those already needed to generate the basic group. Highly selective filters were used to deliver only pure carrier tones for each modulation step, and the 12-channel band at each stage of modulation was defined and limited by filters. At a repeater, new filters were required to separate the J carrier band from the voice and C carrier circuits, and to separate the directional bands within J carrier.

V. SUMMARY

Type K carrier was a great success by any standard. Transmission was excellent [Fig. 5-28]. The system was installed and used about as fast as it could be produced. The wartime demand for circuits and the pent-up commercial demand following the war produced an explosion in plant capacity. Channel mileage doubled from 1940 to 1947; K carrier furnished about 70 percent of the increase. In the latter year, carrier-derived circuit mileage first exceeded voice-frequency mileage in the toll plant. K carrier accounted for 60 percent of the carrier capacity and one-third of total Bell System toll voice-circuit mileage. It was a remarkably durable system. Installed capacity reached 7 million voice-circuit miles by 1950 and remained above that figure until 1970. Thereafter, it was rapidly retired until, in 1980, only a tiny vestige remained.

J carrier had a less spectacular career; installations ran at about 20 to 25 percent of K carrier capacity. At its peak, about 1950, 1.5 million voice-circuit

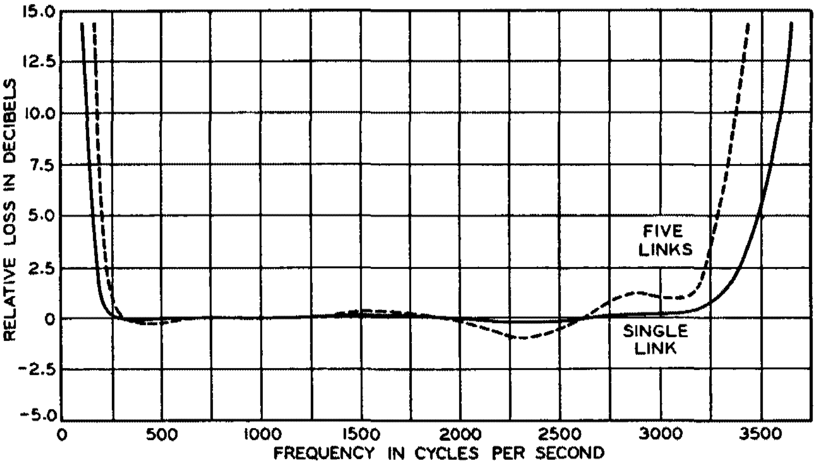


Fig. 5-28. Transmission frequency characteristics of Type K carrier overall circuit.

miles were in service. The reason for its less widespread use, of course, was the preference for and rapid expansion of the cable network, and, in the later stages, the growth of coaxial cable and microwave radio.

It should be noted that these systems (and the many later ones) were never static. A constant stream of improvements was made, and the systems were frequently completely revamped to take advantage of advancing technology. Thus, feedback amplifiers were refitted into the C carrier design, and automatic pilot-controlled regulation fitted into both C and K carriers as these techniques became available. An important development was the standardization of circuit-terminal arrangements. C carrier, for example, although an equivalent four-wire system, was originally regarded as an insert in an otherwise two-wire circuit. The C carrier terminals brought transmission back to a two-wire basis for extension over voice-frequency lines. When J and K carriers were developed, a standardized four-wire terminal arrangement was provided in C carriers as well as in the new systems, providing for ready interconnection at standard levels at a four-wire patch board. Similar standardized arrangements were provided for group interconnection at 60 to 108 kHz, and, in later years, at higher levels in the multiplex hierarchy.

Finally, for systems with an application life of decades, it was usual for two or three major physical reshapings to occur, resulting in large reductions in size, power consumption, and cost, and they were usually accompanied by improvements in performance. These changes were always made part of new production, but were also frequently made available as field kits by means of which existing installed systems could be upgraded.

REFERENCES

1. A. B. Clark and B. W. Kendall, "Carrier in Cable," *Bell Syst. Tech. J.* **12** (July 1933), pp. 251-263.

2. See reference 1, p. 262.
3. See reference 1, p. 263.
4. See reference 1, p. 263.
5. C. W. Green and E. I. Green, "A Carrier Telephone System for Toll Cables," *Bell Syst. Tech. J.* **17** (January 1938), pp. 80-105.
6. R. W. Chesnut, L. M. Ilgenfritz, and A. Kenner, "Cable Carrier Telephone Terminals," *Bell Syst. Tech. J.* **17** (January 1938), pp. 106-124.
7. C. E. Lane, "Crystal Channel Filters for the Cable Carrier System," *Bell Syst. Tech. J.* **17** (January 1938), pp. 125-136.
8. M. A. Weaver, R. S. Tucker, and P. S. Darnell, "Crosstalk and Noise Features of Cable Carrier Telephone System," *Bell Syst. Tech. J.* **17** (January 1938), pp. 137-161.
9. See reference 6.
10. B. W. Kendall and H. A. Affel, "A Twelve-Channel Carrier Telephone System for Open-Wire Lines," *Bell Syst. Tech. J.* **18** (January 1939), pp. 119-142.
11. F. A. Brooks, "The K2 Carrier System," *Bell Lab. Rec.* **27** (April 1949), pp. 141-145.
12. R. S. Caruthers, "Regulation for the J-2 Carrier System," *Bell Lab. Rec.* **19** (November 1940), pp. 74-79.

Chapter 6

Coaxial Cable

I. FIRST CONCEPTS—FIELD AND LABORATORY WORK

1.1 First Concepts

Interest in the coaxial structure, that is, a transmission line consisting of concentric cylindrical conductors, dates from the very early days of electromagnetic theory. The symmetrical circular geometry lent itself to exact analysis. J. C. Maxwell considered its properties, as did Lord Kelvin, O. Heaviside, Lord Rayleigh, and J. J. Thomson. In 1909, A. Russell published an analysis that consolidated much of the earlier work and expressed his results in a form convenient for the calculation of current distribution in the conductors and the resulting loss and impedance as functions of frequency.[1] All the early analyses, including Russell's, were confined to the ideal case of perfect and perfectly concentric conductors.

The first practical use of the coaxial-cable structure was in submarine cable systems. In submarine telegraph lines, the center conductor was encased in a cylindrical insulator with sea water constituting the outer coaxial conductor. Obviously, only the low-frequency properties were of concern in that application. With the advent of radio broadcasting in the 1920s, problems with the open-wire links between the transmitters and their antennas led to a new consideration of the structure. The concept evolved from a cage of wires to confine the field to a continuous tube with a disk-supported center conductor [Fig. 6-1].[2,3] While structures of the latter type were used for high-frequency connections, their use was confined to a relatively narrow band. Emphasis was on low capacitance between the conductors; little consideration was given to their properties as long lines for a very broad band of frequencies.

The initial demonstration of television transmission over open-wire lines and radio in 1927 was limited to a band of only 20 or 30 kHz. It was generally recognized, however, even at that early date, that much wider bands, probably extending to 1 MHz or more, were going to be needed for higher-definition television. In Bell Laboratories, attention was directed almost immediately to the coaxial cable as a possible transmission medium. In a series of unpublished memoranda in 1927 and 1928, E. I. Green considered possible practical structures and analyzed their transmission and impedance characteristics, while R. G. McCurdy developed early results for their shielding and crosstalk properties.

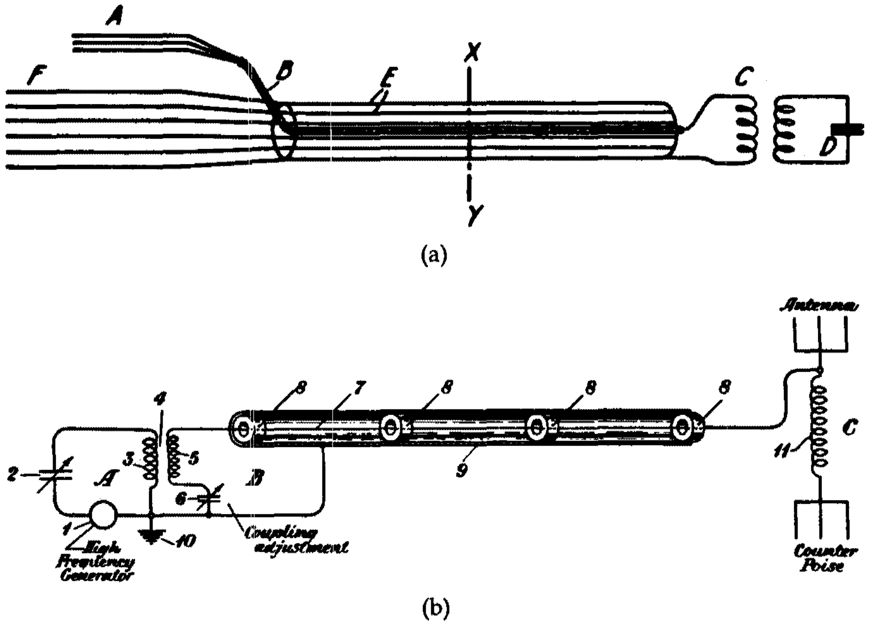


Fig. 6-1. Antenna feed concepts. (a) Shield of surrounding wires in patent filed by H. J. Round of RCA in December 1921. (b) Outer conductor as a continuous tube with disk-supported inner conductor. Patent filed by R. S. Ohl of AT&T, December 1924.

An interesting property of the coaxial structure was recognized at that time. It was obvious that, in general, the larger the conductors, the lower the loss. However, for a given inner diameter of the outer tube, there is an optimum diameter for the inner conductor. In the radio-antenna field application, the emphasis was on keeping the inner-to-outer conductor capacitance low, which favored a small-diameter inner conductor. But for very small diameters, the loss will be very high because of the high resistance. On the other hand, as the diameter of the inner conductor is increased, lowering the resistance, the increasing capacitance will also ultimately result in very high loss. It is evident that there is an optimum between high loss due to the resistance of a small inner conductor and high loss due to the high capacitance of a large one. In England in 1928, C. S. Franklin derived this optimum, arriving at the ratio of 3.6 to 1 for the conductor diameters in coaxials of any size.[4]

If a coaxial cable could be used as a broadband medium for television, it was naturally also a candidate for carrier telephony. In May of 1929, H. A. Affel and L. Espenschied filed a patent embodying their concept of a multichannel carrier telephone communications system using coaxial cable as the transmission medium [Fig. 6-2].[5]

The potential attractions of the coaxial cable were apparent to anyone concerned with the problem of long-distance multichannel carrier transmission. Like all cables, it could be completely protected from the external environment.

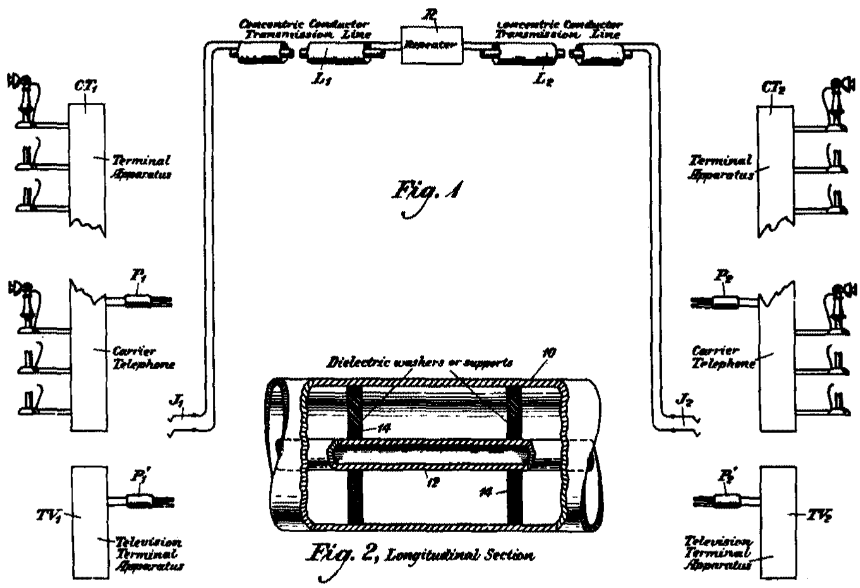


Fig. 6-2. Coaxial-cable carrier telephone system. Patent filed by L. Espenschied and H. A. Affel, May 1929.

The loss could be made as low as desired by using suitably large tubes for the conductors. No loading was required, as was necessary with lower-frequency transmission over wire paths. Perhaps most important, the cable would be self-shielding. Crosstalk coupling would decrease with frequency and could be as low as desired, permitting multiple coaxial tubes to be placed in close proximity in an outer protective sheath. The absence of any coupling to other lines and the continuous structure of a single type would provide a loss that varied smoothly with frequency and impedance that changed hardly at all. In sum, it appeared that here, at last, was a medium where the transmission line would not limit the bandwidth or the number of channels that could be transmitted. Only the terminals and the amplifiers of the repeaters would limit the possibilities. It remained to be seen, of course, whether the structure could be built with materials and methods suitable for large-scale manufacture and whether a practical structure in a realistic field installation would indeed have the desired properties.

1.2 The Phoenixville Tests

In late 1929, the Development and Research (D&R) Department at AT&T headquarters, working with Bell Laboratories, set out to confirm the properties of the coaxial cable as a telephone-transmission medium. As noted previously, formulas were available for loss and other properties, but the physicists and mathematicians of the classical era had dealt only with idealized structures. They did not deal with the inevitable imperfections or translate their analyses to the language of engineers. The D&R people decided, in addition to continuing analysis and laboratory work, to hold a field test to confirm the theoretical properties. This test was held at Phoenixville, Pennsylvania, where a tract that had been used for other field tests was available.

A special line was constructed consisting of a double coaxial [Fig. 6-3].[6] The outer conductor was a 2.5-inch copper tube within which a five-eighth inch outside-diameter tube was suspended. This intermediate tube, in turn, contained a smaller copper wire. This structure thus made two coaxial lines of different size available, with the intermediate tube as a common conductor between them. The choice of tube sizes was such that each line approximated the 3.6-to-1 diameter ratio for minimum loss. As the conductors were fairly rigid, the supporting spacers were well separated so that the dielectric was almost entirely air. Two of these lines, each 2600 feet long, were suspended on wooden fixtures

(a)

(b)

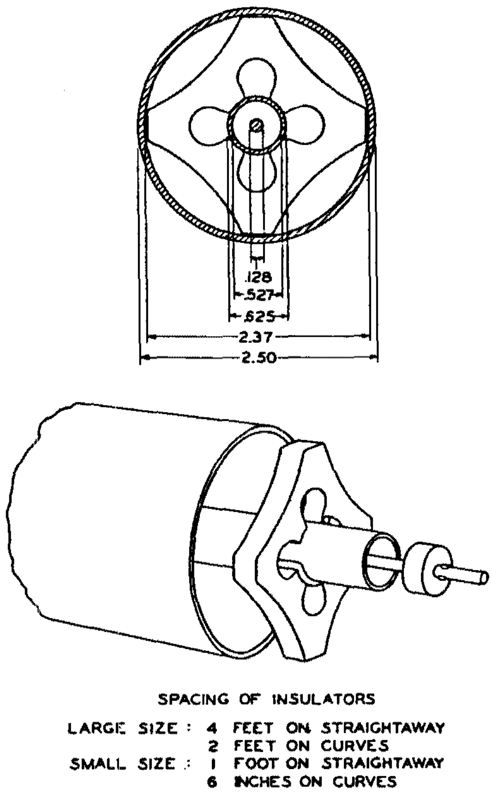


Fig. 6-3. Coaxial structure used in the Phoenixville installation. (a) Cross section. (b) Spacing of insulators. A double coaxial was built, in which the outer conductor of the inner coaxial of approximately five-eighths inch in outer diameter, served as the inner conductor of the larger coaxial, with an outer diameter of 2-1/2 inches.

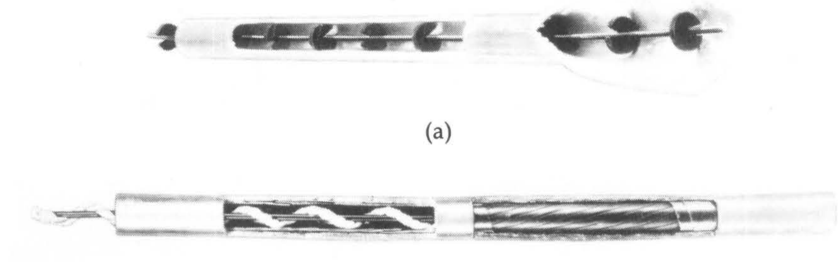
Fig. 6-4. Phoenixville installation, showing coaxial conductors entering test house.

around the test site, with their ends brought in to a test hut for access and measurement [Fig. 6-4].

The tests at Phoenixville continued for about a year. Extensive measurements were made at frequencies from 0.1 to 10.0 MHz. Transmission loss was found to conform closely to the theory with remarkably little effect from imperfections due to joints and departures from perfect concentricity. Crosstalk coupling also agreed closely with theoretical values. The common intermediate tube was included specifically to check the shielding expected from skin effect. At frequencies above 100 to 150 kHz the induced crosstalk was below the measuring capability and below thermal noise, even with disturbing signals much greater than would be encountered in practice. The tests confirmed that lines constructed with reasonable tolerances would behave in every respect as anticipated.

1.3 Laboratory Work and Analysis

Upon completion of the Phoenixville field test, the focus shifted to the laboratory where work was under way to determine the parameters of a practical system. This effort included not only the design of a practical cable structure that could be manufactured and installed, but also the wideband electronics required. Many forms of coaxial cable were built and tested [Fig. 6-5]. Size was a special consideration. Large diameters would provide very low loss and perhaps even permit the use of less expensive material, such as lead, for the outer conductor. The techniques for forming long seamless tubes of lead were well established and would require less factory preparation and cost for the new cable. (Low-loss coaxial cables, two inches or so in diameter, were still a topic for economic studies as late as the mid-1930s.) Various forms of shielded pairs were also considered. The virtues of their balance with respect to ground continued to be appealing, especially at low frequencies. In the end, the simplicity



(a)

(b)

Fig. 6-5. Experimental coaxial structures of the early 1930s. (a) Coaxial in lead tube with rubber-disk insulators. (b) Small flexible coaxial with spiral-string insulating support.

and excellent high-frequency properties of the coaxial structure won out for long-haul transmission. But much smaller forms than the Phoenixville trial type were necessary in order that the cable could be manufactured continuously in large quantities, stored on reels, delivered to a site, and pulled into cable ducts. Interest began to focus on tubes of 0.5 inch or less in diameter. Experimental lengths of various kinds were fabricated in the basement of Bell Laboratories on improvised equipment [Fig. 6-6]. The cable finally adopted for further field tests had an outer conductor of only 0.27-inch inner diameter. The loss of the small coaxial structure was large compared to the loss of open wire, but the perfect shielding made it possible to allow the signals to sink to very low levels and to use very-high-gain amplifiers [Fig. 6-7].

At the same time as the experimental structures were being built and tested, analysis of the coaxial geometry continued. S. A. Schelkunoff of the mathematics research group at Bell Laboratories, building on earlier work of J. R. Carson at AT&T, extended the analysis from the idealized structures to the practical forms then being realized.[7] He further wished to adapt the rather formidable mathematics of the early analysts to engineering uses and translate it into the concepts and language of electric-circuit theory. Like his predecessors, he found the investigation made relatively easy, as the conductor shapes fit perfectly into the cylindrical system of coordinates, thereby making it feasible to carry out a rigorous discussion on the basis of electromagnetic theory. From his initial results, he then introduced the effects of eccentricity and other departures from the ideal, and deduced the effects on loss, impedance, and the other properties of interest. Fortunately, it turned out that the results obtained by field theory could be expressed readily in the familiar language of circuit theory, thereby gaining all the simplicity of the latter combined with the accuracy of the former.

The characteristic impedance, Z, and attenuation, α, of a coaxial circuit are given by the formulas: $Z = \sqrt{L/C}$ and $\alpha = (R/2)\sqrt{C/L} + (G/2)\sqrt{L/C}$, where R, L, C, and G are the so-called primary constants, that is, the resistance, inductance, capacitance, and conductance per unit length. The first term in the attenuation formula is the loss in the conductors; the second is the loss in the dielectric. If the dielectric losses are small, as would be the case with a well-

Fig. 6-6. J. F. Wentz making experimental coaxial conductors on an improvised stranding machine at Bell Laboratories.

constructed cable, the second term is negligible and the loss is given by the first term alone. Attenuation increases as the square root of frequency, as a result of the increase in resistance caused by the skin effect. The combination of the variation of loss with frequency and coaxial diameter led to two major conclusions. First, for a given size of conductor and length of line, the available bandwidth increases roughly as the square of the number of repeater points; and, second, for a given repeater spacing, the bandwidth increases approximately as the square of the coaxial diameter. The validity of the first rule was extensively tested over several decades in the successive series of coaxial-system designs. For a number of practical reasons, the increase of bandwidth with shorter repeater spacing, while substantial (the Type L5 coaxial-cable system has a 60-MHz band), did not quite keep pace with the "square law." As for the second rule, while larger diameters were often considered, only one moderate increase in size was made shortly after World War II.

The velocity of propagation in a coaxial cable is given by the formula: $v = 1/(\sqrt{LC})$, where L and C are the same unit inductance and capacitance used in the transmission equation. In an ideal coaxial, with very thin conductor walls and only air for the dielectric, the velocity would equal the velocity of light at all frequencies. In the early structure, the insulators added about 13 percent to the capacitance and the finite conductor thicknesses added from 2 to 14 percent

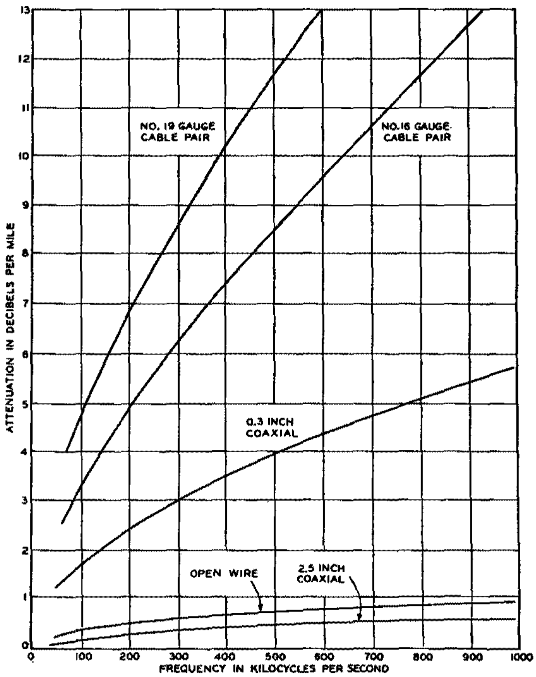


Fig. 6-7. Attenuation-frequency characteristics of coaxial cable and other circuits.

to the inductance. Nevertheless, the actual velocity was quite high and fairly uniform with frequency [Fig. 6-8]. The high velocity was advantageous for telephony, compared to the well-known problems of slow-speed loaded circuits

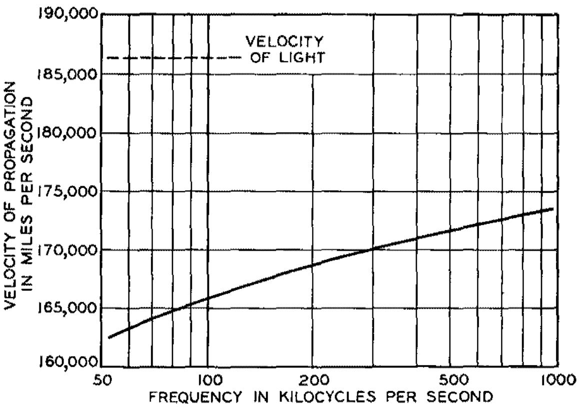


Fig. 6-8. Velocity-frequency characteristic for the coaxial structure.

and the relatively small change with frequency minimized the delay equalization required for high-definition television. The characteristic impedance of a coaxial line of optimum-diameter ratio was close to 75 Ω, almost purely resistive, and essentially constant with frequency.

1.4 System Design

By 1933, the experimental and analytical work established that an excellent medium with well-understood properties was available. At the same time, work had been going forward on the remaining components needed for a complete system. Planning was based on coaxials of one-half inch diameter or less, with repeaters at about ten-mile intervals and a bandwidth of 1 MHz.[8] The band was to be usable for 240 telephone channels, multiplexed by suitable terminals, or used in its entirety for television. The repeater amplifier would have to provide gain to match the line-loss shape (equalization) and compensate for changes in loss due to temperature (regulation); it had to be extremely stable and linear as well.

In the concept that was emerging, a 4000-mile system required 400 amplifiers in tandem. As the end-to-end requirements were not relaxed, but in fact were becoming more exacting, the requirements on each amplifier for equalization, regulation, stability, and linearity were far tougher than anything previously attempted. The line amplifier became, and remained through several generations of design, the critical component limiting what could be done in coaxial systems.

1.5 Repeaters

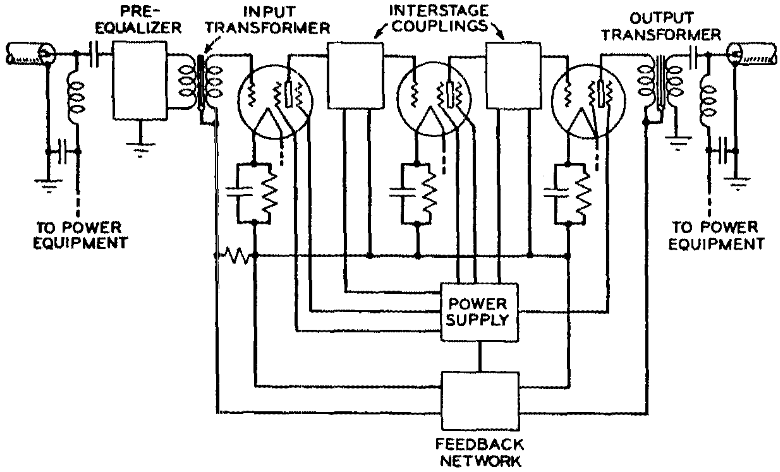
With effective shielding from all other signals, the channel signals could be attenuated to the limit set by thermal and tube noise and amplified to levels limited by the power capacity and linearity of the repeaters. Preliminary analysis indicated that at 1 MHz the lower-level limit was about 55 dB below the transmitting voice-frequency testboard level and the upper about 5 dB above that level. What was needed, therefore, was an amplifier with about 60-dB gain, shaped to match the line loss over a 1-MHz band, with as much feedback as possible. The limited shielding of the coaxial cable at low frequencies, as well as the problem of designing an amplifier suitable over many octaves of frequency, limited the lowest usable frequency to about 50 kHz. An amplifier for the entire band was a difficult design problem. Components of the size and with the properties required for the low end were difficult to control at frequencies 20 times higher, and, as known from feedback theory, had to be manageable up to much higher frequencies. Some consideration was given and experiments actually carried out on splitting the band and transmitting the portions separately. This approach was, apparently, more to ease problems anticipated with terminals for the television signal than to simplify the repeater, but, in any case, was abandoned at an early stage.

High-gain amplifiers at radio frequencies had been built for many years for broadcast radio. However, these amplifiers achieved their gain by sharply tuned high-impedance circuits over relatively narrow bands. The large parasitic capacitance of the tubes and other components made it impossible to maintain a high impedance over the wide coaxial band. Nor was it possible, within the constraints imposed by feedback stability, to make up for this by additional stages. A great deal of emphasis was placed on new high-gain, low-capacitance (high-figure-of-merit) vacuum tubes.

The actual gain required varied from about 15 dB at the low frequencies to 60 dB at the top end of the band. There was much experimenting with equalization. The shape could be achieved by an equalizer preceding an otherwise flat amplifier, or the compensation could be effected by a combination of shaped equalizer and shaped gain in the amplifier. A combination was sought that would match the line loss to the desired precision without playing havoc with the signal-to-noise performance.

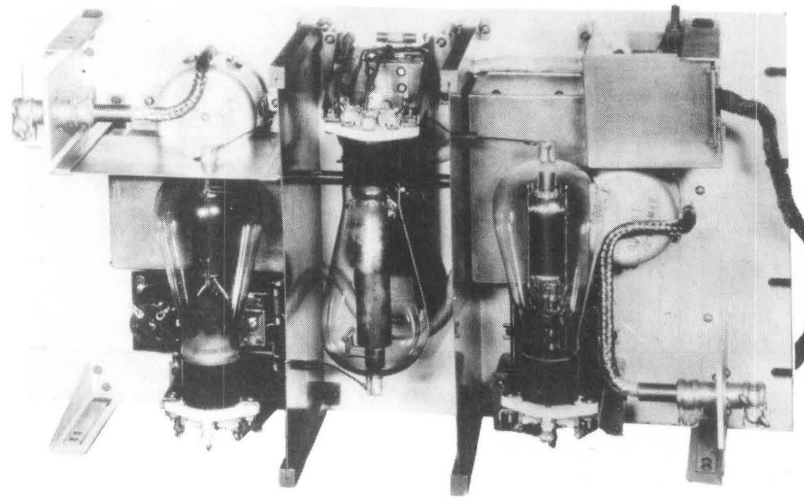
Many experimental amplifiers were built and tested with highly encouraging results [Fig. 6-9]. Developers explored the range well beyond the initial objectives, building amplifiers of both feedback and nonfeedback types to frequencies as high as 5 MHz. This may seem strange in view of the difficulties encountered in meeting the basic objectives, but, even before the first repeatered line was built, broader-band systems were being projected, especially for television. The most promising lines of attack had to be investigated.

During the laboratory experimental stage techniques were also tested for temperature compensation by pilot-actuated regulators with considerable success. Two pilots, one near each end of the band, were felt to be necessary. The low-frequency pilot could be used for overall signal level adjustment, and the high-end pilot used to control a tilt equalizer that would compensate for the varying loss across the band due to temperature changes.

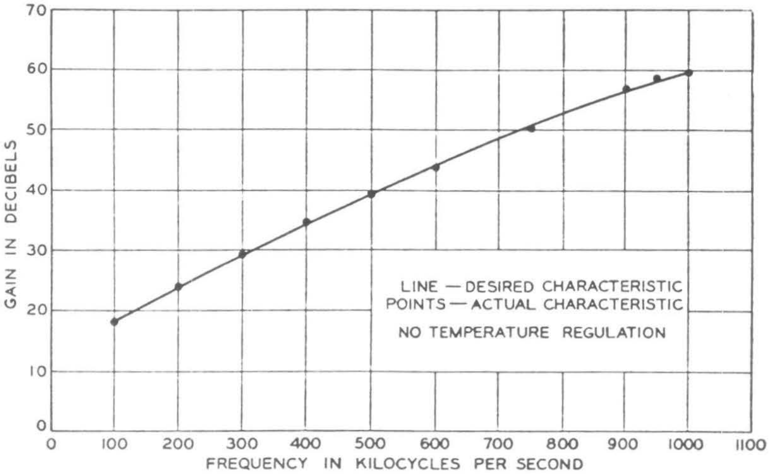


(a)

Fig. 6-9. Experimental coaxial cable amplifiers. (a) Circuit of 1-MHz three-stage feedback amplifier.



(b)



(c)

Fig. 6–9. (b) An early 1-MHz amplifier. (c) Gain of 1-MHz amplifier compared with coaxial line characteristic.

1.6 Terminals for 240 Telephone Channels

The multiplex equipment to form a basic group of 12 channels in the range of 60 to 108 kHz was intended to be a common terminal element for carrier systems on open wire and coaxial cable as well as on paired cable. (See Chapter 5, Section III.) This was, therefore, adopted without modification. To shift the

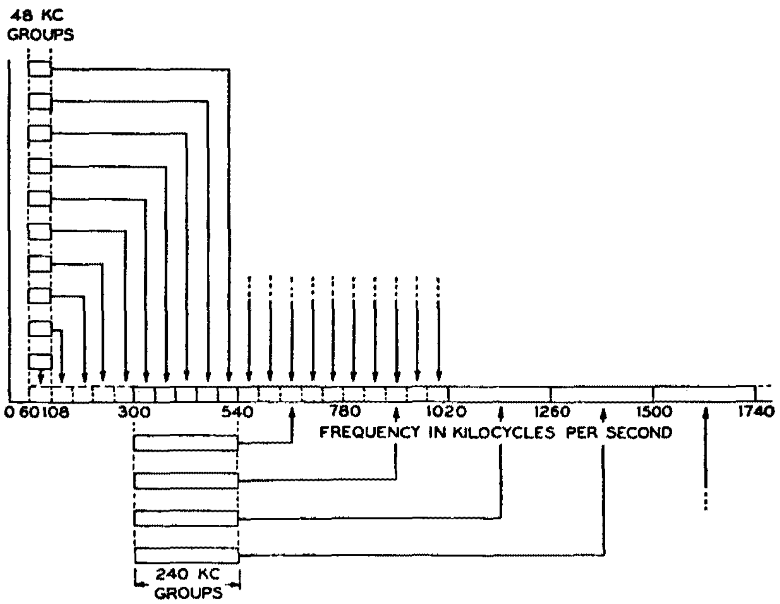


Fig. 6-10. Frequency translation options to form the 240-channel line signal of the experimental coaxial system: *top,* direct translation of twenty 12-channel groups; *bottom,* translation via a 60-channel supergroup.

basic group to the line frequency, designers considered two options [Fig. 6-10]. In the first option, the first 12-channel group was connected directly to the line and up to 19 additional groups were stacked side by side by a second stage of modulation with appropriately selected carriers. In the second option, a supergroup of 60 channels was formed in the range from 300 to 540 kHz by the same method used in the first option. This supergroup could then be shifted as a unit and placed on the line in a third stage of modulation. The supergroup approach required fewer different modulator designs and made the 60-channel block available as an entity, but also involved the greater complexity of a three-stage multiplex. At the experimental stage, the two-stage approach was favored, since fully equipped systems were not needed in any case.

II. THE 1936 NEW YORK-PHILADELPHIA TRIAL

By late 1933, enough progress had been made in the laboratory to consider a system field trial. During 1934, plans were further developed and a decision was made to install about 100 miles of factory-made coaxial cable between New York and Philadelphia. The length was chosen to be long enough to permit extrapolation of the performance to lines of 1000 miles or more with confidence. An important objective was that the equipment should be operable and maintained by regular operating-plant personnel. The design, while still experimental, was to be close enough to a commercial product to permit reliable cost estimates. In early 1935, a petition for the construction was submitted to the newly established Federal Communications Commission and granted in due course. Construction was started in March and completed by November 1936. The chosen route ran between terminals in the Long Lines building in downtown New York City to the Bourse Building in Philadelphia, a distance of 94.5 miles [Fig. 6-11].[9] Main stations with repeaters and power feeds for the line were established in typical attended offices at New York, Newark, Princeton, and Philadelphia [Fig. 6-12]. Unattended auxiliary repeaters in watertight cases were placed in manholes or in simple above-ground enclosures at about ten-mile intervals between the main stations [Fig. 6-13].

The cable for the field trial consisted of two coaxial tubes, one for each direction of transmission, and a number of ordinary pairs for order wire and alarm transmission from intermediate points. Each coaxial tube was formed with an outer conductor formed of overlapping specially formed copper tapes. The tapes were wrapped in a helical lay around hard rubber disks spaced at three-quarter inch intervals, supporting a No. 13 gauge (.072-inch diameter) solid-copper-center conductor. The tube was then wrapped in an outer layer of two steel tapes for added physical strength, protection, and additional shielding [Fig. 6-14].

This structure presented no small challenge to the cable designers and fabricators.[10] It was essential to preserve the low loss and smooth transmission

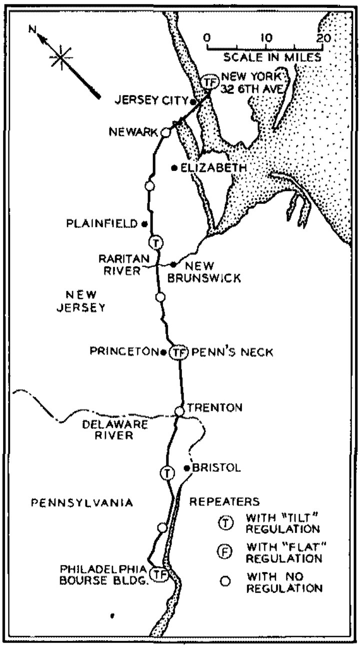


Fig. 6-11. Map of the experimental-coaxial-line route showing repeater locations.

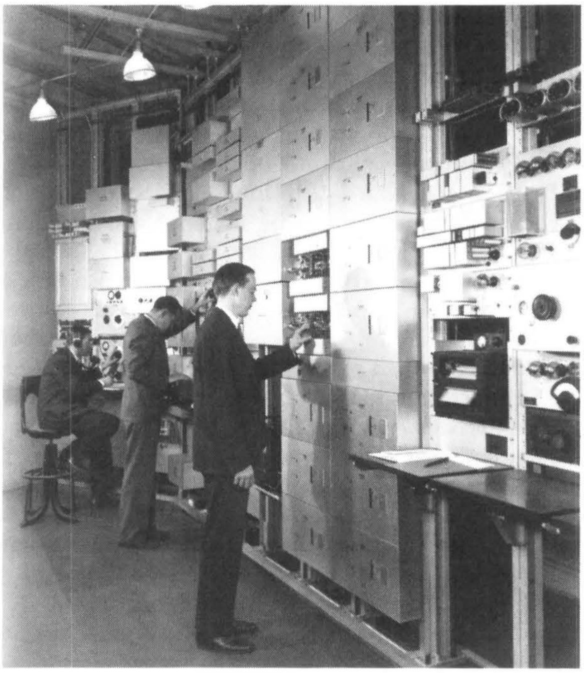


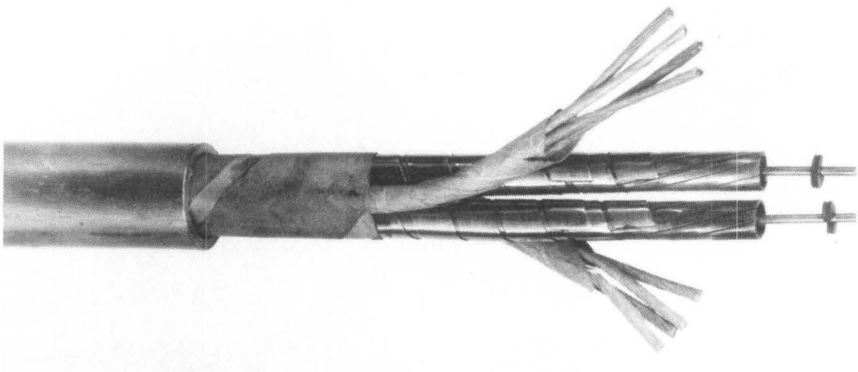
Fig. 6-12. The New York terminal of the 1936 experimental coaxial system.

properties of the coaxial structure. At the same time, the cable had to be in a form that could be manufactured continuously in sections of considerable length and that would be flexible and rugged enough to withstand reeling and installation in a reasonably standard way. A typical problem, for example, was the design of the rubber supporting disks. They had to have low dielectric loss, and it was desirable to make them as thin as possible to keep the dielectric mostly air. They had to be strong enough to support the center conductor during flexing but, at the same time, elastic enough to be snapped over the center conductor without nicking the wire or cracking the disk. Finally, they had to be able to withstand the heat associated with the application of the outer lead sheath. The solution was found in a new, improved, hard rubber compound formulated specifically for this application by the Bell Laboratories chemical research group. Two coaxials and two No. 19 gauge quads were enclosed in a lead sheath with an outer diameter of seven-eighths of an inch. The cable was manufactured by Western Electric at the Baltimore Works and delivered on reels in 1500-foot lengths. These lengths were installed in ducts in the normal way, the small size and light weight of the cable making it possible to pull in as much as 1500 feet, compared to the 500 feet between splicing manholes that was the normal maximum length for a full-sized cable. The long reel lengths provided a desirable reduction in the number of splices.

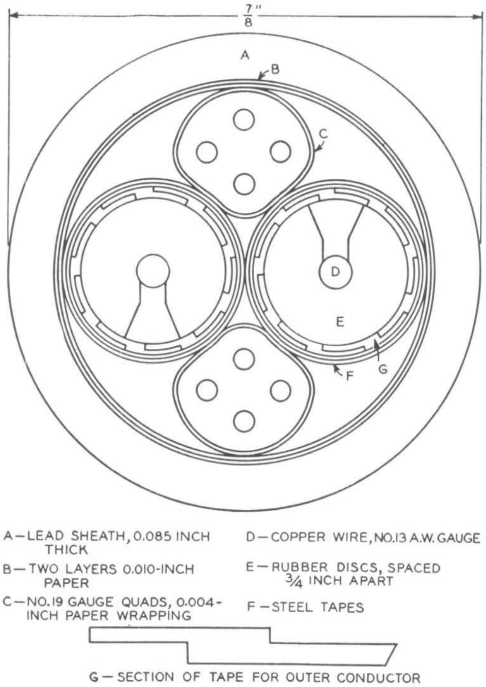
Fig. 6-13. Lowering a repeater into a manhole.

The telephone multiplex terminals for the trial were of the two-stage modulation type; basic 60- to 108-kHz groups were shifted directly to the assigned line frequency in the range of 60 to 1020 kHz. These terminals were only partially equipped, as all important properties could be adequately tested without incurring the expense of equipping all 20 twelve-channel groups [Fig. 6-15]. Three channel banks provided 36 channel terminations, but provision of six group modems made it possible to place the groups at a choice of locations over the entire range of line frequencies. Appropriate patching at the terminals made it possible to traverse the line many times by looping to form a very long single telephone circuit.

The trial repeaters used three-stage amplifiers with feedback around the last two stages and a pilot-controlled regulating network between the first and second stage.[11] Flat regulation, controlled by a 60-kHz pilot, was provided at New York, Princeton, and Philadelphia, while tilt regulation, to correct for change in cable loss due to temperature changes, was provided at those stations and at two intermediate points in addition [Fig. 6-16]. Power for the unattended auxiliary repeaters was fed over the coaxials from the stations at Newark, Princeton, and Philadelphia. The 60-Hz voltage was applied between the center conductors of the two cables, the outer conductor of each being grounded. At



(a)



(b)

Fig. 6-14. Experimental coaxial cable for the 1936 New York-Philadelphia trial. (a) Photograph of cable. (b) Detail of cross section.

each repeater, the AC was rectified and the required DC voltage furnished to the amplifiers. The wide separation in frequency between the 60-Hz power currents and the signal components at frequencies of 60 kHz or higher made it possible to split them by simple power separation filters at each repeater.

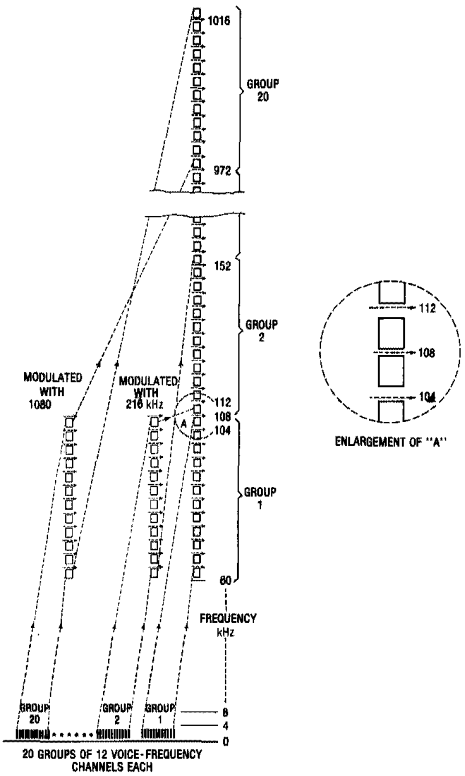
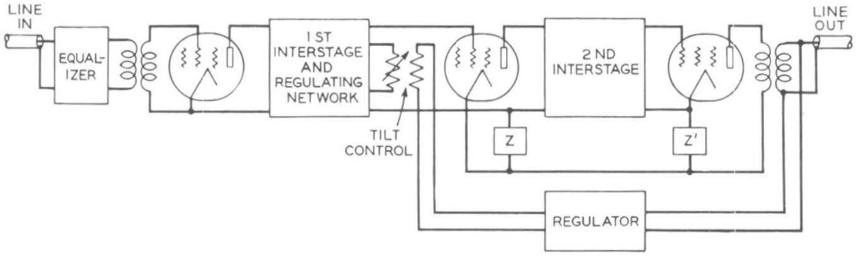


Fig. 6-15. Frequency allocation and translations for 1936 New York–Philadelphia coaxial system trial. Only three 12-channel groups were completely equipped, but these could be transmitted anywhere in the 1-MHz line spectrum.

Tests were started and the first conversations carried shortly after the line was powered up. Circuits were established through the cable and extended to Washington over other facilities for demonstration calls. Patching at the voice-frequency terminals of the channel modems allowed loops to be set up that traversed the system 20 times, for a total channel length of 3800 miles. In each direction, the conversation passed through 70 stages of frequency shift in the terminal modems, far more than would be experienced in any long working system. In April 1937, 27 circuits were established through the cable, 16 to Philadelphia and 11 to points farther south, for a service trial. The operation was continued for several weeks during which carrier telegraph and telephoto were transmitted as well as speech.

The talking and transmission tests and many others indicated that the system basically met the performance projected, although, as might be expected, some weaknesses were revealed, especially in equalization and temperature regu-



(a)

(b)

Fig. 6-16. The New York–Philadelphia trial amplifier. (a) Schematic of pilot-controlled regulating amplifier for experimental coaxial system. (b) Installed regulating amplifier in a manhole. The three-stage amplifier is made up of the vacuum tubes in the top compartment. The tubes in the lower section are part of the regulating network.

lation. The tests pointed to areas where additional development was needed before a fully commercial system could be installed. In general, the troubles were minor, and the trial was terminated with the developers confident they were well on the way to the practical application of a new and exciting technology.

III. TELEVISION AND THE COAXIAL CABLE

Interest in the electrical transmission of images, both stationary and moving, is as old as the telephone. Discovery of the photoresistive effect in selenium generated rudimentary proposals as early as 1875. The principle of scanning a picture line by line with a rapid succession of frames, relying on the persistence of vision to create the illusion of continuous motion, was proposed around 1880. In 1884 in Germany, P. Nipkow patented a system incorporating a rotating perforated disk to accomplish the scanning. K. F. Braun improved the cathode-ray tube around 1900, making it a potential line-by-line picture generator. By 1907, surprisingly modern concepts were put forth by B. Rosing in Russia and A. A. Campbell-Swinton in England. Their proposals included sequential scanning of a picture, transmission of electrical signals that were analogs of the picture-element light values, and image displays by means of a cathode-ray tube. There is little evidence that any of the ideas were reduced to even laboratory demonstration. Realization, unfortunately, was far beyond the technology of that early time.

The invention and perfection of the vacuum tube and the advent and rapid growth of radio broadcasting in the 1920s created a renewed interest and a powerful stimulus to inventors in several places. By the mid-1920s, many experimenters in Europe and in the United States were attempting to transmit images by electrical signals.[12] In the United States, C. F. Jenkins, a contributor to motion picture and telephotography development, transmitted and recreated simple animated figures. Credit for the first achievement of true television appears to belong to J. L. Baird in England, who succeeded in 1926 in generating, transmitting, and recreating recognizable images with tonal gradation. Most of the early efforts used rotating disks, or their equivalent, to scan the image. V. K. Zworykin, a pupil of Rosing, persisted in efforts to realize an all-electronic method and patented a precursor to his later, highly successful, iconoscope in 1923.[13] The 1927 demonstration by Bell Laboratories used a rotating disk to scan and neon glow lamps to recreate the image. The Bell Laboratories tests featured, in addition, transmitting the signal over open-wire lines between New York City and Washington DC, and by radio from Whippany, New Jersey to New York City. The image was scanned in 50 lines at a 16-frame rate, generating a signal in a 20-to-30-kHz band.[14]

In all these early efforts, the definition, by any standard, was poor. Pictures composed of only 50 scanning lines were recognizable, but hardly sharp. Adequate synchronization between transmitter and receiver was a problem. The images generated by Bell Laboratories, while judged by one observer to be "fairly clear," were further described as "comparable to faded daguerreotypes held in a shaky hand."

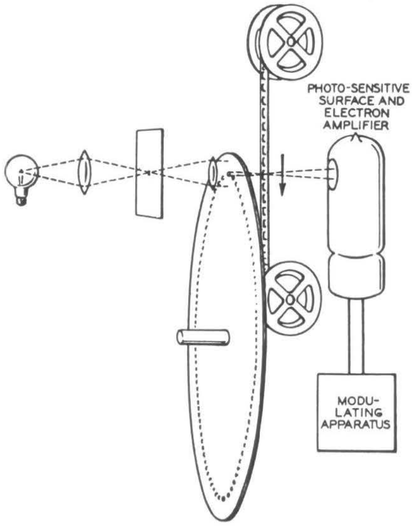
Nevertheless, vigorous work continued in several laboratories. For some time, the work in Bell Laboratories continued to focus on television as a two-way, face-to-face adjunct to telephony. Rudimentary color transmission was demonstrated in 1929. By 1930, an improved two-way system with better definition was in experimental use between Bell Laboratories and an AT&T location in Manhattan. Broadcasters became interested, and, by 1932, a dozen or more stations in the United States were licensed for experimental broadcast transmission and several, both in the United States and in Europe, were making occasional transmissions using the mechanical systems available. The really significant advance, however, was in the work toward an all-electronic system by P. T. Farnsworth in his San Francisco laboratory and, especially, by Zworykin at Radio Corporation of America (RCA). RCA demonstrated a working iconoscope in 1932 and, by 1934, generated pictures of 343 lines at a rate of 30 frames per second and displayed them on a cathode-ray-tube kinescope, completely eliminating the scanning wheels and all other moving parts.[15]

In 1932, the growing number of radio stations showing an interest and the success of the all-electronic system with its much greater bandwidth led Bell Laboratories to reconsider its objectives. The emphasis changed from two-way, telephony-adjunct systems with bandwidth in the range of 50 to 100 kHz to work on circuits for broadband television networking. As noted in contemporary conference records, the circuits were to be "capable of transmitting television programs having sufficient resolution to be of real entertainment value." The frequency band necessary was "considered to be not less than 300,000 cycles and probably not greater than 1,000,000." The objective was to furnish channels for television networking in a manner analogous to the Bell System's established role in furnishing audio program channels for the radio networks, with coaxial cable as the likely transmission medium. In 1932, NBC discussed with AT&T the prospects for furnishing wideband television networking by coaxial cable. A demonstration of wideband television transmission became one of the major objectives of the New York–Philadelphia tests.

3.1 The Transmission Problem

In addition to requiring a very wide band, the television signal had several other properties that posed new problems for the transmission-system designers. The low-frequency components in particular presented difficulties. Broadband line repeaters could be built to amplify only a limited range of frequencies. In addition to a limit on the highest frequencies that could be transmitted, the ratio of the lowest to highest frequencies could not be excessive. Delay equalization, i.e., having all the signal components arrive in the correct phase relationship, which is essential for television, became especially difficult for a wide range of frequencies. Finally, in contrast to its excellent shielding properties at high frequencies, coaxial cable, with its unbalanced structure, is a very poor shielding medium at low frequency. Even short lengths were not satisfactory for the baseband video signal. For all these reasons, it was necessary to shift the television signal to higher frequencies for transmission over the line. At the same time, the required high frequencies stressed the amplifier-design capabilities. A compromise was reached, making the lowest frequencies about 100 kHz.[16]

A picture generator was built at Bell Laboratories consisting of a scanning disk six feet in diameter, providing 240 lines at 24 frames per second [Fig. 6-17]. This arrangement gave fairly good resolution, and, while the flicker was perceptible, it was adequate for the experiment. The number of lines, of course, determined the vertical resolution. For equal horizontal resolution, in a picture with a horizontal-to-vertical ratio of seven to six, an elementary analysis of



(a)

(b)

Fig. 6-17. The mechanical scanning disk for the 1937 demonstration of television transmission over a coaxial line. (a) Schematic diagram. (b) Disk and associated experimental apparatus.

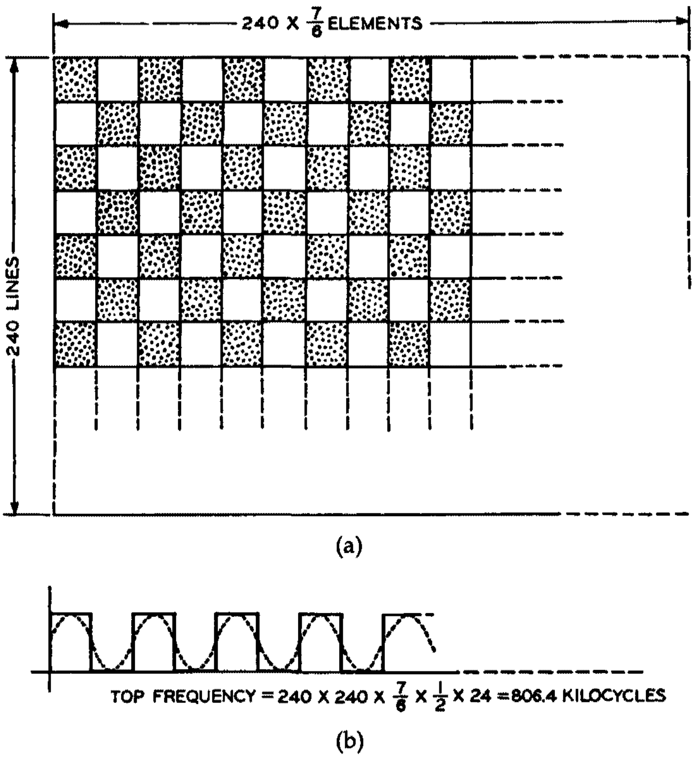


Fig. 6-18. Television transmission. (a) Diagram illustrating the resolution of an image into picture elements. (b) The derivation of maximum frequency required for transmission at 24 frames per second.

picture elements led to a frequency of 806 kHz needed for transmission, a bandwidth compatible with the design of the existing coaxial line [Fig. 6-18]. Display at the receiver was by means of a specially constructed cathode-ray tube designed to match the capabilities of the other system elements [Fig. 6-19]. The subject transmitted was a moving-picture film scanned by the transmitter. The object was to generate and recreate a picture under precisely controlled conditions. By using a film, exactly the same scene could be transmitted repeatedly while other elements were changed or adjusted.

The sequence of modulation steps is shown in Fig. 6-20. The first step in the two-stage process was to shift the signal with a carrier at 2376 kHz, creating a double-sideband signal in the range from 1570 to 3182 kHz. A carrier at a frequency of at least twice the baseband frequency was required to prevent the sidebands from falling back on the original signal. At this point the lower sideband was selected by a specially designed filter. Since all frequencies down to zero are involved in baseband video transmission, the two sidebands "touch" at the carrier, and it is not possible to select one and completely reject the other. The filter used passed essentially all of the lower sideband, but started

Fig. 6-19. The cathode-ray receiving tube used for the 1937 demonstration of television transmission over coaxial cable. The dimensions were chosen to obtain high linearity.

to cut off close to the carrier and passed a small vestige of the upper sideband as well. The second stage of modulation, with a carrier at 2520 kHz, translated this signal to the desired range within the passband of the line. The carrier was at 144 kHz, the desired sideband occupied the band from 144 to 950 kHz and the vestigial sideband extended down to 120 kHz. The widely separated upper sideband generated in this modulation step was easily suppressed.

At the receiver, the process was reversed. The transmitted signal was first moved up in frequency by modulation with a 2520-kHz carrier and then translated back to baseband video with a carrier of 2376 kHz. Unlike other modulation steps, where the carrier was at the edge or further removed from the signals to be shifted, the 2376-kHz carrier was within the signal band at the junction of the major sideband and the vestige. In the recovered baseband signal, the carrier was the zero frequency and the vestige was folded over by the modulation process to complement the attenuated section of the main sideband and thus, fully recovered the original signal.

This was a tricky arrangement. The filter to shape the desired sideband and vestige had to be very precisely designed. The carrier was very carefully placed

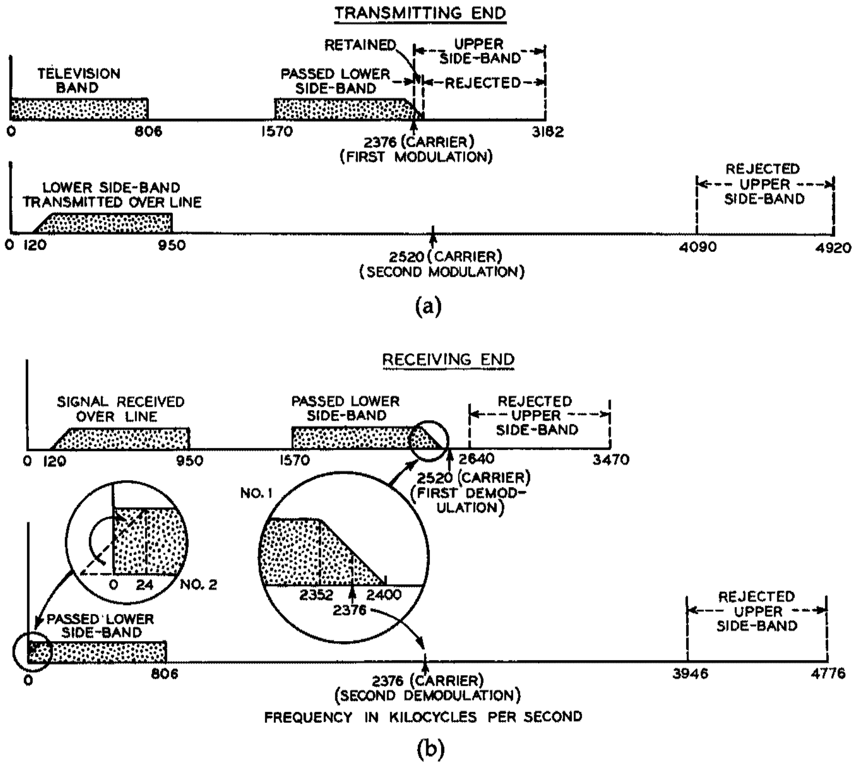


Fig. 6-20. Modulating and demodulating scheme for the 1937 television transmission over coaxial cable. (a) Modulation sequence at New York transmitting terminal. (b) Demodulation sequence in Philadelphia.

in the cuttoff region and had to be very closely held in frequency. There was some skepticism at first about even the concept of single-sideband transmission of video signals, since waveforms transmitted by that means appeared to be unduly distorted. However, H. Nyquist, building on earlier analyses he had carried out for carrier telegraphy, showed that the picture distortion was not as great as anticipated, and the advantage in resolution, due to the increased bandwidth possible with single sideband, more than compensated for its distortions.[17,18]

3.2 New York-Philadelphia Television Transmission

With the picture signal in the frequency range that could be transmitted, several characteristics of the coaxial line remained to be addressed. It was necessary that both the amplitude and phase of all the picture signal components be closely preserved. Amplitude equalization, while more exacting than that required for telephony, was relatively straightforward. Deviations of no greater than about ±1 dB were obtained. Phase or delay equalization was a more difficult matter. It was first necessary to determine a requirement, since none existed at that time. Preliminary objectives were set as part of an extensive program to analyze the signal and its properties. To reproduce the picture perfectly the various components had to arrive at the receiver in precisely the same time relationship in which they were generated. The requirement was most exacting for the fine detail and highest frequencies, where only a slight time displacement would be damaging. From a consideration of the smallest picture elements transmitted, it was judged that a displacement of no more than one-half of an element should be permitted. This translated to maintaining a delay displacement no greater than 0.3 μs for baseband frequencies from 3 to 806 kHz. The requirement on the line, of course, was for the shifted frequencies of 120 kHz and up and was just about met over the 95-mile line.

In addition to equalization, requirements for noise and interference were also derived. It was determined by laboratory tests that random noise should be at least 40 dB below the picture-signal amplitude. Pattern interference, however, is much more perceptible, and it was found necessary to hold this to 55 dB below the signal level. This situation left the system designers with their usual optimization problem. Too low a transmission level would exceed the signal-to-random-noise limit; too high a level would generate distortions, often showing up as patterned impairments. It was not too difficult to find this optimum point though, and, on November 9, 1937, motion pictures were transmitted from New York and displayed in Philadelphia. The tests were quite successful; no unsurmountable difficulties were met and no significant impairments were introduced by transmission over the line, compared to a picture generated adjacent to the transmitting terminal in New York.

Some additional signal processing contributed to this result. Part of the signal-analysis program included a characterization of the scanned signal spectrum [Fig. 6-21].[19] The energy of a television signal is concentrated in a line spectrum related to the line-scanning and frame rates. The relatively slowly changing picture content is contained in quite low-level sidebands around each of the spectrum lines. The high frequencies, required to reproduce the fine detail, contain very little energy, while the lower frequencies, extending all the way to zero, are required to reproduce larger areas of light and shade and contain most of the signal power. This signal-energy distribution, coupled to the result of studies of viewers' perception and tolerance of noise and impairments of different types, made it advantageous to pre-emphasize the signal. Pre-emphasis consisted of raising the level of the high-frequency components relative to the low frequencies, with a complementary restoration at the receiving end. A better optimum balance between noise and nonlinear distortion was thus achieved.

The signal analysis also revealed that there was essentially no energy midway between the spectrum lines. F. Gray pointed out that it was thus possible to transmit additional signals in the same overall band, providing they were adequately separated before detection at a receiver.[20] In early television transmission over coaxial cables, advantage was taken of this property to insert a control pilot in the otherwise empty frequency space. Later, it was the basis

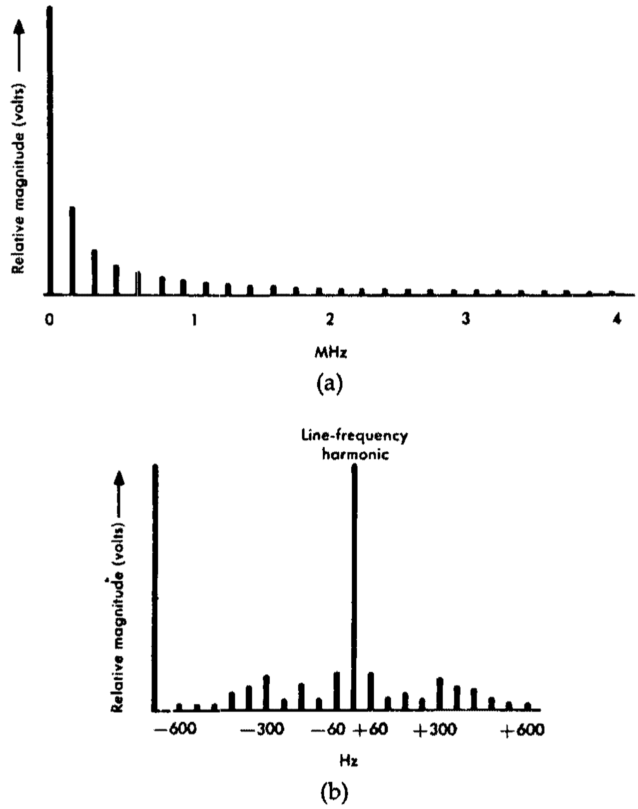
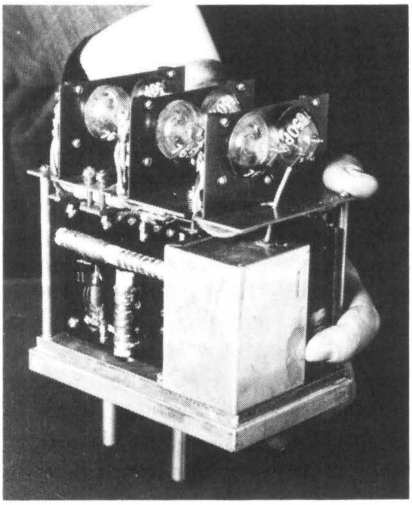


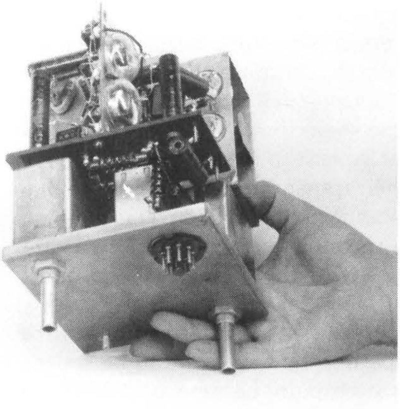
Fig. 6-21. Signal frequency composition of the monochrome television spectrum. (a) Typical spectrum showing every tenth line-frequency harmonic. (b) Typical sidebands around each line-frequency harmonic. The frequencies are for picture standards adopted in the early 1940s. Except for scaling factors due to the lower frequencies and scanning rates, the signal spectrum in the 1937 tests would be similar.

for inserting a color signal without extending the band. The color signal was not detectable by black-and-white receivers and a compatible system was thus realized.

The New York–Philadelphia transmissions provided a sound basis for the further development of coaxial cable for network television. Meanwhile, all-electronic systems were advancing rapidly. By 1934, RCA demonstrated a system using the iconoscope with 343 lines and 30 frames per second. A few years later, 441 lines became a commonly accepted interim standard. In the early 1940s, this was again modified to 525 lines and adopted by the National Television System Committee and Federal Communications Commission as a national standard, which it remains to this day. The increased number of lines implied much wider bands than those achieved in the New York–Philadelphia tests. The standard signal contained components to about 4.2 MHz. While there was little prospect that the first-generation coaxial design could reach that range, efforts were started, even before the first television tests were completed, to extend the frequency band. A new generation of line amplifiers, good for 2 MHz, was installed on the New York–Philadelphia line at 5-mile intervals in the summer of 1938 [Fig. 6-22]. The upgraded line was carefully tested for intermodulation, load capacity, thermal noise, crosstalk, and all the features



(a)



(b)

Fig. 6-22. Coaxial-line amplifier evolution. (a) Two-MHz amplifier, 1938. (b) Three-MHz amplifier, 1939.

judged necessary in a commercial system. A year later, prototypes of amplifiers for the first commercial system, with a further increase in bandwidth to 3 MHz, were installed at the same locations. This so-called 3-MHz system provided a useful band for television of about 2.7 MHz by terminal modulation and transmission techniques similar, except for the frequencies used, to those already described.

3.3 Local Links

From the start, broadcasters aimed to cover live events of the day as well as to transmit studio programs. Some form of link between the scene of the action and the television transmitter or network port was necessary. On May 20, 1939, the National Broadcasting Company (NBC), a subsidiary of RCA, covered the six-day bicycle races then in progress at Madison Square Garden in New York City and telecast them live in a half-hour program of sight and sound. The link between Madison Square Garden and the NBC transmitter at the Empire State Building was furnished by New York Telephone and Bell Laboratories as an experiment in local television transmission. Transmission was via existing No. 19 gauge telephone cable pairs extending from the arena to a central office and from the central office to the NBC studio.[21] Special amplifiers and gain and phase equalizers were used in the central office and at the terminals. Illumination was low compared to studios, but the results were judged to be distinctly satisfactory.[22] This test demonstrated the use of ordinary cable pairs, specially handled, for short links from the site of an event to the studio. (It might be noted in passing that selecting the six-day bicycle race as a subject was quite analogous to the use of a repetitive film strip for the first demonstrations between New York and Philadelphia!)

In June of 1940, the upgraded 3-MHz New York–Philadelphia line was used to transmit television coverage of the Republican National Convention in Philadelphia to New York City where it was broadcast live.[23] The small audience of television-set owners in New York at that early period were thus able to see as well as hear the galleries chanting "We want Willkie." This, too, was a joint undertaking by NBC and the Bell System. NBC generated the pictures in Philadelphia and delivered them to AT&T, who transported them over local wire paths and the coaxial system to New York and delivered them back to NBC for broadcast from their transmitter on the Empire State Building. The picture was at the 441-line standard and occupied, as generated, a nominal 3.5 MHz band. For short-haul transmission within the two cities, existing pairs in telephone cables were used. At the New York end, local transmission was also tested over new, special, shielded pairs with wider-spaced conductors [Fig. 6-23]. The special video pairs had less than one-fourth of the loss of ordinary No. 19 gauge pairs, permitting amplifiers to be spaced as much as five miles apart. In later years, such pairs became part of the standard equipment for local television transmission.

The transmission of the convention was a great success, and the last stage before a commercial coaxial system for television or multichannel telephony was ready for service. World War II intervened and suspended work on television broadcasting, as well as on coaxials for networking, but the system was

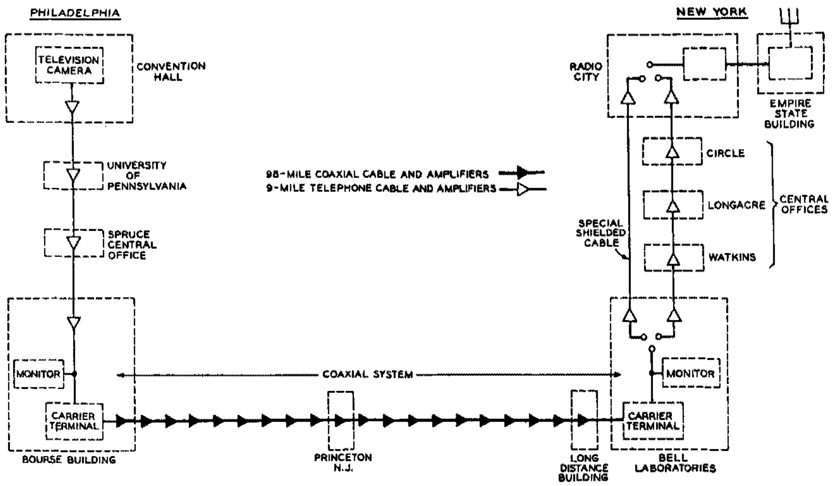


Fig. 6-23. The New York–Philadelphia network used to transmit the 1940 Republican Convention.

ready and spread rapidly when television returned following the war. For several years, in the mind of the public, coaxial cable became synonymous with television networking, even after microwave radio became a competitive, and even predominant, medium. Many a battle-scarred coaxial-development veteran listened with dismay as the announcer attributed the loss of a remote picture, for whatever reason, to "a failure of the coaxial cable."

IV. The Commercial System—L1 Coaxial Cable

After the 1936 tests and subsequent television transmission, the New York–Philadelphia line became a field laboratory for the improved equipment necessary for a commercial system. By the end of 1938, the designers were satisfied that they had a system suitable for regular service. Preparations were made for larger-scale manufacture, and, in May 1939, FCC approval was obtained for the construction of a line with four coaxial tubes from Stevens Point, Wisconsin to Minneapolis, a distance of 195 miles. Construction was started later that year.[24]

The line was divided into four main station sections of about 50 miles each [Fig. 6-24]. The first section west of Stevens Point was of aerial construction, but the open-country sections of the remaining 150 miles were directly buried by a cable plow to an average depth of 30 inches (surely a testimonial, both to the capability of the plow and to the quality of the Wisconsin/Minnesota soil) [Fig. 6-25]. In later construction, direct burial, usually to a depth of four feet in excavated trenches, which were then backfilled, was the usual procedure. The added depth not only reduced temperature changes, but provided highly valued additional physical protection from small-scale excavation work.

Development work had continued on the coaxial structure since the New

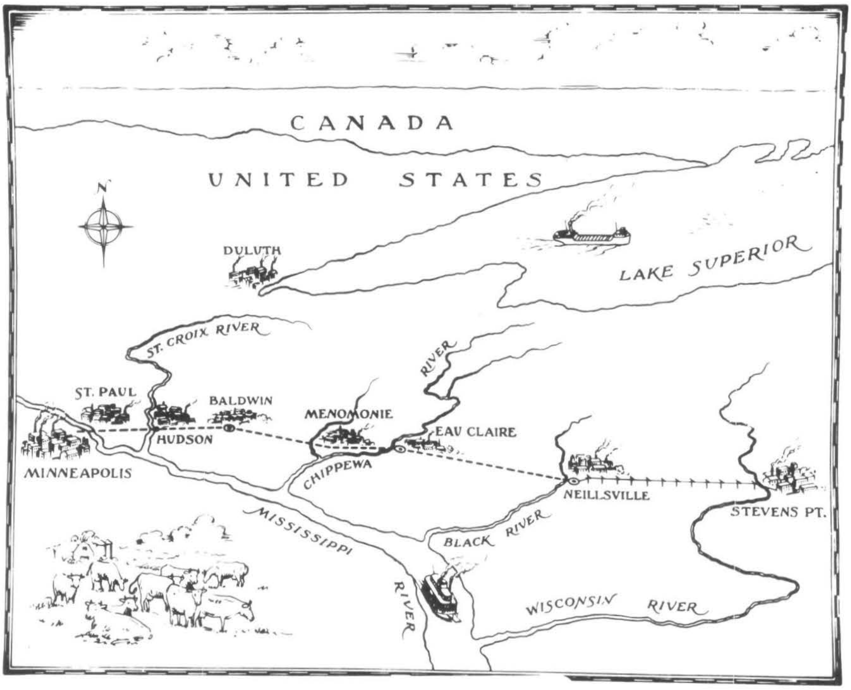


Fig. 6-24. The Stevens Point–Minneapolis coaxial-cable route.

Fig. 6-25. Plowing in the coaxial cable on the Stevens Point–Minneapolis route, 1939.

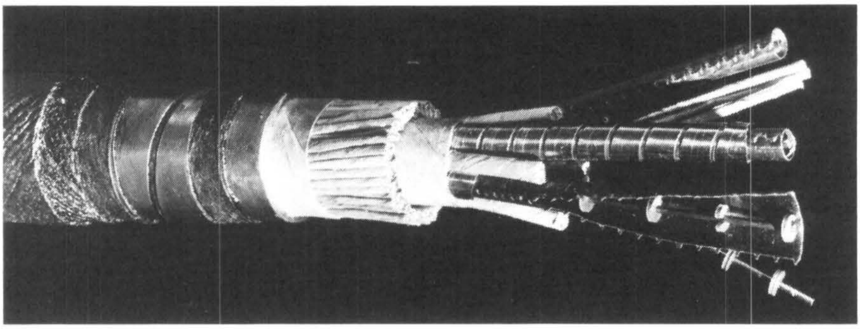


Fig. 6-26. Coaxial tubes with serrated longitudinal seam used in the Stevens Point–Minneapolis cable.

York–Philadelphia experimental installation. For the Stevens Point–Minneapolis installation, a new form of coaxial structure and cable was used.[25] The outer conductor of each coaxial tube was constructed of a single copper ribbon with serrated edges. This was formed into a cylinder around the disk-supported inner wire with the notched edges crimped together in a continuous longitudinal seam [Fig. 6-26]. The structure proved to be both less expensive and superior in transmission properties to the earlier multitape form. The complete cable was made up of four coaxial tubes and a number of pairs. The Stevens Point–Minneapolis line was pressed into preliminary service on two occasions in 1940—once when a storm disrupted the existing lines, and again, to help with heavy Christmas traffic. Regular service was initiated on June 9, 1941. The new system was designated the L1 coaxial system.

While television service was not planned on this route, the line was capable of transmitting the full 3-MHz band required. New terminal equipment, with a film scanner capable of generating 441-line images had been tested on a New York–Philadelphia–New York loop in January 1941 to demonstrate its television capability. In May of that year, the terminals were shipped to Minneapolis and applied to the Stevens Point–Minneapolis line with the four coaxials looped in series for a distance of 780 miles, the longest television circuit established up to that time. As in all loop tests, it was possible to compare the looped signal with a local picture. Experienced observers could detect the effect of the long line on the pictures, but to most observers they were hardly perceptible.

4.1 L1 Terminals

The commercial L1 telephone terminals differed from the earlier experimental versions by using three stages of modulation and a different approach to line-frequency assignments [Fig. 6-27]. Instead of modulating each 12 channel group directly to the line frequency, five groups were shifted to form a 60-channel block, designated a supergroup, in the band from 312 to 552 kHz, and this was then translated to the line frequency in a third modulation step. Within the supergroup, the channels continued to be stacked side by side at 4-kHz intervals. The five 12-channel groups making up the supergroup were divided

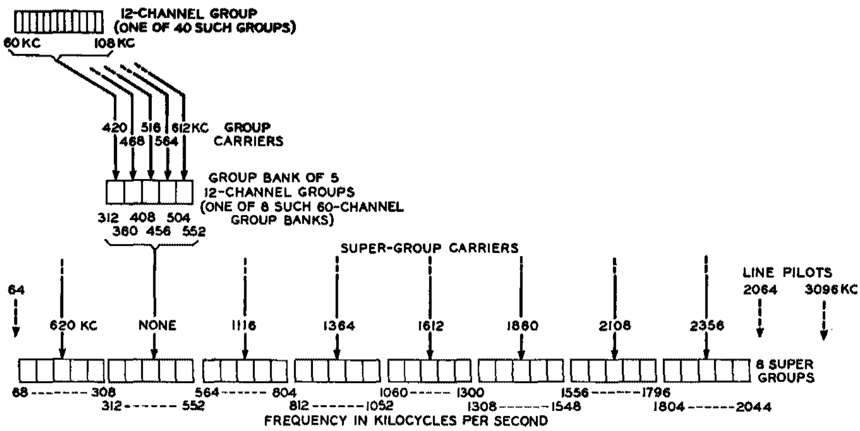


Fig. 6-27. Frequency allocations and modulation steps for 480 telephone channels on the L1 coaxial system.

into odd and even groups, separated in frequency, so they could be easily isolated by inexpensive filters. They were then combined into the supergroup by means of a hybrid [Fig. 6-28]. In contrast with the 1936 line-frequency assignments where the groups were placed chockablock with a channel every 4 kHz over the entire line frequency range, the supergroups were spaced in the line-frequency spectrum with an 8-kHz gap between each one. The spacing permitted selective filtering for blocking and branching of each supergroup without disturbing the transmission of the adjacent supergroups. The coaxial designers were learning that their superhighway needed occasional interchanges with other highways, without bringing the entire stream of traffic down to individual voice channels. Blocking and branching had been successfully tested on the New York–Philadelphia line during the testing of the final L1 design.

Pilots were supplied to the line at 64, 556, 2064, and 3096 kHz. The 2064-

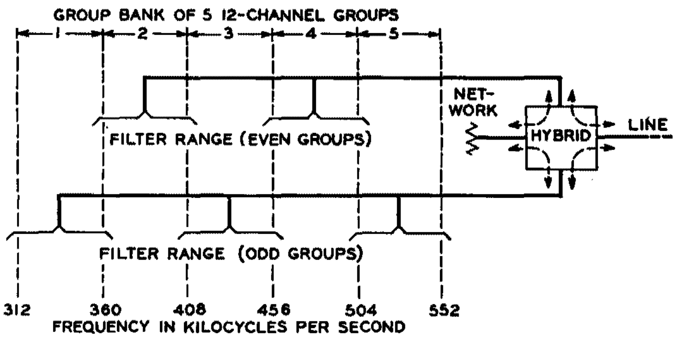


Fig. 6-28. Frequency allocation and combining arrangements for the five groups of each supergroup.

kHz pilot, just above the top telephone channels, was used for basic temperature regulation at every auxiliary repeater. (This was absolutely necessary on the aerial link). At main stations, the 556-kHz pilot was used to control an equalizer to correct for accumulated errors in line-temperature regulation. At the same locations, the 3096-kHz pilot was used in an automatic equalizer to compensate for transmission deviations caused by equipment temperature changes. In the initial installation, the 64-kHz pilot was used only for supervision and some manual adjustments. In a longer system, it was expected to be of use for additional automatic equalization. In addition to the line pilots, a 92-kHz pilot was inserted in each group, in place of the balanced-out carrier, to provide a test tone for group continuity.

As in other carrier systems, the L1 design reflected attention to reliability and continuity of service, of special concern on coaxial cable because of the large number of circuits. A single coaxial pair far exceeded in capacity the moderate needs of the Stevens Point–Minneapolis route, even allowing for growth, but the additional tubes were furnished as hot spares to be switched into service in the event of a working-line failure. A complete set of spare modems, filters, and hybrids was provided to form a 60-channel supergroup from any five group banks, as well as a full complement of equipment necessary to modulate the basic supergroup to any line frequency. The carrier supply that furnished the array of modulation tones was completely duplicated. Power was furnished from the commercial AC source but, in the event of a power failure, a DC-to-AC inverter, driven from the office battery, was automatically switched in.

The substitution of any spare for a working unit was arranged, as far as possible, to cause no break in service. The transmission path was duplicated in parallel, via hybrids, before any working unit was removed. In addition to a patchboard at voice frequency for all circuits, the input and output of all modems were brought to high-frequency patchboards to permit rapid and convenient substitution. Jackboards were provided for testing on either a "terminated" or "bridged" basis. Terminating tests removed equipment from service, with the test set terminating the transmission path and matching its impedance. Bridging tests were made with high-impedance sets shunted across the line for in-service testing. Patching and test jacks were centrally located in the terminal bay lineup [Fig. 6-29].

All these measures involved expense, of course, but experience has fully justified their value. Over many decades and generations of transmission facilities of every type, the continuity and reliability of Bell System telephone service was a source of satisfaction to the public and a matter of pride to its designers, manufacturers, and operators.

4.2 Post-World War II—L1 Carrier

While the war interrupted further development work on coaxial cable as well as on all other civilian transmission projects, a limited amount of construction was needed to meet wartime needs. Planning for the postwar era continued, especially as the war neared its end. In December 1944, with the end of the war in sight, AT&T announced an ambitious plan for the extension of its coaxial

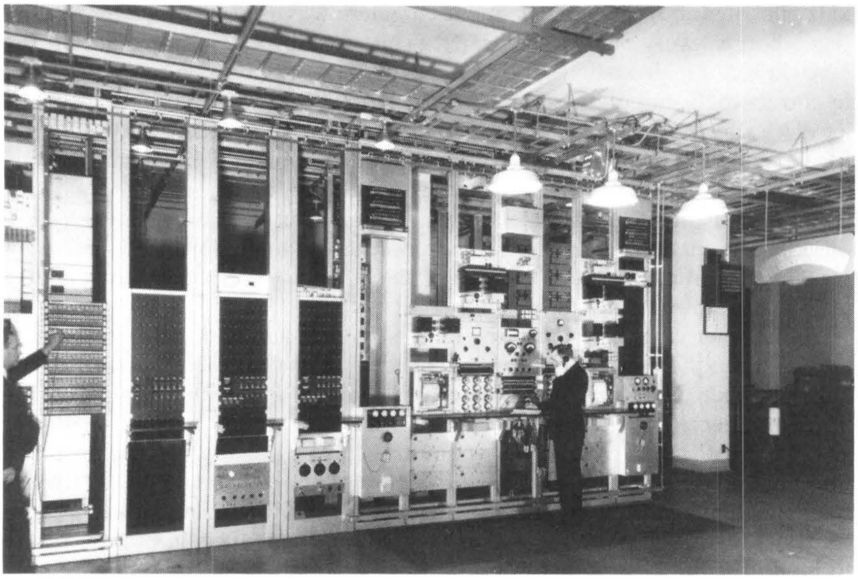


Fig. 6-29. L1 terminal equipment showing centrally located patching and test jacks.

network. At that date, in addition to the Stevens Point–Minneapolis link, the New York–Philadelphia line had been fitted for commercial service and extended through Baltimore to Washington. Lines from Atlanta to Jacksonville and from Terre Haute to Saint Louis were under construction. The 1944 plan called for construction of 6000 to 7000 miles of new line over the subsequent five years. The plan called for a line down the East Coast from Boston to Miami, a southern transcontinental route from Atlanta to Los Angeles, and major links from the East Coast to Chicago, St. Louis, and New Orleans. All the new links were to provide at least six coaxials in a sheath.

In the same announcement, AT&T indicated its intention to provide extensive television network capability over the coaxial lines and to provide local links to the intercity coaxial network access points and between studios and transmitters. At the same time, plans were revealed for a new system with a 7-MHz band, capable of transmitting a full 4-MHz, 525-line television signal, plus 480 telephone channels simultaneously. In the postwar years, this projected development became the L3 coaxial system with a band of over 8 MHz, able to transmit 600 telephone channels in addition to television. (Whenever a gap occurs in the otherwise sequential series of system designations, it is a fairly safe assumption that a development was abandoned or that a system concept was studied for a time and rejected. L2 was a small-scale attempt to adapt coaxial lines to light traffic and shorter routes, a constant matter of interest, because of the enormous number of such situations. It was abandoned at an early stage.)

The deployment plan was carried out essentially as projected. Implementation was substantially helped by two major improvements in the coaxial structure design. The first was a substitution of polyethylene for the hard rubber in the supporting disks. Prewar experiments had established the electrical superiority of polyethylene to rubber as well as its mechanical adequacy, but wartime conditions delayed its introduction [Fig. 6-30]. (See another volume in this series, *Physical Sciences (1925-1980)*, Chapter 14, Section III.) Following the war, its use was most fortunate because, for a time, the rubber became almost unavailable. The other major change was an increase in coaxial-tube diameter with a consequent decrease in loss. The inside diameter of the outer conductor was increased to 0.375 inch, and the inner conductor diameter, maintaining the 3.6-to-1.0 ratio, became 0.1 inch. The outer conductor thickness was also increased slightly, to 0.012 inch, to improve low-frequency transmission. The larger tube, first used in an eight-coaxial sheath between Dallas and Fort Worth in 1946, had substantially lower loss, permitting eight-mile spacing for L1 repeaters instead of the approximately five miles required with the 0.27-inch tube. The lower loss and wider repeater spacing were especially important in reducing the number of repeaters required in the much wider-band systems that lay in the future. Not the least virtue of the new design was its superiority for AC transmission. The larger-center conductor permitted power feeds to be

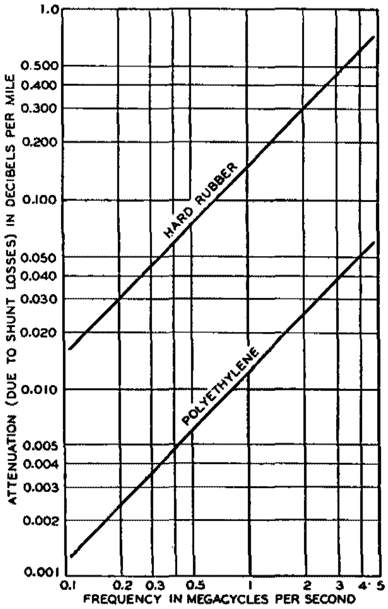


Fig. 6-30. Attenuation in coaxials caused by shunt losses in polyethylene insulators compared to hard rubber.

spaced as much as 165 miles apart. The latitude in power-feed spacing was such that the locations were generally determined by other considerations.

In later years, even larger-diameter structures were often considered. Assuming that manufacturing and handling problems could be solved, this had a great attraction for new line construction. But a major application for (and a constraint on) new systems was always the need to upgrade existing lines, with the repeater spacings fixed by existing equipment huts or manholes, and the consequent advantage of making new spacings submultiples of existing spans. In addition, with microwave radio growing even more rapidly than coaxial cable, the extent of new cable installations was always uncertain. The three-eighths inch coaxial remained the standard in 1984.

As the systems were installed and grew in length, much new information was learned in the field.[26] Despite rapid advances in the ease and precision of measurements, equalization and measurement continued to be a problem. With end-to-end system requirements established and divided down (by a factor of several hundred) to determine the requirements to be met in individual repeaters, the accuracy needed often exceeded the capability of the best measuring equipment available. A build-and-learn approach was necessary. Amplifiers were designed and built as close to the desired characteristic as could be determined by measurement, but, in addition, special pains were taken to make them as alike as possible in every feature that could affect transmission. The long multirepeater lines were then carefully measured to determine the multiplied deviations. This occasionally led to modifications in the repeater design, but was usually more helpful in the design of equalizers to mop up the accumulated deviations. At each main station, so-called A equalizers were provided to correct deviations that were fairly well predicted. B equalizers were provided at longer intervals, their design determined more by the process just described.

By these methods, the systems were equalized to a degree entirely adequate for telephony. Residual deviations from desired levels could always be corrected in the terminals by gain adjustments at supergroup-, group-, or even individual-channel levels. However, transmission for long distances at levels higher or lower than the optimum or with prolonged exposure to an interfering signal was harmful to the signal-to-noise ratio. A further measure helped to meet long-distance telephony objectives. At 800-mile intervals, the recommended practice was to demodulate the entire array of line signals to supergroup or even group level. The line signal could then be reconstructed with the pieces assigned to new frequency slots, thus averaging out impairments due to noise, interference, and transmission deviations in the different parts of the frequency band. This frogging, or scrambling, was not possible on a long television connection, and good equalization for television always remained a problem on the L1 system.

As the system spread, maintenance and operational methods were developed and adapted to experience. The typical auxiliary station was a simple concrete block hut without heat, light, or power except for that on the coaxial cable. Only a minimum of checks and tests to identify faulty equipment were made at the site. Defective amplifiers or other equipment was replaced and the removed unit taken to a main station where further tests were possible. These stations were equipped for more extensive tests and to make simple repairs, such as replacing tubes. For more elaborate repairs, including the means to reconstruct an amplifier completely, the equipment was shipped to one of four or five regional repair centers strategically located around the country.

Later, as the reliability of the system was established, confidence grew. Continuous power was furnished by AC generators driven by commercially powered AC motors with DC motors on the same shaft. If commercial AC power failed, the inertia of the rotating machines maintained line power until the DC motor, connected to the power feed station's battery, could pick up the load. Even main stations became only partially attended. Full-time attendance was required only at major stations at intervals of 600 miles or so. At these points, as many as 150 remote alarms and indications could be displayed via the maintenance pairs that accompanied the coaxials.

For television networking, L1 dominated until microwave radio became a significant factor after 1950. By 1948, over 5000 route miles were in service for television, with established television networks along the East Coast (the New York–Boston link by experimental microwave radio) and from St. Louis to Buffalo.

In time, a ninth and tenth supergroup, with somewhat higher noise, but adequate for shorter distances, were added to L1 in the frequency band above 2 MHz. At its peak, about 1954, before widespread installation of TD radio and L3 coaxial absorbed the growth, L1 provided over 10 million miles of voice-circuit capacity and about 20,000 television-channel miles.

REFERENCES

1. A. Russell, "The Effective Resistance and Inductance of a Concentric Main, and Methods of Computing the Ber and Bei and Allied Functions," *Phil. Mag.* **17** (April 1909), pp. 524-552.
2. H. J. Round, U.S. Patent No. 1,452,849; filed December 13, 1921; issued April 24, 1923.
3. R. S. Ohl, U.S. Patent No. 1,619,882; filed December 4, 1924; issued March 8, 1927.
4. C. S. Franklin, British Patent No. 284,005; filed October 19, 1926; issued January 19, 1928.
5. L. Espenschied and H. A. Affel, U.S. Patent No. 1,835,031; filed May 23, 1929; issued December 8, 1931.
6. L. Espenschied and M. E. Strieby, "Systems for Wide-Band Transmission Over Coaxial Lines," *Bell Syst. Tech. J.* **13** (October 1934), pp. 654-679.
7. S. A. Schelkunoff, "The Electromagnetic Theory of Coaxial Transmission Lines and Cylindrical Shields," *Bell Syst. Tech. J.* **13** (October 1934), pp. 532-579.
8. M. E. Strieby, "A Million-Cycle Telephone System," *Bell Syst. Tech. J.* **16** (January 1937), pp. 1-9.
9. E. I. Green, "The Coaxial Cable System," *Bell Lab. Rec.* **15** (May 1937), pp. 274-280.
10. C. Kreisher, "Construction of the Coaxial Cable," *Bell Lab. Rec.* **15** (June 1937), pp. 325-328.
11. K. C. Black, "Repeaters for the Coaxial System," *Bell Lab. Rec.* **15** (August 1937), pp. 385-390.
12. *Television: Collected Addresses and Papers on the Future of the New Art and Its Recent Technical Developments*, Volume I (New York: RCA Institutes Technical Press, 1936); *Television: Collected Addresses and Papers on the Future of the New Art and Its Recent Technical Developments*, Volume II (New York: RCA Institutes Technical Press, 1937); *Television: Volume III (1938-1941)* (Princeton, New Jersey: RCA Review, 1946); *Television: Volume IV (1942-1946)* (Princeton, New Jersey: RCA Review, 1947);

Television: Volume V (1947-1948) (Princeton, New Jersey: RCA Review, 1950); and *Television: Volume VI (1949-1950)* (Princeton, New Jersey: RCA Review, 1950).
13. See reference 12, Vol. I, pp. 202–219.
14. H. E. Ives, "Television," *Bell Syst. Tech. J.* **6** (October 1927), pp. 551-559.
15. See reference 12.
16. M. E. Strieby, "Television Over the Coaxial Cable," *Bell Lab. Rec.* **16** (February 1938), pp. 188-195.
17. H. Nyquist and K. W. Pfleger, "Effect of the Quadrature Component in Single-Sideband Transmission," *Bell Syst. Tech. J.* **19** (January 1940), pp. 63-73.
18. S. Goldman, "Television Detail and Selective-Sideband Transmission," *Proc. IRE* **27** (November 1939), pp. 725-732.
19. P. Mertz and F. Gray, "A Theory of Scanning and Its Relation to the Characteristics of the Transmitted Signal in Telephotography and Television," *Bell Syst. Tech. J.* **13** (July 1934), pp. 464-515.
20. See reference 19.
21. C. L. Weiss, Jr., "Television Transmission Over Telephone Cables," *Bell Lab. Rec.* **18** (October 1939), pp. 34-37.
22. "Television Pick-Up Over Telephone Cable Pairs," *Bell Lab. Rec.* **17** (June 1939), p. 313.
23. "Television for National Republican Convention," *Bell Lab. Rec.* **18** (July 1940), p. 336.
24. K. C. Black, "Stevens Point and Minneapolis Linked by Coaxial System," *Bell Lab. Rec.* **20** (January 1942), pp. 127-132.
25. O. S. Markuson, "Stevens Point-Minneapolis Coaxial Cable," *Bell Lab. Rec.* **19** (January 1941), pp. 138-142.
26. L. G. Abraham, "Progress in Coaxial Telephone and Television Systems," *Trans. Amer. Inst. Elec. Eng.* **67** (1948), pp. 1520-1527.

Chapter 7

Ultrahigh Frequencies, Microwaves, and Radio Relays

I. EARLY CONCEPTS AND DEVELOPMENT

The idea of receiving a weak radio signal, amplifying or regenerating it, and retransmitting it at higher power occurred to inventors almost from the start of wireless telegraphy. Proposals for relaying radiotelegraph signals are found in United States patents issued to J. Stone as early as December 1902. Apparently few, if any, radiotelegraph stations were built for the specific purpose of relaying, although it was probably a fairly common practice for an intermediate operator to receive and retransmit a message that was too faint at a more distant station.

No attempt to relay a speech message by radio is known prior to 1915, but an entry in November of that year in a Western Electric laboratory notebook proposes the idea, no doubt stimulated by the work on the first repeatered telephone wire lines then going into service. One of the first commercial two-way radiotelephone links was established between Catalina Island and the mainland of California in 1920. Shortly after it went into service, a two-way connection was established by radio from the SS *Gloucester*, some distance out in the Atlantic Ocean, to an East Coast shore station. From there, it was extended by wire over the transcontinental line and, by radio again, to Catalina Island. Perhaps as much as any branch of technology, the history of radio transmission is punctuated with demonstration spectaculars. We need not ask why anyone on Catalina Island should want to talk to the *Gloucester*, and the transmission must have been terrible, but it was a radio relay of a sort.

The technology of the 1920s, however, was not advanced to the state where widespread radio relaying of telephone messages could become practical. To many people, on the other hand, it continued to seem only reasonable that radio developments would parallel the repeatered and multiplexed telephone wire line work, and numerous patents were issued for repeatered radio links. Even the earliest of these, in 1921, proposed multichannel multiplexing and recognized the need to shift the radio frequency at the repeater to avoid singing around the local transmitter-receiver loop.[1]

It has been characteristic of radio that, as advancing technology opens up new bands, there is a rush to fill the newly available spectrum with a host of services competing for frequencies. Following Marconi's demonstration that relatively long waves were suitable for long-distance transmission (his first transatlantic transmission was at 1100 meters [270 kHz]), it was soon observed that the lower the frequency, the better the long-distance transmission. Frequencies as low as 15 or 20 kHz became the favored choice for long-distance radiotelegraph service. Frequencies of several hundred kilohertz were much poorer, and, above 1500 kHz, transmission beyond the horizon became very poor indeed. (In 1925, H. W. Nichols and J. C. Schelleng at Bell Laboratories showed this to be due to a resonance phenomenon of the charged particles of the ionosphere in the earth's magnetic field in the range of 1.5 to 2 MHz.)[2] The frequencies above long-distance telegraphy and below the absorption range were ideal for local coverage, and, in the 1920s, the radio broadcast services filled the band from 500 to 1500 kHz with their signals.

Marconi, the father of long-wave radio, realized that if new services were to be had, they would have to be in a new and unused range—that is, at higher frequencies. As early as 1916, he directed his organization to start experiments at frequencies of 5 MHz and higher. One of the advantages, of which he was well aware, was the ability to realize more directive antennas at the shorter wavelengths. To the surprise of everyone, these shortwaves (high frequencies) were discovered to have useful long-distance transmission properties. They were above the absorption resonance and were returned to earth by the ionosphere. As the properties of the ionosphere and the facts governing shortwave transmission became better known in the 1920s, these frequencies were also quickly occupied for long-distance communication. The first transatlantic telephone circuit, opened in 1927, was at a frequency of about 60 kHz, but development at high frequencies (HF), in the range of 5 to 25 MHz, was already under way.[3] The band available at the high frequencies made many more links possible than in the long-wave region developed earlier.

Shortwave propagation and directive antennas were quickly adopted by inventors seeking radio relays. For a time, pioneer Station KDKA in Pittsburgh transmitted at 5 MHz as well as at the usual broadcast frequencies of the day, and its programs were occasionally received and rebroadcast in England. The Westinghouse Electric Corp., seeking to establish its own radio-program network, established a shortwave radio-relay station in Hastings, Nebraska to relay its programs to stations on the West Coast. Unfortunately, the antennas, while moderately directive, also had significant levels of radiation in directions other than the one desired, and propagation was not at all limited to the range of the intended receiving station. The programs were often heard in Europe and South America as well, where, it may be assumed, they were not always welcome.

Bell System inventors, too, seized on the possibility of HF for radio relay. Patents issued in the 1920s attempted to utilize the shortwave frequencies.[4] Novel use was made of nulls in the antenna pattern and zigzag routes to reduce the problem of "overreach" to the next station [Figure 7-1]. It was hoped by this means to reuse the radio frequencies, reducing the number required in a multihop chain. In the early 1930s, the possibility of improving transatlantic HF transmission by means of island or ship-based HF relays was also studied in the Bell System.

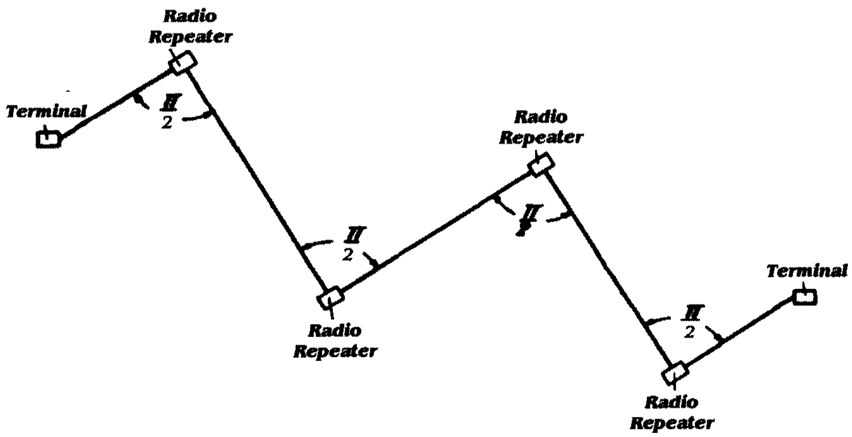


Fig. 7-1. Patent figure for proposed directive radio repeating system, 1928.

None of these relaying schemes came to anything. For overland routes, costs were prohibitively high for a single telephone channel that could be realized at much lower cost by a wire pair. With limited antenna gain, path loss was high, and tubes had to be pushed close to their load limit. Attempts to reduce the cost per channel by multiplex transmission were thwarted by the nonlinearity and intermodulation in the transmitters. In the 1930s, while some success was achieved in telephone multiplexing over radio, in general, radiotelephony continued to be restricted to overseas and shorter links to otherwise isolated localities. Quality was relatively poor compared to the high-grade wire circuits coming into common use. The radio art was not yet ready for large-scale telephone relaying.

II. WORK ABOVE 30 MEGAHERTZ—ULTRAHIGH FREQUENCIES

By the early 1930s, the radio research and development work at Bell Laboratories was divided into three categories. Two of these—continuing development for broadcast radio and HF transmission in the range from 5 to 25 MHz for overseas two-way telephony—had been established in the 1920s and continued to be pursued vigorously. The third category was the investigation of the relatively unexplored area of ultrahigh frequencies (UHF)—frequencies above 30 MHz. This exploration was stimulated, both by a developing need for transmission in this range and by advances in the means for accomplishing it.

Perhaps foremost was the growing awareness of the prospect for television, along with the knowledge that large frequency bands would be required for its transmission. By the early 1930s, analysts understood that a single television signal of even moderate definition would require a radio band larger than that assigned to the entire audiobroadcast industry. Since the right to use even a small piece of the existing band was usually the subject of an intense contest, it was clear that the large blocks needed for television (multiple channels were assumed from the outset) would have to be in unoccupied bands. Fortunately, the technology tends to make the signal bandwidth that can be managed about proportional to the frequency in which the band is located. That is, it is about equally difficult (or easy) to handle (amplify, modulate, transmit, or broadcast) a 10-MHz-wide band at 100 MHz as to do the same for a 1-MHz band at 10 MHz. Both require a ten-percent band. All factors pointed to the UHF frequencies as the natural home for television. To a lesser degree, the emerging needs for FM broadcasting with wide deviation also fitted well with a UHF assignment.

AT&T's interest was primarily with the prospect for radio relaying. The properties of radio waves at UHF were especially attractive for this application. The short wavelength made possible highly directive antennas of moderate size, with the corresponding prospect of low transmitter power and better linearity. The ionosphere did not normally return these frequencies to the earth as it did those in the HF band. Transmission was generally restricted to a line-of-sight optical path, with little propagation beyond the horizon. The restricted transmission range and narrow beams made possible the reuse of frequencies in fairly small geographic areas. The broad band available and the prospect for reuse promised to make a very large communications resource available. Directivity, low power, limited transmission range, broad band, and reuse of frequencies looked exactly like the combination for which the radio-relay people had been searching.

Others too, of course, were conscious of these factors. The Radio Corporation of America (RCA), whose primary interest was in broadcast television, also understood the need to transmit expensive program material for simultaneous broadcast in more than one city. As part of their early television experiments about 1934, they established a relay link from New York to Philadelphia.[5] Signals broadcast at 44 MHz from the Empire State Building in Manhattan were received at Arney's Mount, a high point in southern New Jersey. From there, they were transmitted at 79 MHz over a relatively short line-of-sight path to a receiver in Camden, New Jersey for rebroadcast in the Philadelphia area.

In addition to two-way telephony, the Bell System wished to play the same role in television-program networking that it already occupied in the well-established service for audiobroadcast programs. The coaxial cable had been designed with both objectives in mind; radio relay was expected to follow the same path. That interest in Bell Laboratories was intense is indicated by the files of that period. A casual survey of the system development files reveals that, in the 1930s, more than 90 memos were addressed to one aspect or another of radio relaying.

2.1 Vacuum Tubes

Some of the advantages for relaying, especially antenna directivity, increased more or less proportionately with frequency. For this and other reasons, a considerable number of workers began to investigate the upper frequency limits of vacuum tubes. The experiments typically explored the highest frequency at which a tube would oscillate. This was closely related to the maximum frequency at which amplification was possible, and oscillation was much easier to detect and measure than amplification. Oscillating frequencies up to 100 MHz had been obtained with conventional negative-grid biasing before 1920. Using tubes from which the bases had been removed and thus reducing a limiting capacitance, G. C. Southworth produced oscillations up to 300 MHz in 1920. Throughout the 1920s, a number of experimenters achieved frequencies as high as 300 to 600 MHz using, in some cases, the harmonics of a fundamental frequency.

The work at Bell Laboratories consisted of analyzing and understanding the mechanisms that established a limit, as well as working in the laboratory to produce tubes that extended it. In 1927, C. R. Englund analyzed the structure of a number of readily available tubes and found the upper limit, due to lead inductance and element capacitance, to be about 200 MHz.[6] W. E. Benham, in England, noted the significance of the transit time of the electrons in their passage from the cathode to the plate in 1928.[7] In 1934 and 1935, F. B. Llewellyn extended the analysis of oscillation limits to include the effect of the transit time and dissipation in the leads, concluding that tubes of the then available construction would be limited to frequencies of several hundred megahertz.[8,9]

In the meantime, in a number of laboratories, efforts were under way to reduce the limiting conditions as far as materials and methods would permit. One approach was to abandon the concentric cylindrical cathode, grid, and plate structure, commonly used at the time, in favor of a closely spaced parallel-plane structure. At RCA, this approach was adopted; and, by shrinking the dimensions as far as available tooling permitted, designers developed a series of miniature low-power receiving-type acorn tubes, capable of oscillation up to 1000 MHz [Fig. 7-2].[10] In Bell Laboratories, effort was directed more to higher-

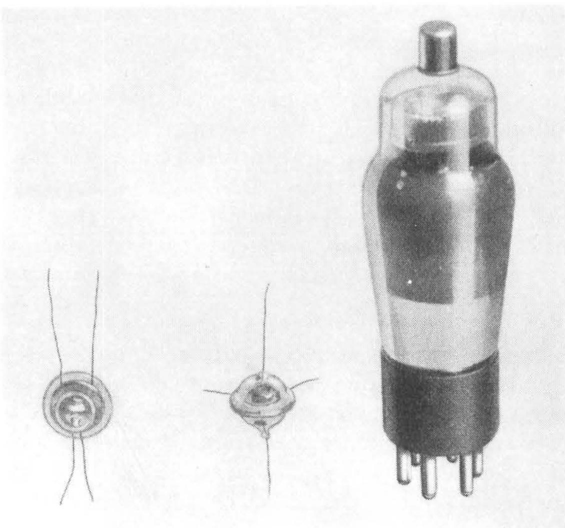


Fig. 7-2. Receiving tubes of extremely small dimensions. *Right,* a conventional receiving-type tube is shown for comparison with, *left,* miniature acorn tubes.

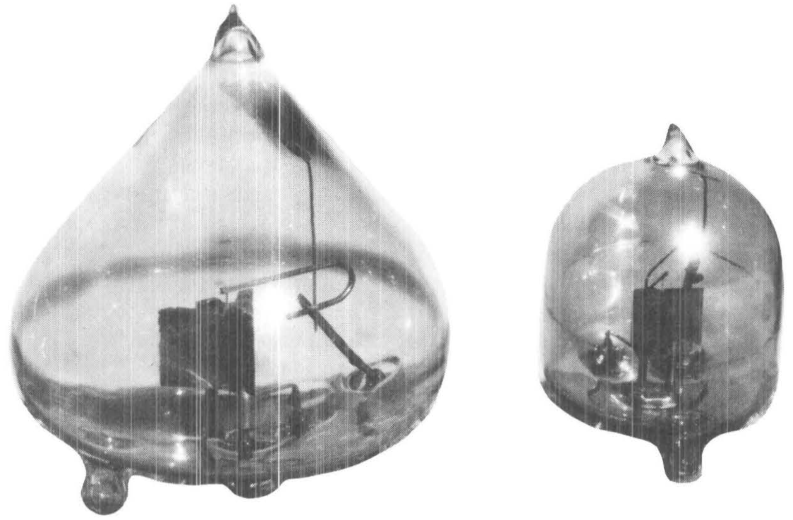


Fig. 7-3. High-power, high-efficiency UHF vacuum tubes developed at Bell Laboratories, 1935. *Left*, a tube that generated 10 w at 670 MHz with 20-percent efficiency. *Right*, a tube that generated 1 w at 1200 MHz and 10-percent efficiency.

power transmitting-type tubes. By 1935, tubes were produced capable of delivering 10 w at 670 MHz with 20-percent efficiency as oscillators and others capable of 1 w at 1200 MHz and 10-percent efficiency [Fig. 7-3].[11] Later in the decade and in the early 1940s, work at the General Electric Co. using very close spacings and low-inductance disk seals to connect to the grid, produced negative-grid triodes capable of working even further into the microwave range. The availability of tubes of suitable capability was not, in general, a limitation on the use of ultrahigh frequencies. The performance figures quoted, however, were usually obtained under carefully controlled laboratory conditions. Incorporating the tubes into effective working circuits and practical systems imposed many additional limitations, and system experiments were generally at much lower frequencies than the oscillation limits.

2.2 Propagation

Along with the work on generating, amplifying, and detecting signals in the new wavelength ranges, an extensive program was carried on by Bell Laboratories to determine the propagation properties of these waves through the atmosphere. (See another volume in this series, *Communications Sciences (1925-1980)*, Chapter 5, Section I.) The experience leading to the effective utilization of the HF signals was still relatively fresh in people's minds. HF frequencies that were not expected to propagate much beyond a line of sight were found to be capable of transoceanic transmission, and, by the early 1930s, were providing the backbone of worldwide radio communications. It was believed that the limit to ionospheric reflection had been reached in the range from 30 to 40 MHz, but experimenters were leaving no possibility untested and were alert for evidence of any new phenomenon.

In 1932, Bell Laboratories conducted propagation tests at frequencies of 17 and 80 MHz over a variety of paths in the vicinity of their New Jersey laboratories.[12] The paths were chosen to be both at and well beyond line of sight and over both land and water. In general, the method was to transmit from a fixed location convenient to one of the laboratories and to receive with portable equipment at various remote points over a period of time. The receivers were calibrated by application of a known voltage to the receiving antenna terminals. On occasion, traverses were made, measuring the received signal as a function of distance from the transmitter [Fig. 7-4]. Some traverses were made to a receiver in the Bell Laboratories airplane, a Ford Tri-motor *Tin Goose* [Fig. 7-5]. The alternate reinforcement and cancellation between the direct path and the wave reflected from the earth was clearly observed. In other tests, a sharply defined interference pattern was observed due to reflection from an isolated tree on flat ground [Fig. 7-6].

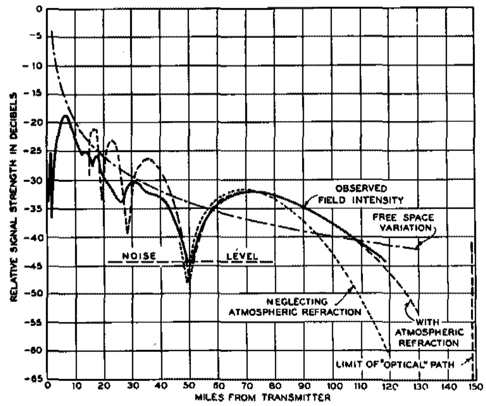
In most of these tests, quite good agreement was obtained with the theory

Fig. 7-4. Automobile-mounted receiving equipment for measuring the strength of ultrashortwave fields during overland traverses.

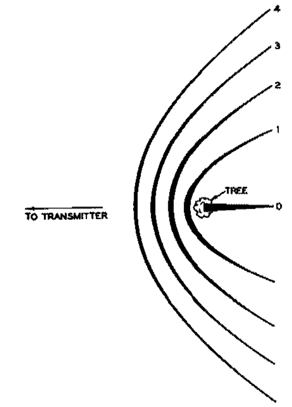
Fig. 7-5. The Bell Laboratories Ford Tri-motor airplane with fixed V and strut antennas, 1930.

developed over several years, both in the United States and abroad. Results were explainable in terms of inverse distance loss over the direct optical path modified by reflection, refraction, and diffraction, the latter two being especially important for paths below the horizon, beyond line of sight. Among the noteworthy results of this work was that reflection was not only quite regular over water paths as expected, but was also surprisingly so over land of moderate relief, where the irregularities were large compared with the wavelength. On further consideration, it was realized that this was to be expected at near-grazing incidence. In later work, reflection coefficients as high as 0.8 or 0.9 for grazing paths over wooded terrain were found even at frequencies into the microwave range, above 1000 MHz. For flat fields, deserts, or over-water paths, strong reflections were found down to wavelengths of a few centimeters or less. One result of this work was rapidly incorporated into radio lore worldwide. It was found that the refraction of the waves due to the decreasing density of the air with altitude could be accounted for by assigning a fictitious radius to the earth equal to four-thirds its actual radius and then treating the paths as if there were no atmospheric refraction. (Atmospheric refraction is well known at optical frequencies. The setting or rising of a star occurs when it is about 35 minutes of arc below the geometric horizon.)

In late 1933, the propagation tests were extended with measurements at 65 and 190 MHz.[13] For long over-water traverses of 100 miles or more, tests were again made using the airplane. Received-signal levels from a fixed transmitter on the New Jersey shore were recorded while the airplane flew at various altitudes away from and toward the transmitter, over the ocean along the south shore of Long Island. Results were obtained at altitudes from 2500 to 8000 feet and at ranges out to 120 miles [Fig. 7-7]. Again, fair agreement with theory



(a)



(b)

Fig. 7-6. UHF field-strength measurements, 1933. (a) Cancellations and reinforcements between the direct path and the wave reflected from the earth during an airplane traverse. Measurements are represented by the solid line; calculated values are dotted and dashed lines. (b) Interference pattern around an isolated tree showing the first five lines of minimum field strength.

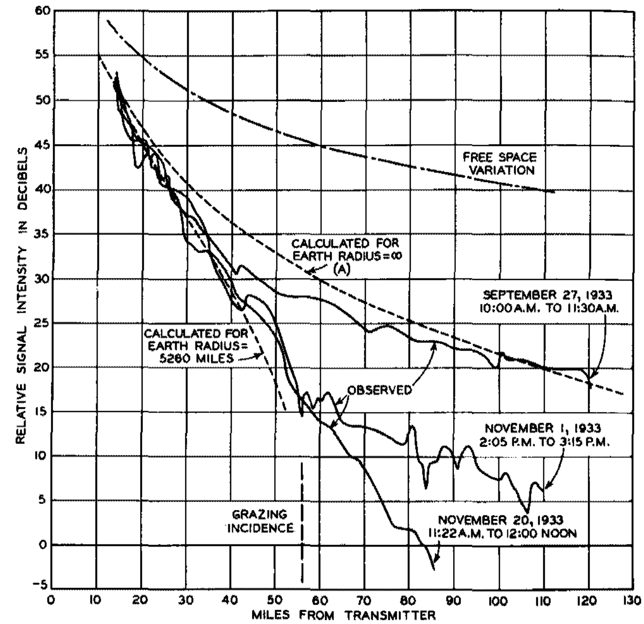


Fig. 7-7. Received signal at 4.6-m wavelength, altitude 1000 feet, September and November 1933. The path loss beyond the horizon was unexpectedly low and variable.

was obtained, especially for flights at the highest altitude, which were entirely in line of sight. Reception at lower altitudes also agreed reasonably well with theory for line of sight; but, for paths beyond the horizon, even allowing for diffraction around the bulge of the earth, unexpected large variations were observed at different times under what appeared to be essentially the same conditions. The experimenters were forced to conclude that other factors, such as the varying moisture content of the atmosphere, entered in and propagation beyond the horizon was not as predictable as they had hoped. The path loss for the below-horizon paths was very high, and, at one point, the plane was given a thorough going-over to tighten all bolts and reduce radio noise suspected to be due to loose joints. Even with the bolts tightened, no measurements were made at altitudes below 1000 feet. The Ford Tri-motor was a land plane and, when it was over water, it was desired to keep within gliding distance of the Long Island shore—the zeal for research did not exclude prudence.

About 1934, extensive measurements at 34 MHz were also taken in the urban environment of Boston.[14] Propagation was found to follow the inverse distance relationship, with interference effects due to reflections from large buildings about as expected. Considerable noise was observed from automobile ignition, but none from electric streetcars. Overhead trolley wires, however, reduced signal levels about 15 dB at intersecting lines. The motivation for this work was the growing interest in UHF for two-way use by the police and in the harbor.

2.3 Field Systems

In addition to the research on tubes and propagation, there was a considerable urge to get in the field with a working system. This was partly to learn by actual experience about some of the real-world engineering problems of telephone links in the UHF band and partly to "establish a presence." Others were eager to enter the field. There was some uncertainty as to how the Federal Communications Commission (FCC) would regard the Bell System's reentry into the domestic radio business, even if it were confined to two-way telephony and television networking. It was felt that it would add to the persuasiveness of the case for telephone company use of radio to have a system in operation. In the early 1930s, Bell Laboratories set up a relatively short experimental link for two-way telephony between the Holmdel and Deal laboratories in New Jersey. Operation at first was at about 60 MHz and later at 150 MHz. In 1934, using the techniques and types of equipment used in these laboratory installations, a radio link was established in Massachusetts across Cape Cod Bay from Green Harbor on the mainland to Provincetown at the tip of Cape Cod [Fig. 7-8].[15]

A single two-way telephone channel was provided with transmission at 65 MHz from Green Harbor and 63 MHz from Provincetown. The receivers were powered continuously and the transmitters arranged for fast start-up (one second) when an operator connected to the circuit. For the convenience of test personnel during the experimental stage, transmitters and receivers were housed in a hut at the base of the antenna poles, but were designed into simple weatherproof cabinets intended for pole mounting outdoors in commercial service [Fig. 7-9]. Identical antennas for transmitting and receiving were pole-mounted

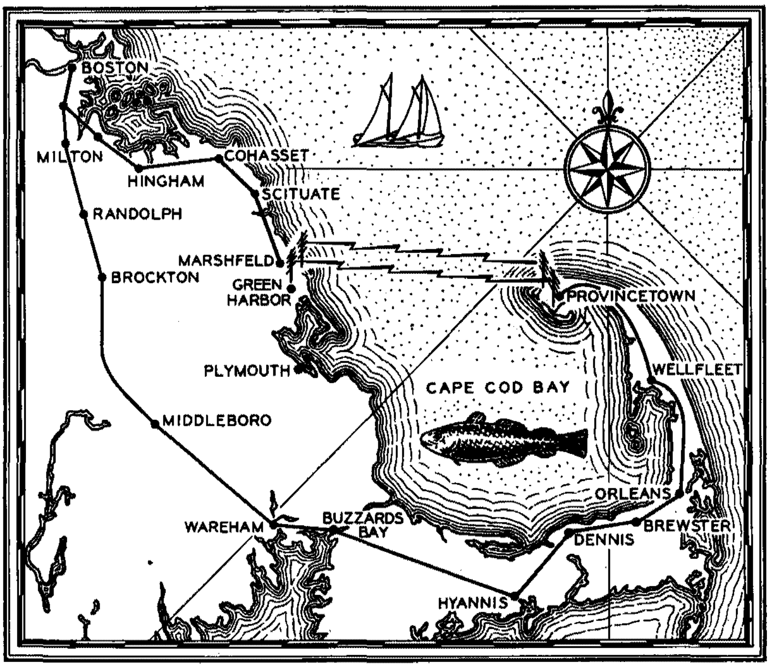


Fig. 7-8. The two-way radio link at 63 and 65 MHz between Green Harbor and Provincetown, Massachusetts, 1934.

arrays of half-wave dipoles. Each antenna consisted of four levels of two dipoles each with a similar reflector array a quarter wavelength behind the exciter elements to make the array unidirectional [Fig. 7-10]. All the equipment was designed to operate unattended and performed well in service. During a severe hurricane in 1938, when the land links were knocked out, the radio link provided the only telephone communication to Provincetown for several days.

As the virtues of UHF radio for point-to-point line-of-sight telephony became known, there was considerable interest in similar links to isolated locations in both the United States and Europe. The Bell System established radio connections to islands in the Chesapeake Bay, and RCA designed a UHF network for Mutual Telephone of Hawaii linking several of the islands. The British established similar links to small offshore islands and even one across the Irish Sea to Ireland.

These were essentially all one-hop links with a single voice circuit, but interest in multichannel, multihop relaying remained high. In 1938, RCA established a three-hop link from New York to Philadelphia using frequencies around 90 MHz capable of carrying two-way-telegraph, teletype-printer, and facsimile signals.[16] This system used line-of-sight hops with dipole-antenna arrays. Four antennas were needed at each repeater, as separate transmit and receive arrays were used for each direction of transmission. In 1939, RCA also established an experimental multihop, one-way relay system to transmit television from

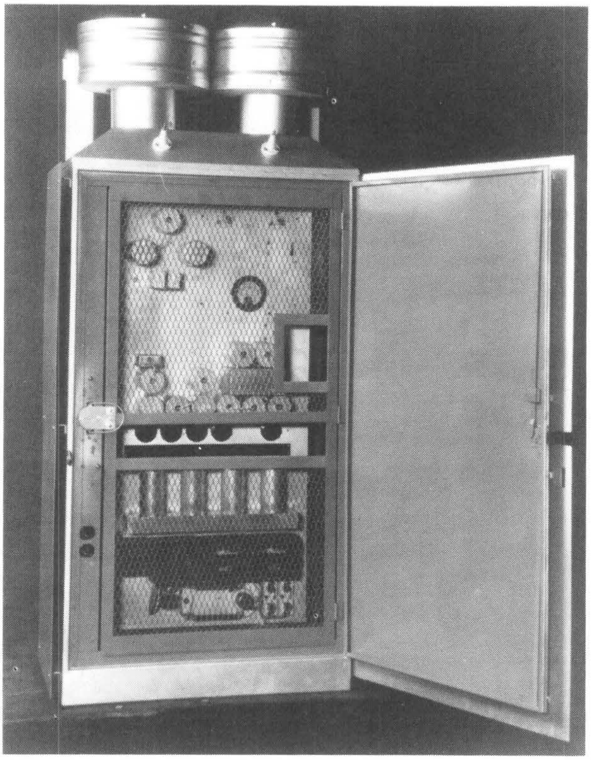


Fig. 7-9. The 65-MHz ultrashortwave transmitter for the Cape Cod Bay link. The cabinet was designed for outdoor pole mounting with no additional protection from weather.

the Empire State Building transmitter to Riverhead, Long Island.[17] Transmission from New York City was at the broadcast frequency (45 MHz) to a receiving station at Hauppauge, Long Island, then at frequencies of 460 and 474 MHz for two additional hops to Riverhead. The repeaters provided about 8 MHz of flat band. The higher-frequency hops used FM, with a transmitted power of 0.7 w, and obtained 20 dB of gain from cylindrical parabolic reflectors. Satisfactory transmission of the then standard 441-line picture was demonstrated over this link.

In Bell Laboratories, as elsewhere, the highest possible frequencies were pursued to provide the maximum advantages of UHF transmission. Offsetting this was the increasing intractability of the components as frequencies were raised. In the mid-1930s, frequencies around 150 MHz became the favored choice as a range in which most of the advantages of high frequency could be obtained and for which the apparatus was essentially in hand. In 1935 and 1936, considerable research and development effort was directed toward a repeater at 150 MHz with a bandwidth of 2 MHz and a power output of 2.5 w.[18] This was comparable to the band expected in the concurrent coaxial-

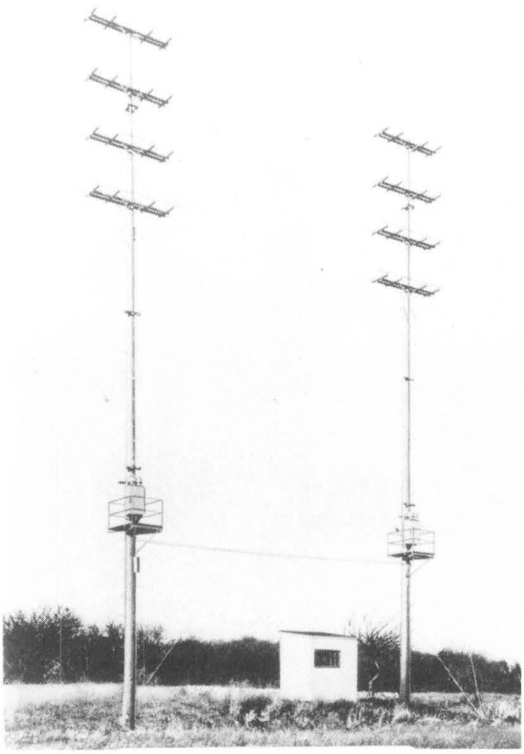
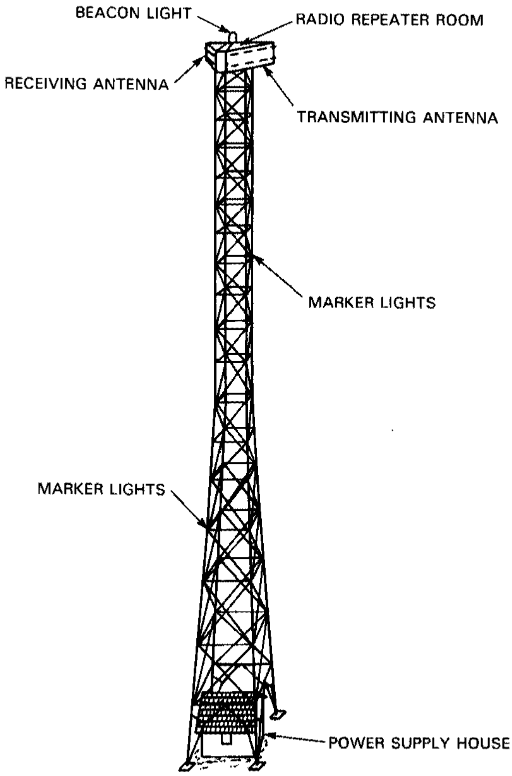
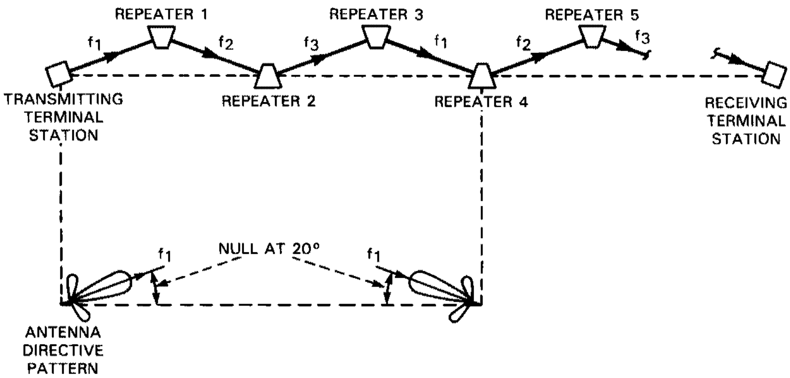


Fig. 7-10. Pine-tree antennas used on the Provincetown–Green Harbor 63- and 65-MHz circuit.

cable development and was felt to be adequate, although minimally so, for the television of that date. System concepts were developed, taking advantage of the narrower beams and pattern nulls closer in angle to the main lobe to narrow the angle of the zigzag route from the 90 degrees proposed in 1921 [Fig. 7-11]. Studies of a potential commercial 150-MHz relay system were made on routes from New York to Boston and Philadelphia to Pittsburgh. Specific repeater sites were chosen, at least as far as this could be done from detailed topographic maps, and cost estimates made. The cost estimates were found to be favorable (as such estimates for new systems are likely to be) compared to a coaxial system. The coaxial cost target, however, would not stand still. The demonstrated telephone capacity of the cable was constantly increased, and, for telephony, it continued to challenge radio for many years. Bell System interests in radio relaying were expressed in a letter to the FCC in February 1936, but relaying at UHF never got off the ground. Radio relay continued to look attractive for television, but not for multiplex telephony, where the linearity requirements always appeared to put very restrictive limits on the number of channels that could be transmitted.



(a)



(b)

Fig. 7-11. (a) Concept for 150-MHz radio-relay station, 1935. (b) Proposed pattern of angular offsets and frequency use in the 1935 150-MHz radio-relay system.

One final UHF field installation in the years immediately preceding the war was significant in demonstrating a start in multichannel operation. A radio link was established across the 26-mile-wide mouth of the Chesapeake Bay from Cape Charles, Maryland to Cape Henry, Virginia. Prior to the establishment of this link, telephone communication between these points was available only by wire lines over a 400-mile-long route via Baltimore and Washington. Radio transmission was at frequencies of 156 and 161 MHz in the two directions. The radio carriers were amplitude-modulated by the multiplexed voice signals from a Type K carrier channel bank in the range of 12 to 60 kHz. This route was opened for service with five circuits equipped in October 1941.[19]

III. MICROWAVES AND WAVEGUIDE

To telecommunications engineers trained in the middle decades of the 20th century and conscious of the struggle over many years to utilize higher and higher radio frequencies, it is surprising to learn how rapidly the early pioneers explored the spectrum.[20] The surprise, no doubt, stems in part from how little was done, until much later, to make effective use of large parts of the frequency area explored. H. R. Hertz wished to confine his demonstration of the existence of the electromagnetic waves, postulated by J. C. Maxwell, to the inside of his laboratory and chose the dimensions of his apparatus accordingly. His original experiments using spark gaps and resonant loops were carried out in the range from 6 m to 50 cm (50 to 600 MHz).* Before 1900, using basically the same techniques, experimenters had generated frequencies above 15 GHz. By the 1920s, frequencies in excess of 3700 GHz (82 μm in wavelength) had been generated and detected.[21] The gap from short radio waves through the infrared to near-visible-light frequencies had been closed. These experiments covered the range of microwaves, generally taken to mean frequencies above 1 GHz, with a good deal to spare; but to generate microwaves was one thing, to do so at usable levels and control them for communications was quite another.

3.1 Electron Oscillators

By the late 1930s, conventional negative-grid vacuum tubes were producing power at frequencies of 1000 MHz and higher. But a more rapid advance toward usable amounts of microwave power lay with tubes operating on quite

---

* Physicists often prefer to use the term *wavelength*, as it lends itself to the experiments they carry out. Communications engineers liked to think in terms of *frequency*, as their useful signals required familiar frequency bands, e.g., 4 kHz for good-quality speech and 4 MHz for television. This caused little difficulty as long as the main interest of the two disciplines lay well separated in wavelength or frequency. In recent years, their range of interests has tended to overlap, especially in such areas as optical communications. Since this is a history and telecommunications developed from lower frequencies to higher, we will generally (but not invariably) stick to frequency. The translation is simple. The velocity of light in space is 3×10^8 m per second; a 1-m wave is therefore 3×10^8 Hz, that is, 300 MHz. Prior to the 1960s, the unit *hertz* had not replaced cycles per second, and contemporary references are usually in terms of cycles or megacycles per second.

different principles. In 1919, H. Barkhausen and K. Kurz had observed oscillations at extraordinarily high frequencies in tubes with grids that were positive with respect to the cathode and plates at zero or negative voltage.[22] These were identified as electron oscillations, that is, oscillations not caused by the periodic exchange of energy between the electric and magnetic fields of capacitors and inductors, but due to the ballistics of electrons in the electric field of the tube. Their frequency was shown to be entirely due to the fields and the tube geometry, entirely independent of external elements. In the early 1920s, E. W. B. Gill and J. H. Morrell showed that such oscillations could, however, be coupled to and controlled by external circuit elements as well.[23] Barkhausen's experiments were conducted in the range from 300 to 600 MHz. In 1926, coupling a number of oscillating tubes to a Lecher wire as the tuning element generated frequencies up to 1.7 GHz.

In 1928, H. E. Hollman devised a means to modulate a Barkhausen-type oscillator and succeeded in transmitting telephone speech over a range of "500 wavelengths," presumably a few hundred meters.[24] Also in 1928, S. Uda and H. Yagi, working in Japan at frequencies from 150 to 660 MHz, modulated a Barkhausen oscillator and transmitted telephone signals over 10 km, in experiments equally noteworthy for the invention of Yagi's "director array" antenna.[25] In France in 1930, G. Beauvais transmitted telephone signals at frequencies close to 2 GHz from the Eiffel Tower to receivers up to 23 km distant, using paraboloids as directive antennas.[26] Equipment similar to this was used the following year in a link across the Strait of Dover.

The microwave link across the Strait of Dover deserves special notice because of its influence on the work of the 1930s. On March 31, 1931, two ITT affiliates, Standard Telephones and Cables Ltd. in England and Les Laboratoires Le Matériel Téléphonique in Paris, established a two-way "microray" telephone circuit of about 21 miles between stations in the vicinity of Dover, England and Calais, France. (The technical descriptions of this link are rather sketchy. A more complete description was published of the somewhat later, but apparently almost identical, link between airfields close to the original route.)[27] Transmission was at 17.4 cm (1720 MHz), produced by a Barkhausen-Kurz oscillator with a power of 0.5 w. Transmitting and receiving antennas were paraboloids about 3 m in diameter, yielding a gain of 26 dB each. The large antenna gain made transmission possible at the low power used [Fig. 7-12]. This was a noteworthy accomplishment and was widely hailed at the time, but a stampede to microwaves did not occur. Although the link was announced as ready for commercial service in 1931, it does not appear that much traffic was carried. A second link, not far from the first and using somewhat improved equipment, was set up between airfields on opposite sides of the Strait of Dover two years later.[28] Its use appears to have been restricted to teleprinter messages to advise of airplane flights across the channel. A few years later, Bell Laboratories established a similar link between Bell System buildings in the New York area, using Barkhausen oscillators at 1000 MHz.[29] In at least one test, this was modulated by a three-channel Type C carrier, but, in this case too, little follow-up occurred.

The problem impeding rapid development evidently was the very embryonic stage of the technology. The new components were difficult to use effectively. To be effective, the electron oscillators required the total emission from the

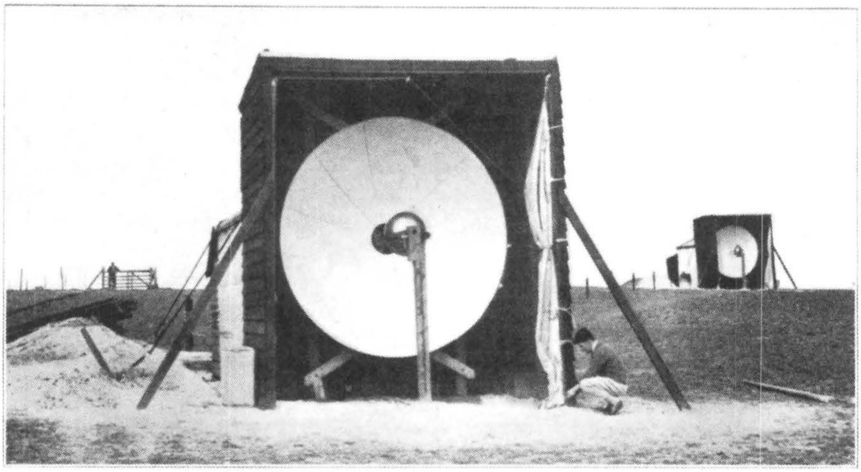


Fig. 7-12. Front-fed parabolic antennas near Dover, England for 1720-MHz link across the Strait of Dover, 1931. [*Post Office Elec. Eng. J.* **24** (1931): 153.]

cathode, unlike the negative-grid tubes, which generally operated under space-charge-limited conditions. As a consequence, the heaters were pushed hard and life tended to be short, even at rather feeble output. In addition, even short transmission lines were lossy at 1700 MHz. In the initial Calais-Dover installation, the oscillator and a tiny radiating antenna were suspended at the focus of the paraboloid, clearly not the handiest place for replacement and adjustment. Later, the oscillator was mounted behind the dish and connected to a dipole at the focus by a tapered coaxial line. This arrangement introduced some loss, and the separation between the source and the feed antenna would make tuning and matching touchy.

The Barkhausen and other electron oscillators were not easy to modulate in a useful way. Changing the electrode voltages tended to change both the amplitude and frequency of the output. The early accounts of the Strait of Dover link are silent on the method of modulation used. It was presumably some form of FM or combination of FM and AM. In 1933, the ability to produce 40-percent amplitude modulation at constant frequency by simultaneously driving two electrodes in the proper ratio was described as an improvement. The difficulty of modulating microwave tubes was to plague experimenters for many years. In 1934, experimenters at RCA avoided troublesome incidental FM by passing the radiated output beam from a 9-cm (3300-MHz) magnetron through a grid of gas-discharge tubes. The loss of the grid was varied by modulating the gas-discharge current by the desired signal—a sort of microwave heliograph. Such expedients would not have been used if any reasonable option had been available.

All this was bad enough, but perhaps the most serious shortcoming of the electron oscillators was that they were just that—oscillators. They did not have the ability to amplify, in the sense of producing an output that was an enhanced replica of an input at the same frequency. It was not until about 1935, when the brothers R. H. and S. F. Varian invented the klystron at Stanford University, that a true microwave amplifier, based on the electron-velocity-variation principle, became available.[30] The shortcomings of electron oscillators were summed up by one research team, "...all suffer from one or more serious faults when considered from the standpoint of practical use. Nearly all of the methods are wasteful of plate power. Many are insensitive in the extreme. The more sensitive are unstable, in general. Tuning is broad. The most serious faults shared by all are limitation of sensitivity, due to the fact that no amplification may be had...."[31]

For all these reasons, most researchers continued to concentrate on what was felt to be the more tractable components and promising approach at UHF. In one area of work, however, interest in microwaves remained high. There was no practical way of working with waveguides except at frequencies over 1000 MHz.

3.2 Waveguides

As frequencies were pushed higher and higher, the conventional circuit elements of lumped inductors and capacitors became smaller and smaller until they had all but vanished. Transmission lines of conventional types became very lossy, and even widely spaced, parallel, tubular, Lecher wires acted more like antennas than waveguides confining the waves. If frequencies above 500 or 1000 MHz were to be used, an altogether different circuit technique was needed. In 1897, Lord Rayleigh had shown that certain solutions of Maxwell's equations indicated that the transmission of electromagnetic waves through hollow conducting tubes was physically feasible.[32] Rayleigh had thus invented, or at least shown the prospect for, what came to be known as waveguides. By that time, however, the two-conductor concept of guiding waves was so firmly embedded in practitioners' minds that an unusual combination of circumstances was required for such waveguide transmission to be more generally recognized and proven a reality. Indeed, with very few exceptions, the communications engineers of the early 20th century did not think in terms of guided electromagnetic waves at all and tended to keep radio and wire transmission in completely separate compartments of their minds.

A waveguide is a hollow metal tube filled with a low-loss dielectric, usually air.* The cross section of the tube may be any of a variety of shapes, although rectangular tubes and circular cylinders are the most common and are of the greatest interest. Such waveguides can propagate electromagnetic energy along the axis of the tube, providing the frequency is high enough. The wavelength must be comparable to or shorter than the cross-section dimensions of the tube. Below a critical frequency, the waves are cut off and will not propagate. The waves are confined entirely to the interior of the waveguide. At the frequencies required for propagation, the waveguide is a perfect shield and there is no external field.

* There are waveguides of many other types, as all transmission lines are waveguides of one form or another. The term is used here and in what follows for the hollow conducting-tube type, its generally-accepted meaning since the 1930s.

Transmission occurs via configurations of the electric and magnetic fields, which are called modes. Many modes are possible; there is essentially no upper limit to the number that may exist, but only a few are of practical interest. Each mode has a unique configuration by which the mode is identified and which remains unchanged as the wave travels along the waveguide. In a given waveguide, each mode also has its own cutoff frequency, below which it cannot propagate. The cutoff is a function only of the waveguide dimensions and the dielectric constant.

In a waveguide of circular cross section and for other simple geometries, the electric and magnetic field lines of the modes can be described by precise mathematical formulations [Fig. 7-13]. For more complex (higher-order) modes, the cutoff frequency is higher, because more half-wavelength variations in field must occur across the waveguide cross section. The wavelength for a given size waveguide must be smaller and the frequency higher. Similarly, since the simplest mode must have at least one field variation across the waveguide to avoid being short-circuited, at least one of the waveguide dimensions must be greater than approximately a half wavelength for any mode to propagate.

Little interest was shown in Rayleigh's analysis, and nothing was done to confirm the existence of electromagnetic modes in waveguide for many years. In large part, this situation was due to the almost total lack of experimental observations. Since an air-filled waveguide six inches in diameter has a cutoff frequency of 1150 MHz for the first mode that will propagate, the lack is understandable. The common research at high frequency, even as late as the 1920s, was at frequencies only about one-tenth of that required, so there was little chance of an accidental observation that would stimulate further investigation.

The subject, however, was not entirely ignored; some theoretical work was published in the early 1900s. In 1920, O. Schriever at the University of Kiel had analyzed the propagation of waves along dielectric lines;[33] and, in 1924, J. R. Carson at AT&T rediscovered the cutoff characteristics of waveguides. In an unpublished memorandum, Carson suggested that this characteristic might be utilized to achieve a high-pass filter. The first experimental evidence came from work by G. C. Southworth at Yale University in 1920. Southworth was experimenting with high frequencies on a transmission line immersed in a trough of water when he observed some anomalous behavior. The water, with a dielectric constant of 80, reduced the electrical wavelength of his signals by a factor of almost nine compared to the wavelength in the air, thus bringing the cutoff frequency of the trough within his measuring range. Southworth observed resonances, measured wavelengths, and recognized similarities to the waves predicted by Schriever. In 1931, Southworth, then a research engineer with AT&T, remembered his 1920 observations and returned to the subject. He again began experiments with water-filled pipes; and, on November 10, 1931, made his first notebook entry proposing the use of dielectric-filled metal pipe as a transmission medium. He recognized early in his work the importance of verifying the existence of the electromagnetic modes and identifying their propagation characteristics. In March 1932 at Netcong, New Jersey, Southworth constructed water-filled waveguides six and ten inches in diameter and four

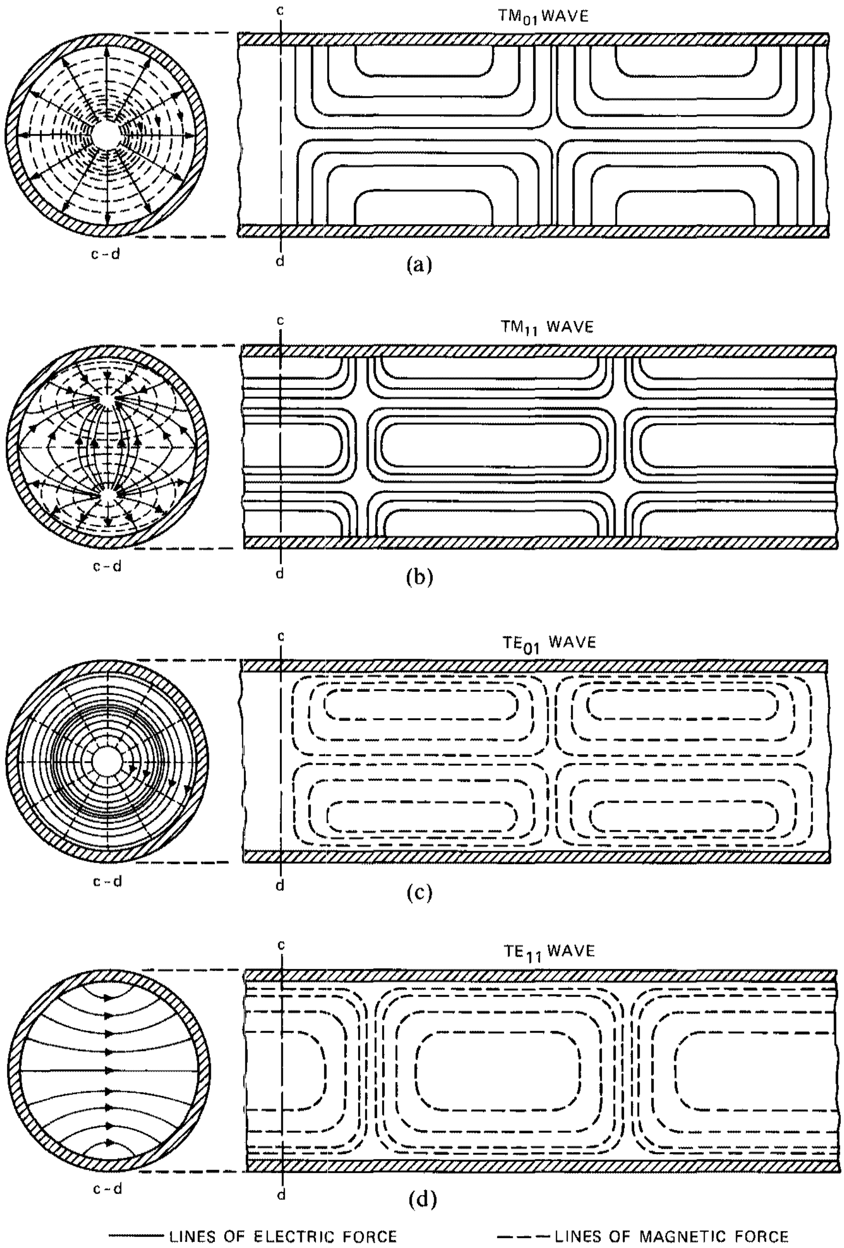


Fig. 7-13. Electric and magnetic fields in circular waveguide. [Southworth, *Principles and Applications of Waveguide Transmission* (1950): 120.]

feet long, an oscillator that would operate from 150 to 250 MHz, the means for launching and receiving the modes, and various measuring apparatus.[34]

The work was almost nipped in the bud when a memo by Southworth came to Carson's attention and he concluded, based on his earlier analysis, that the work was not practical and directed that it be terminated. However, after reviewing his own work and Southworth's early results, Carson reconsidered and, with his assistant, S. P. Mead, began active work on analysis of the transmission characteristics.

By August 1932, Southworth had experimentally identified what came to be designated the TE_{11} (transverse-electric) and TM_{01} (transverse-magnetic) modes, both by field configuration and cutoff characteristic, and made numerous measurements of waveguide phase constants and other quantities. During 1932, several theoretical and experimental memoranda had been written, and the physical reality of the guided waves was well established.

The 1932 memoranda brought the subject to the attention of S. A. Schelkunoff, who had just made important contributions to the understanding of transmission on coaxial cable, and he also began working on the problem. By the end of May 1933, both Mead at AT&T and Schelkunoff at Bell Laboratories, working independently, had issued memoranda giving theoretical propagation characteristics, including attenuation, for certain types of modes. In June, they apparently talked together for the first time, discussing certain differences in their results, and in early July, within three days of each other, issued separate memoranda, each of which included an astonishing result for the attenuation of what are now designated circular-electric waves. Schelkunoff's abstract, dated July 7, 1933, states it concisely, "Among several types of electromagnetic waves that can be shot along a hollow cylindrical conductor exists a wave with a circular line of electromotive intensity tangential to the conductor. For sufficiently high frequencies, its attenuation is inversely proportional to the 3/2's power of the frequency." (Although the credit for this discovery was disputed at the time, a compromise was reached that resulted in the first publication of the theory of waveguide transmission, including the attenuation formulae, in a joint paper by Carson, Mead, and Schelkunoff.[35] It appears that Mead, working with Carson, was probably the first to write the correct formulation and Schelkunoff the first to realize the significance of the asymptotic behavior with frequency; the compromise was a fair one.) To put this discovery in perspective, the waveguide tests and analyses came at the same time that other Bell Laboratories groups were struggling to transmit a dozen voice channels on cable pairs or to transmit a few hundred channels or a skimpy television signal over the embryonic coaxial cable. Both groups were constrained by the limited band available, by the high loss, and the fact that the loss increased with frequency. The latter phenomenon had come to be regarded as essentially a basic law of nature. Now it appeared that a transmission line of arbitrarily low loss could be achieved merely by going to a sufficiently high frequency with the equally great attraction of a correspondingly enormous exploitable band.

In the meantime, Southworth was able to demonstrate with higher-frequency oscillators the existence of some of the waveguide modes in sections of air-

Fig. 7-14. G. C. Southworth in front of the circular-electric-mode waveguide transmission line built at Holmdel, New Jersey in 1935. Southworth is holding one of the resonant chambers used for tests of waveguide transmission.

filled copper pipe four and five inches in diameter. In May 1933, he transmitted a first message, in telegraph code, over a section 20 feet long, reading, "Send Money."[36]

In the summer of 1933, Southworth demonstrated experimentally the existence of the circular-electric wave. There was great skepticism at the time as to whether the predicted low attenuation could actually be realized in practice. In fact, many questioned whether any of the waveguide modes would maintain their identity or their usefulness in traveling over long distances. Consequently, in July and August, an air-filled guide five inches in diameter and 875 feet long was built at Netcong, New Jersey, and many tests were run. The results were quantitatively inconclusive, because both the methods of waveguide fabrication and the measuring techniques were still crude. Southworth demonstrated, however, that energy could indeed be transmitted through a long waveguide, and the attenuation was not severe.

Up until this point, Southworth had been carrying out his experiments alone. In 1934, as part of the final consolidation of all Bell System research and development work in Bell Laboratories, his work was transferred to the Bell Laboratories Radio Research Department at Holmdel, New Jersey, and he was assigned the help of two engineers and an assistant. By 1936, when Southworth first published his results as a companion paper to that of Carson, Mead, and Schelkunoff, the techniques were evolving rapidly.[37] In that paper, Southworth described his experiments on the characteristics of the modes in circular waveguides and compared them with theory.

A six-inch copper pipe 1250 feet long and more carefully constructed than the earlier waveguides was built at Holmdel in 1935 [Fig. 7-14]. Using it, Southworth was able to show good agreement between measurement and theory for the attenuation of the mode with the lowest cutoff frequency (later designated the TE_{11} dominant mode). He was still unable to verify the low-loss characteristic of the mode identified by Mead and Schelkunoff (transverse-electric-zero-one, TE_{01} mode), however, because its cutoff frequency is over twice that of the TE_{11} mode and, for the six-inch waveguide, was too close to the frequency limit of his measuring equipment. As higher-frequency signal sources became available in the late 1930s, measurements of the TE_{01} mode attenuation in the 1250-foot test line indicated that, although low, it was considerably more than that calculated. The difference was thought to be caused by a longitudinal seam in the waveguide that transferred power to other lossier modes. The recognition was dawning that, for good TE_{01} transmission, a high degree of perfection in the geometry of the circular cylindrical waveguide was necessary. The theoretical and physical existence of the low-loss mode had been revealed, but many years of research and development were to be invested before the components of a

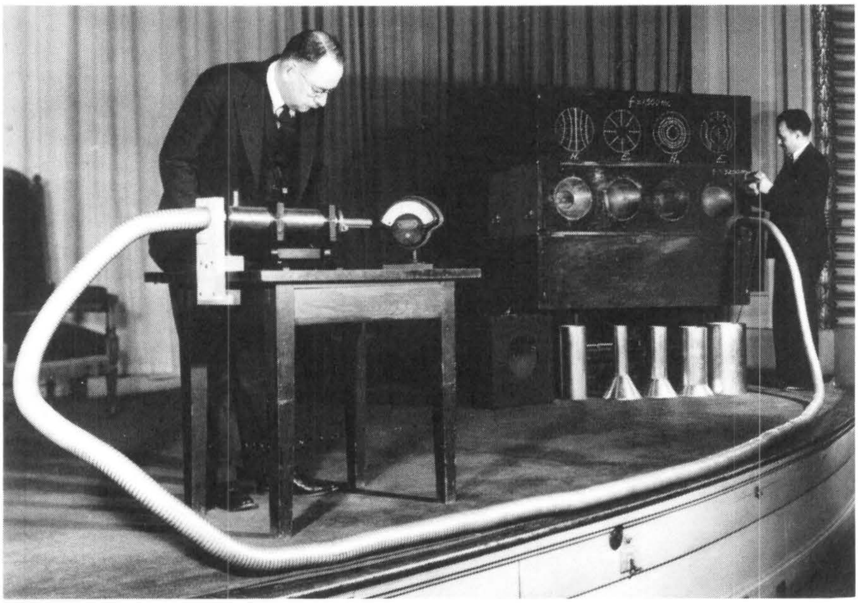


Fig. 7-15. First demonstration of waveguides before the Institute of Radio Engineers, February 2, 1938. This emphasized different modes of transmission and their respective cutoff frequencies.

practical millimeter-wave, circular-electric-mode waveguide system were in hand. Southworth and his associates gave many lectures and demonstrations to publicize the concept of waveguide transmission, but the more immediate impact of the waveguide work of the 1930s lay in its critical importance to microwave radio relay and to the wartime radar developments [Fig. 7-15].

3.3 Microwave Components

The usual circuit elements, such as capacitors and inductors, can be considered as "lumped" only as long as they are very small compared to the wavelength of the signals used. At microwave frequencies, they would have to be vanishingly small and extremely difficult or impossible to design and control. In his early work, difficulties caused by discontinuities in the waveguides had led Southworth to study their effects and to realize that they might be useful in the realization of microwave circuit elements. From his own observations and following an idea of Schelkunoff, he realized that short sections of waveguide and plates in the waveguide with small openings, termed irises, could be made to behave in waveguide circuits like capacitive or inductive reactances. Starting from this point in the mid- and late 1930s, he and his group developed techniques and an array of components that were to prove basic to the development of microwave radio, radar, and the low-loss-mode (millimeter) waveguide. These included tuned terminations, resonant cavities, the use of waveguides as radiating horns, tuned generators and receivers, wave meters, standing-wave indicators, and crystal detectors, all of which became common in the microwave art [Fig. 7-16]. At about this same time, W. L. Barrow at the Mas-

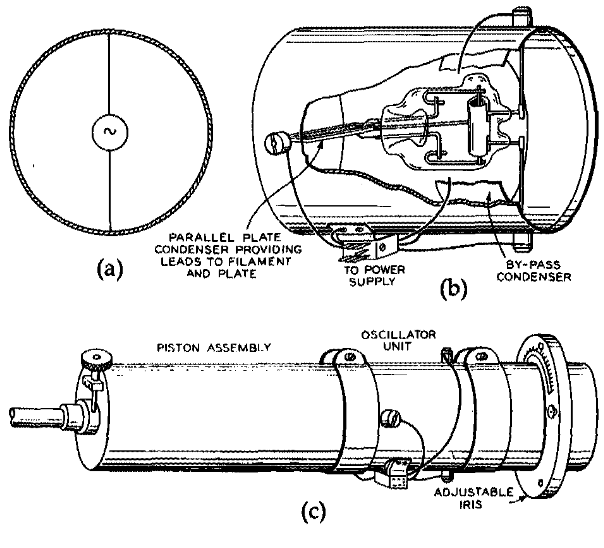


Fig. 7-16. (a through c) Adjustable microwave source of 1936. (a) Schematic. (b) Cutaway view. (c) Complete assembly including adjustable piston and iris.

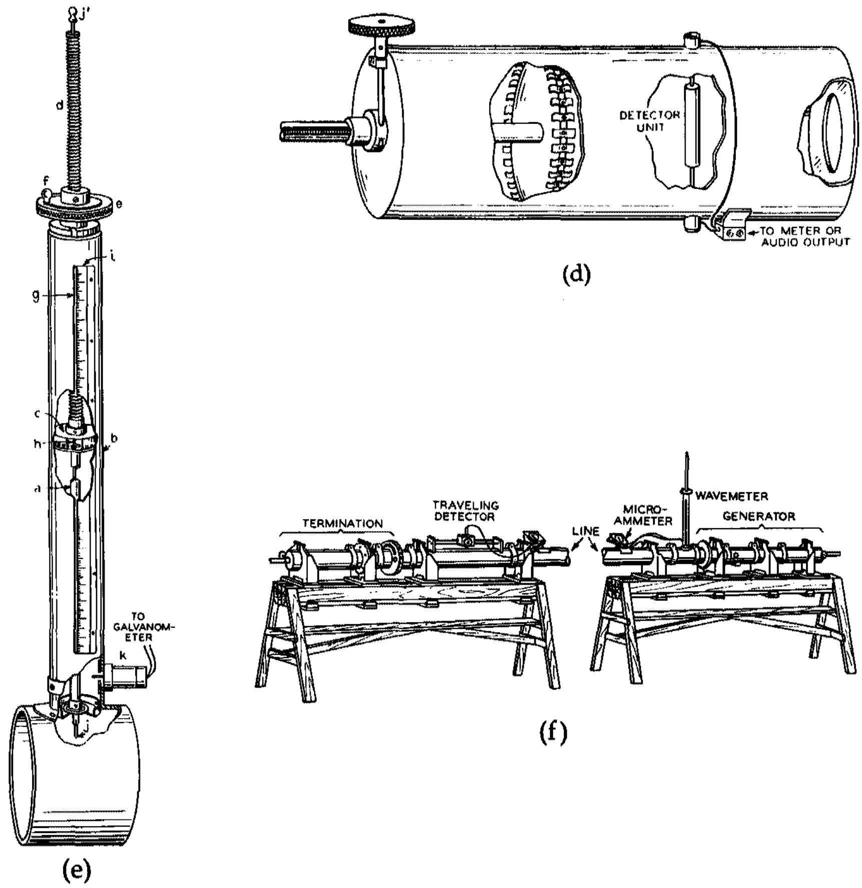


Fig. 7-16. (d through f) Early microwave components of 1936. (d) Tuned receiver based on the resonant-cavity principle. (e) Coaxial-conductor wave meter mounted on waveguide. (f) Bench mountings with typical apparatus at the receiving and transmitting ends of an experimental waveguide.

sachusetts Institute of Technology was also conducting a program of analysis and experimental work with waveguides.[38] He, too, recognized the possibility of realizing filters and other circuit components by using resonant cavities and discontinuities in the waveguides.

As early as 1936, Southworth realized that the inductance and capacitance of even short leads precluded the use of vacuum tubes with waveguides if any sort of external connection was required. If gain from a tube were to be achieved in a waveguide line, the interelectrode spaces would have to be made an integral part of the waveguide component. He obtained a patent on several ideas for accomplishing this[39] and carried out some experiments, but with only limited success. The concept was sound, however, and was successfully applied but only at a much later date.

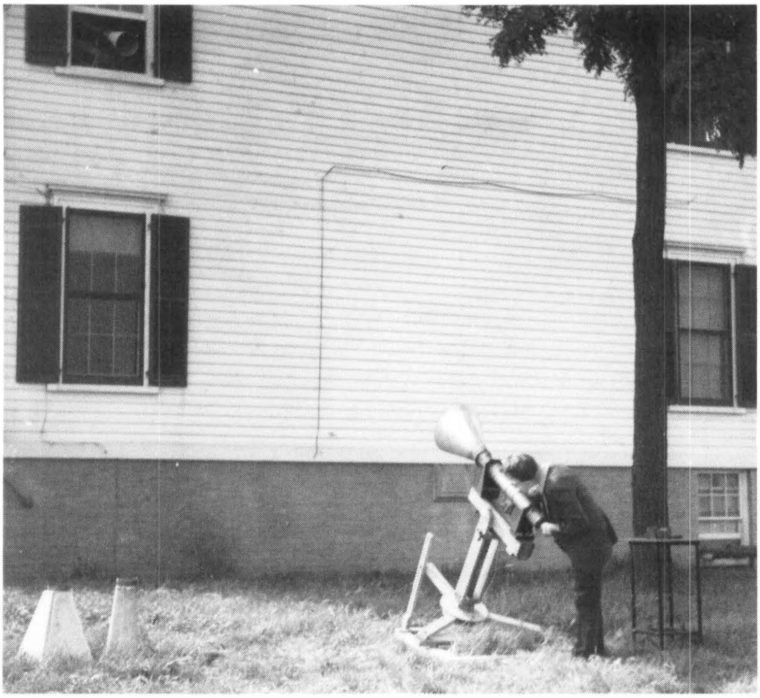


Fig. 7-16. (g) Testing microwave horn radiator at Holmdel, 1940.

Research interest in low-loss waveguides for long-distance transmission lines remained high, but there was also a growing awareness that essentially all the components needed for a microwave radio-relay system had been realized. The invention of the klystron in 1935 provided the crucially-needed amplifier at microwave frequencies.[40] A "microwave party" emerged, with Southworth as its spokesperson, urging the Bell Laboratories management to redirect its radio-relay effort to the higher band. In addition to the klystron amplifier, they were able to point to an impressive array of components: (1) receiving and transmitting antennas in the form of waveguide horns of moderate size with 30-dB gain, (2) low-loss waveguide transmission lines, (3) waveguide filters to subdivide a broad microwave radio-frequency (RF) band into subbands of manageable size, (4) silicon-crystal detectors and frequency converters, (5) intermediate-frequency (IF) amplifiers with bands of 10 to 20 MHz to provide most of the gain required, (6) silicon-crystal up-converters to shift the amplified signal to a new microwave frequency, and (7) klystron velocity-variation microwave amplifiers to raise the RF signal to a suitable level for transmission. In addition, a considerable array of measuring and monitoring equipment was available to assure that the equipment met the objectives of its designers and was performing as intended. Over 100 patents had been issued, with most of the ideas reduced to practice. In response to this and to the rapidly changing technology in other areas, a committee consisting of A. L. Durkee, W. M. Goodall, and A. A. Roetken, with J. C. Schelleng serving as advisor, was formed in mid-1940 to study the options and recommend the most promising approach to a broadband relay system.

3.4 The First System Proposal

The result of the study was by no means a foregone conclusion and the committee gathered data, deliberated, and analyzed costs for several months. A great deal of effort had been invested in the UHF work and most of the components were in an advanced stage of development. By contrast, the microwave components were still somewhat embryonic. The feelings of many people were expressed by one anonymous observer: "It has frequently been remarked that the radio-relay system is a will-o-the-wisp which always recedes to a higher and higher frequency range just as we expect it to materialize." Nevertheless, by January 1941, the committee came out in favor of the microwave option. Two factors were decisive. First, there appeared to be severe ultimate limitations on any UHF system. Television relaying was certainly possible, but the linearity of available repeaters limited the capability for multiplexed telephony to only 50 channels for 300 miles. The long-term prospect appeared to be for no better than 50 to 100 channels for much longer routes. Because of the relatively low frequencies, towers would be very high and correspondingly expensive. UHF radio simply did not look competitive to the rapidly evolving coaxial cable technology. Second, by contrast, the highly directive, high-gain

TREND OF RADIO PERFORMANCE CHARACTERISTICS

		1915	Early 1920's	Late 1920's	1930's	1940's
		Transatlantic Radio-Telephone Demonstration	Ship-Shore Experiments Broadcasting	Transoceanic Service	F. M. Television Virginia Capes Multiplex	Radio Relays and ?
	Ideal Requirements for Wire Type Performance	5000 Meters (60 kHz)	500-200 Meters (0.6-1.5 MHz)	60-15 Meters (5-20 MHz)	7-2 Meters (40-150 MHz)	15-3 Centimeters (2000-12,000 MHz)
Directivity	Sharp	None	None	Some	Moderate	Sharp
Noise	Low and Steady	Severe Static Variable	High Static Variable	Moderate Static Man-Made Interference Important	Low Static Man-Made Interference Serious	None (as yet)
Transmission Disturbances	Small and Slow	Small	Marked Rapid	Severe Rapid	Echoes Troublesome for Television	Small and Slow
Frequency Space Available	Unlimited	Picayune	Meagre	Small	Moderate (Crowded)	Large

Fig. 7-17. Trend of radio performance, indicating favorable characteristics of microwave frequencies for radio relay.

microwave antennas permitted much lower towers and low transmitter power. The enormous band held the promise (but only the promise) of being able to transmit several hundred telephone channels.

Like most retrospective examinations of such issues, it all seems clear and obvious now, but it did not appear so at the time. The committee was most uncomfortable with the problems of multiplexed telephony, since they were well aware of the very demanding linearity requirements it imposed on any repeater, and they were by no means certain about what could be accomplished in this respect with a complex microwave repeater. Various forms of RF modulation were examined with this problem in mind. The committee leaned to FM, but did not exclude single-sideband AM. Basically, they opted for a more open-ended, albeit a much less explored, technology over the much better-known but more narrowly limited field at UHF. R. Bown, the director of research, summed it up in a simple chart [Fig. 7-17].

An agreement was reached to direct work toward a two-link experimental system at 3000 MHz, between the Deal and Holmdel laboratories in New Jersey, to be completed by the summer of 1942. The unhurried schedule was undoubtedly established to permit further study and research and design work on the more difficult components. Some conceptual station and system layouts

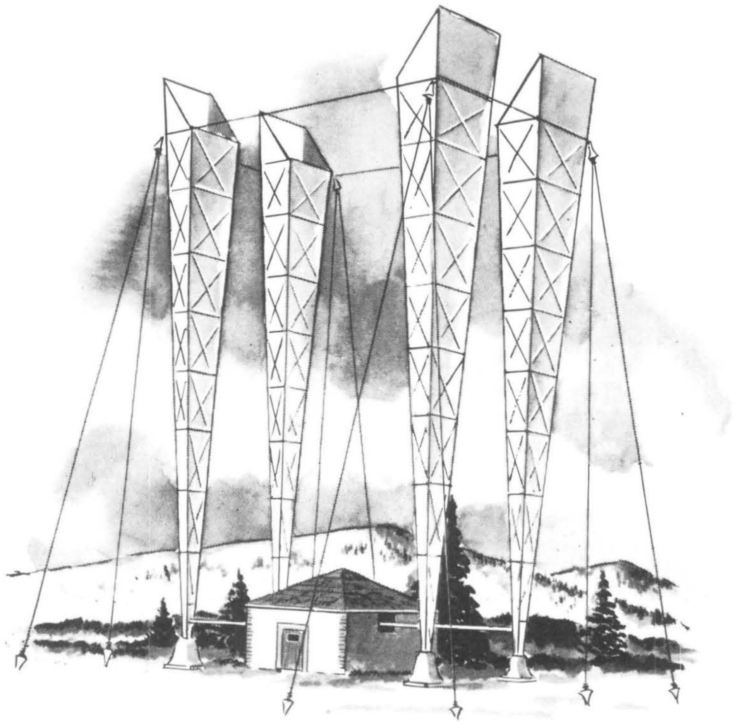


Fig. 7-18. A conception of microwave-radio repeater station, 1941.

were developed early in 1941 that still look surprisingly like a modern radio-relay station [Fig. 7-18]. By the end of 1941, most of the equipment had been built and partially assembled as planned. The war interrupted work, and effort was redirected to more urgent wartime projects. The microwave work was to prove of tremendous importance during the war, especially in its application to radar, but it was to be late in the war before any effort was again directed toward civil communications work. (See another volume in this series, *National Service in War and Peace (1925-1975)*, Chapter 5, Section VI.)

IV. RESEARCH TO DEVELOPMENT—THE TDX SYSTEM

While war work pushed almost all other considerations aside, the problems and prospects of the postwar world were not totally ignored. In mid-1943, some thought was given to how the radio-relay work might be resumed when the war situation permitted. Early that summer, Bell Laboratories proposed to AT&T an initial route from New York to Boston. In conjunction with the existing New York–Philadelphia coaxial line and its extension, the route would extend television transmission from Washington to Boston and permit a comparison of the radio and coaxial systems. The anticipated substantial growth in telephone traffic might also be served. In July 1943, a high-level committee, including technical executives of AT&T and the Long Lines Department, as well as Bell Laboratories, was established to consider the postwar prospect for systems to transmit television and multichannel telephony. It appears from the charter that wire lines and coaxial development were the main concern, but one of eight major subjects to be addressed was repeatered radio. The initiative on this item was assigned to M. J. Kelly, then research vice president of Bell Laboratories. It was clear that radio relay was to be a high-priority postwar research and development project.

Events then began to occur rapidly. Planning meetings, specifically on the radio project, were held before the end of July. By late October, the parent committee had received a tentative plan for a postwar radio-repeatered project from Bell Laboratories and agreed unanimously in favor of authorization to proceed as early as circumstances permitted.

More concrete plans were formulated by the end of 1943. All types of signals were to be transmitted. Multichannel telephony received more attention than in earlier studies. The success of the military AN/TRC-6 radio system in transmitting eight channels by pulse-position modulation was noted, but hopes were held for as many as 500 channels in a commercial microwave system. As noted in a conference report, television quality "equivalent to that of the coaxial used for the 1940 convention" was established as an objective. A repeater station with parabolic-dish antennas and a repeater with most of the gain at an IF of 60 MHz were proposed. Through use of the same local oscillator for the frequency conversion down to IF and back up to RF, a shift of 120 MHz was to be realized between the received and transmitted RF signals [Fig. 7-19].

Little specific laboratory work was possible in 1943, but, in December, a Long Lines Department engineer was brought in to the group to start site surveys on the New York–Boston route. Application for construction was made to the FCC, and a permit for the first station issued on March 16, 1944. The

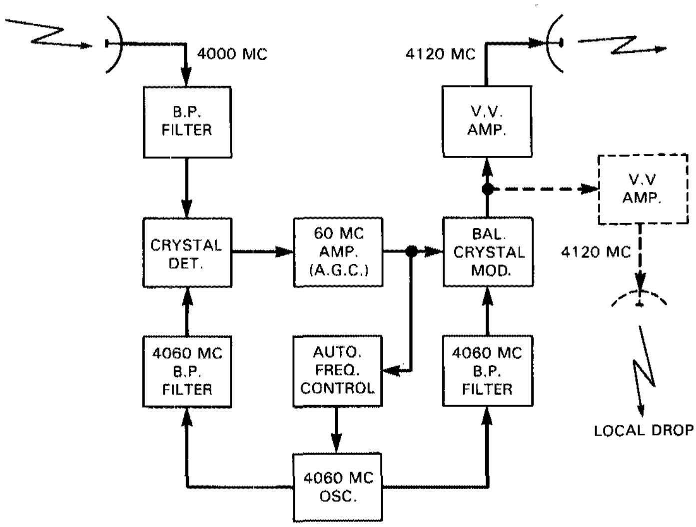


Fig. 7-19. Schematic concept for an IF-type microwave radio-relay repeater, 1943. A common local oscillator 60 MHz above the received signal for both up- and down-frequency conversions realized a shift of 120 MHz between the received and transmitted RF signals.

next day, AT&T publicly announced its plan for the New York–Boston radio-relay route to furnish postwar television networking as well as telephone service. Internally, AT&T pressed hard for a completion date before the end of 1946. They wanted a specific capability and date in connection with the talks on television networking, then under way with the National Broadcasting Company. AT&T was also well aware that several others were gaining microwave background as a result of the wartime radar work and were eager to enter the field. The Bell System management did not want to be anticipated in, and perhaps even excluded from, what looked to be one of the most promising lines of communications development. That their concern about rivals was well founded is evidenced by the fact that RCA, the Philco Corp., and others did indeed establish microwave links as soon as, and in some cases, even before, the war was over.[41] Bell Laboratories was understandably diffident about any hard commitments while wartime work was still the most urgent priority but agreed to do the best it could.

Until this time, the work had been carried on under general research funding. In late-1944, as a group dedicated to the project was formed, formal project funding was approved for two million dollars to be about equally divided between research and development expense and the construction of the eight-hop route. G. N. Thayer of Bell Laboratories, whose entire background had been in development, was transferred to research to head the project, and the initial staffing was made up of people from both research and engineering development. The first route was to be experimental, but was to serve as a prototype for a completely commercial system expected to follow immediately afterward.

4.1 Research

Much good work on the essential components had been accomplished in the late prewar years, but much remained to be done. By making the first route a prototype for commercial systems, something beyond the pure pursuit of understanding was introduced. At the same time, the task was anything but a straightforward development project. The research and development team had to expand the base of essential knowledge and apply it very rapidly in the realization of equipment suitable for the large-scale trial. The crucial applied-research effort was under the direction of H. T. Friis. (Friis, an outstanding researcher and radio engineer, in addition to innumerable personal contributions, served as a mentor to a generation of Bell Laboratories radio scientists and engineers. He was as much loved and respected for his human qualities as for his technical acumen. To those who knew him, the enormous microwave-radio network is a fitting monument to his memory. This section is based almost entirely on his report of the work.)[42]

4.1.1 Propagation

As always, good understanding of propagation was essential. Several test paths, some of which had been used during the war to study microwave propagation for radar projects, were equipped and statistical data gathered to be used as the basis for engineering the links of the trial route [Fig. 7-20]. Past experience had shown that, for frequencies below 1000 MHz, it was necessary to account for ground reflections. It was quickly established that, for microwave line-of-sight paths with adequate clearance, scattering and absorption by typical terrain results in substantially free space propagation. But all antennas, even

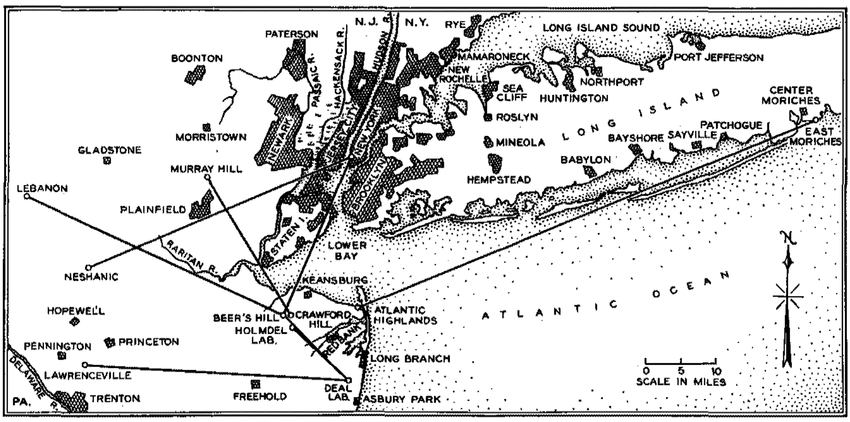


Fig. 7-20. Map showing principal paths tested for microwave propagation, 1945–1947.

highly directive microwave antennas, radiate energy to some extent in all directions; what constituted adequate clearance? It was proposed that sufficient path clearance was provided if the first Fresnel region was free of obstacles. ("The first Fresnel region for a given transmitter and receiver is bounded by points for which the length of the path, transmitter to point to receiver, is greater by one-half wavelength than the direct path from transmitter to receiver: its cross section by any plane perpendicular to the direct path is the first Fresnel zone in the sense used in optics.")[43] This useful rule of thumb was adopted and widely used as the basis for laying out microwave-radio paths in the years that followed. The application of this rule to a path between two Bell Laboratories locations in New Jersey for 100 MHz (3 m) and 10 GHz (3 cm) vividly illustrates the optical line-of-sight nature of microwave radio [Fig. 7-21]. The report of the work noted, however, that conditions such as flat desert terrain or over-water paths might prove to be troublesome as a result of reflections even at microwave frequencies.

The work shed much light on atmospheric refraction effects, including fading. This phenomenon, which continues to plague microwave radio, is altogether different from the fading due to ionospheric changes over long distances at lower frequencies. It was shown conclusively to be due to inhomogeneities in the atmosphere, especially prevalent during clear, calm summer nights when temperature-inversion layers are apt to form. It was observed to be worse for higher frequencies than for lower and, on the 40-mile New York–Neshanic, New Jersey path, to be twice as great in dB for the entire path as on either half section.

It was observed that deep fades usually did not occur simultaneously at widely separated frequencies, a fact made use of later in frequency-diversity switching to maintain continuity of transmission during fading. The phenomenon of "earth-bulge" fading was also observed, in which the normal curving of the path by refraction is reversed, and waves that would normally clear the terrain are bent into the ground. Up-fades, that is, increases in signal strength above free space levels, due to trapping or ducting in a layered atmosphere were also observed. Fading was found to be the same on vertical and horizontal polarization. Measured rain attenuation showed that little difficulty was to be

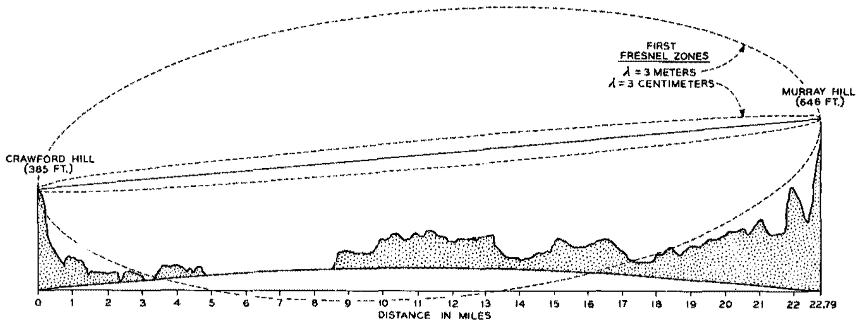
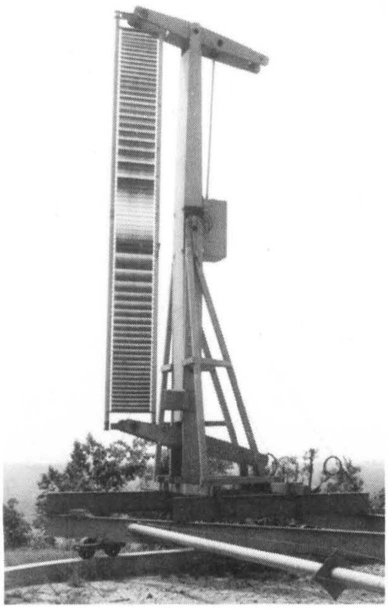


Fig. 7-21. Profile map of Murray Hill–Crawford Hill paths showing first Fresnel regions for wavelengths of 3 m and 3 cm.

(a)



(b)

Fig. 7-22. Antenna with high resolution in the vertical plane used to measure the angle of arrival of microwaves. (a) General view. (b) Antenna face showing element array.

anticipated at 4000 MHz, but that it could be a serious problem for frequencies of 10,000 MHz or more.

Many measurements were made with high-resolution antennas to determine the angle of arrival of wave fronts and did much to explain the multipath nature of fading [Fig. 7-22]. This work also provided the data establishing the maximum useful gain and directivity that could be achieved before the changing angles of arrival reduced the advantage of highly directive antennas.

The researchers were led to conclude that "the mechanism of microwave propagation is certainly a complicated one, and a considerable amount of experimental work in the fields of radio and meteorology will be required to unravel it."[44]

Despite the complexity, the essential data were obtained, at least in statistical form. It was concluded that clear line-of-sight paths of 25 to 30 miles, with allowance for fading from +10 dB to −20 dB, would provide reasonably high continuity of transmission, and this range was made the basis for the repeater for the experimental route. With the spacing, fading range, maximum practical antenna size and gain, and transmitter power (about 1 w) established, the total repeater gain required between antennas was about 75 dB. With the resulting power levels and noise at the receiver input about 14 dB above thermal noise (about that being realized in radar receivers), the signal-to-noise ratio appeared adequate for systems at least several hundred miles long.

4.1.2 Antennas

The front-fed parabolic-dish reflector was a common antenna, widely-used for wartime radar, and consideration was given to its use for the experimental system. It had, however, a number of serious shortcomings. With a typical front feed, a portion of the transmitted signal was reflected back into the feed line, causing a poor impedance match. Electrical isolation between paraboloids in close physical proximity was poor, and they were difficult to protect from the effects of ice and weather [Fig. 7-23]. A different type of antenna that overcame all these disadvantages, the horn-reflector antenna, had been invented by Friis and A. C. Beck at Holmdel,[45] but a shielded horn-lens design was chosen instead for the early installations [Fig. 7-24].

As Southworth had demonstrated, a horn with a large aperture matches free space quite well, and the slight mismatch at the throat can be tuned out over wide frequencies. The emerging wave can be focused to the plane-wave front, needed for high directivity and gain, by either a parabolic reflector or by a microwave lens [Fig. 7-25]. Both the horn-lens and horn-reflector antennas provided good impedance, high gain, and high isolation. The principal argument in favor of the horn-lens antenna was its relative tolerance to warp or twist. Within reasonable limits, these would not significantly impair its beam-forming properties, whereas in any reflector antenna, a departure of the reflecting surface from a true paraboloid by as much as ±1/8 inch would be damaging.

In view of the fact that the horn-reflector antenna was likely to be a quite rigid structure and the extensive wartime experience fabricating microwave reflectors, many for use at higher frequencies than 4 GHz, it is a little difficult

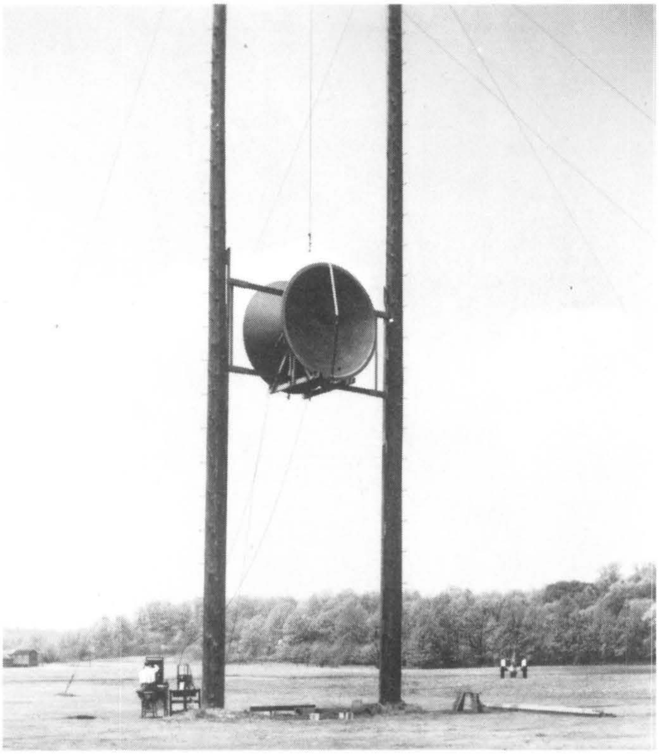


Fig. 7-23. Measuring the back-to-back coupling of two 10-foot parabolic-dish antennas.

to understand the basis for the decision in favor of the horn-lens antenna. It was perhaps of greater significance that the horn-lens antenna could be more or less hand-assembled, while the horn-reflector antenna would require considerable and time-consuming factory tooling. Whatever the basis for the decision, the horn-lens antenna was chosen and used successfully in many of the early installations, but the superb properties of the horn-reflector antenna eventually won out. Its offset feed receives little reflected energy, providing an almost perfect match; it is extremely broadband; single antennas have been used for simultaneous transmission at 4, 6, and 11 GHz; it can be highly isolated from nearly antennas; and it can support orthogonally polarized waves without interaction. The horn-reflector antenna was adopted in almost all high-grade radio-relay systems in later years.

4.1.3 Filters

A recurring dream of the early radio-relay researchers was to discover or devise a truly broadband RF through-repeater, one that would accept the entire

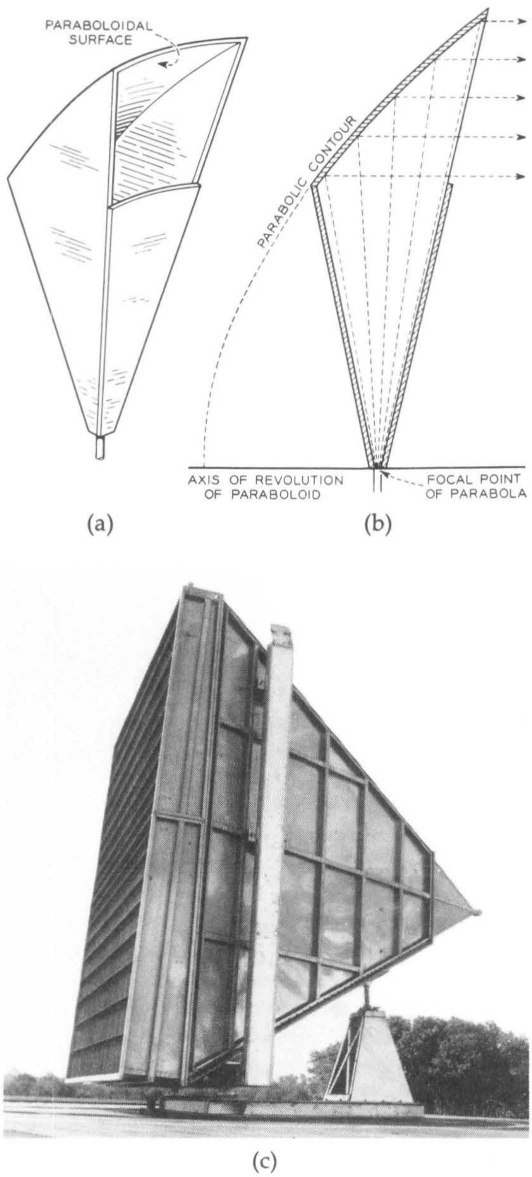


Fig. 7-24. Principle of horn-reflector antenna. (a) Three-dimensional view. (b) Cross section. (c) Shielded metallic horn-lens antenna.

available band, amplify it, and send it on its way. What was sought was a "radio line" analogous in operation to a medium such as repeatered coaxial cable. There was little difficulty in handling a ten-percent band (400 MHz at

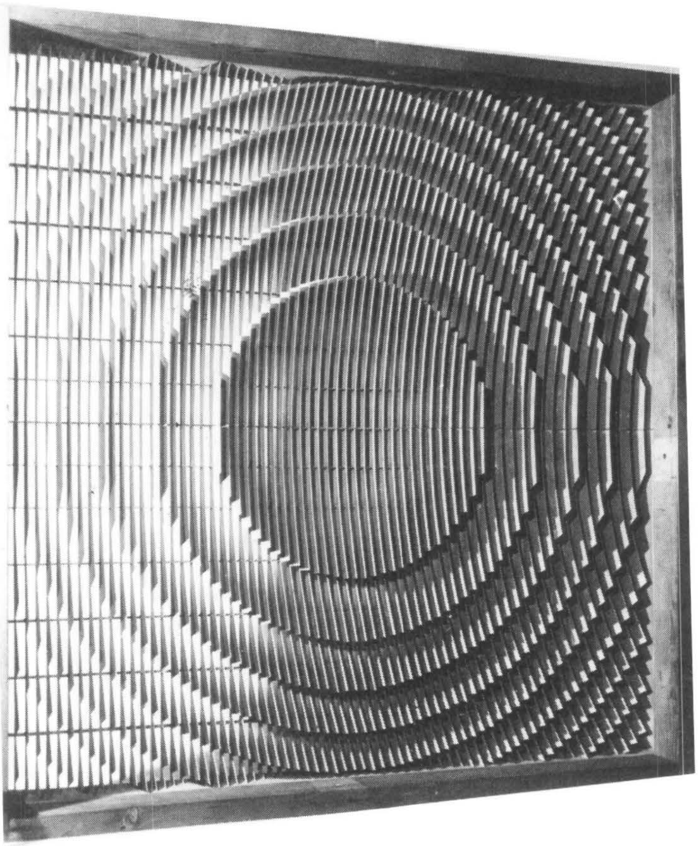


Fig. 7-25. Internal views of a microwave lens for the shielded metallic horn-lens antenna.

4000 MHz) in a single antenna, but the necessary broadband RF amplifier stubbornly refused to appear. In retrospect, it was just as well. Even the earliest workers recognized that a frequency shift was essential to prevent excessive coupling between a transmitter and a sensitive receiver on precisely the same frequency in close physical proximity. Radio, being an open medium, does not provide the means for the high degree of shielding possible between the input and output of a cable repeater. But there were, in addition, other good reasons for dividing a broad band into segments and making the portions individually accessible.

From the outset of the microwave system work, effort was focused on an IF-type repeater. In this repeater, a microwave signal, occupying only a portion of the total band, was separated from other signals in the same band and translated to an IF frequency around 60 MHz, where most of the required gain was realized. After amplification, the signal was translated back to a different microwave frequency, amplified, combined with other microwave signals in the same band, and transmitted through the antenna. The microwave route consists of completely separate sets of equipment, including separate antennas, for the opposite directions of transmission.

The microwave band for the experimental system was subdivided by filters that grew out of the prewar waveguide work. Such filters could, in principle, be paralleled, just as the filters to stack up and divide the 12 voice channels of the basic 48-kHz group band were paralleled in the channel bank. Since only two radio channels were to be used in each direction for the planned system, a simple "Y" based on this principle was adequate for the experimental system [Fig. 7-26]. However, it was of the highest importance to make the maximum use of the band available, and multiple radio channels on a single route were anticipated in later systems. For several channels, paralleling filters would have become extremely cumbersome; instead, an ingenious constant-resistance, microwave-channel dropping and combining array was invented [Fig. 7-27].

The key element was a waveguide equivalent of the hybrid coil, which was so useful at lower frequencies. This structure [Fig. 7-27(a)], dubbed a "Magic T" from one of its forms used in radar, was invented at Bell Laboratories by W. Tyrrell. This useful device found a host of applications, notably in balanced detectors and frequency converters, as well as in the channel filters. In the constant-resistance filter, it was used in the form as shown in Fig. 7-27 (b) and schematically in (c) and (d). An array for dividing a 400-MHz band into five channels is shown in Fig. 7-27 (e). An identical type of structure could be used to combine channels after passage through the separate repeater-amplifiers.

Channel-separating filters provided adequate, but only moderate, discrimination against channels at the adjacent microwave frequencies. Additional waveguide filters were used in the dropped arm to improve selectivity, and much more discrimination was provided in the IF amplifier. In addition, alternate frequency slots for transmitting and receiving and generous guard bands were proposed (and later used in practice) to ease the filter-design problem [Fig. 7-28].

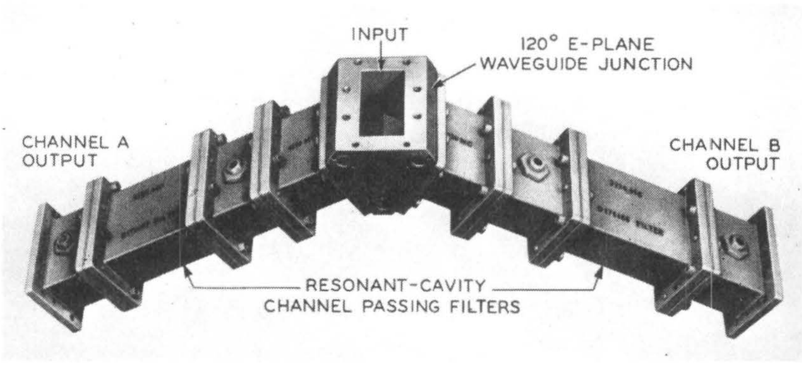


Fig. 7-26. A branching filter for the 1947 experimental radio-relay system.

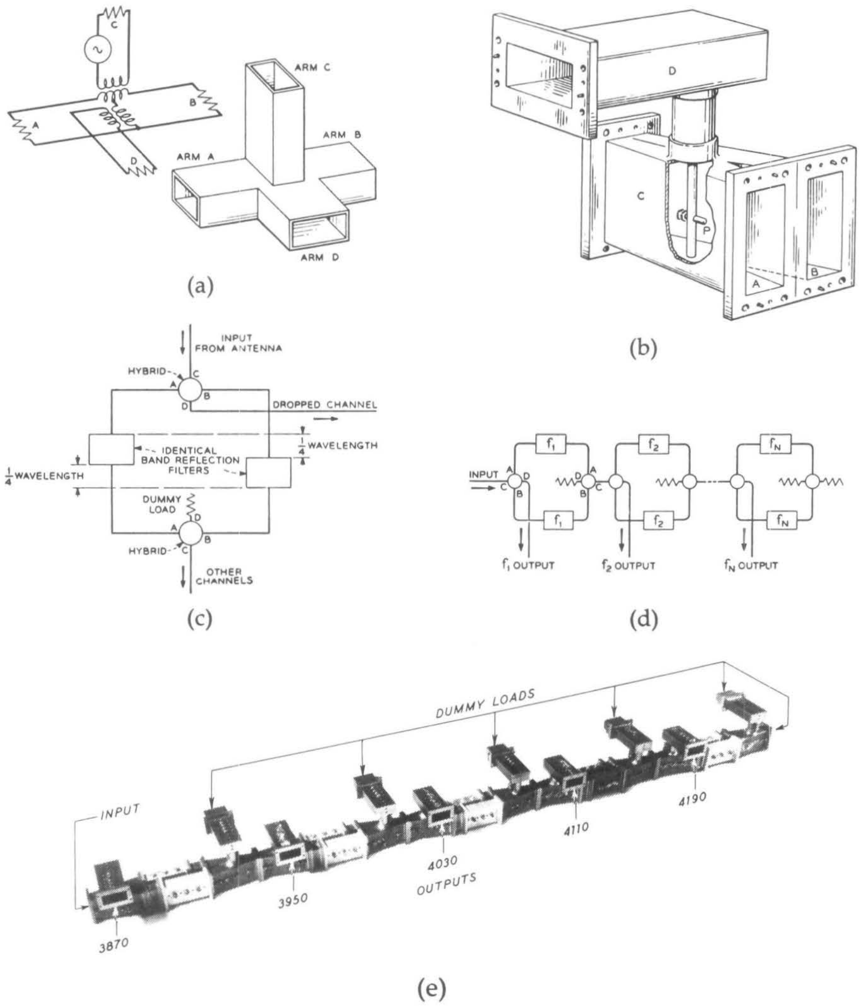


Fig. 7-27. Constant-resistance microwave branching network. (a) Diagrammatic representation of the waveguide hybrid, *right*, with the equivalent wire circuit, *left*. Energy into port (arm) C was equally divided between ports A and B. If all ports were properly terminated, no signal propagated through D. (b) The waveguide hybrid as developed for the constant-resistance branching network. (c) Each section of the network consisted of two hybrids and two band-reflection filters with one-quarter-wavelength difference in spacing. The reflected signals re-entered ports A and B, but with opposite phase. The reflected waves were canceled in port C and summed in port D, the dropped channel port. All other frequencies not reflected passed to the output hybrid, combined in phase, and passed on to filters for other channels. Any slight signal due to unbalance appeared in port D of the output hybrid and was absorbed in the resistive load. (d) Schematic of the branching network consisting of one or more sections connected in series. (e) Photograph of a five-section branching network.

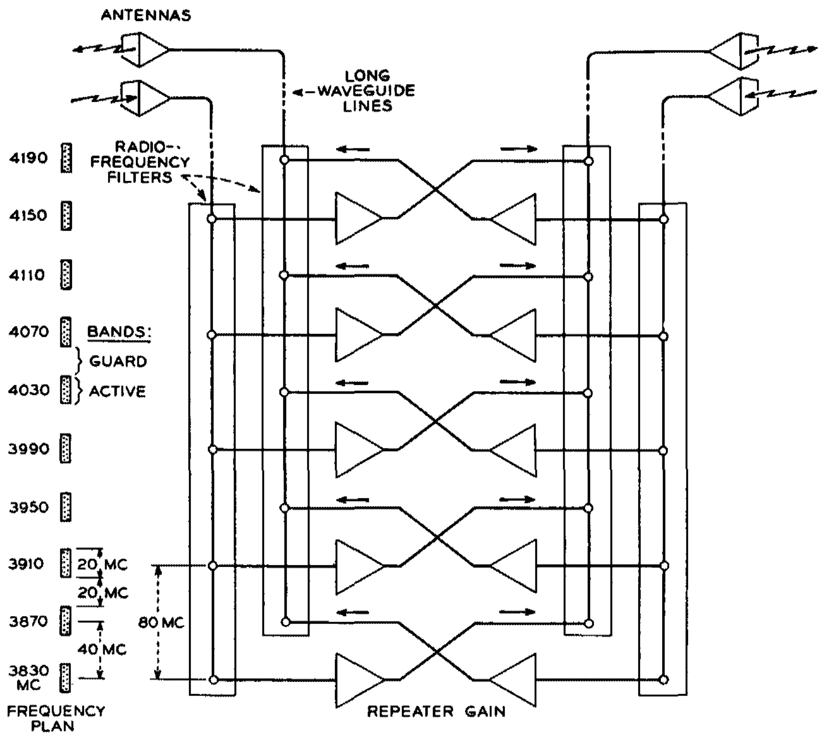


Fig. 7-28. Schematic diagram of a five-channel radio-repeater station proposed in 1948.

4.1.4 Receiving Converters and IF Amplifiers

Unlike earlier carrier systems, there was no low-noise line-frequency (that is, microwave) amplifier available for radio relay. The receiving frequency converter and input to the IF amplifier, therefore, occupied an especially important position in a microwave system, as it was at this point that the signal was at its lowest level, and the system signal-to-noise performance was determined. Considerable experience had been gained from similar converters developed for radar during the war, and with much additional intensive work, they were adapted to the radio-relay application. The converter developed was of the balanced type, using a waveguide hybrid and low-noise silicon point-contact rectifiers [Fig. 7-29]. By mixing the incoming signal and a tone from a local microwave-beat-oscillator offset from the received signal by the intermediate frequency (65 MHz in the experimental system), a difference frequency in the IF range having all the message signal components, was generated. This signal was then amplified in a multistage IF amplifier that provided the main block of gain required, as well as automatic gain control to offset the fading variations in the input signal and maintain a constant output.[46,47,48]

The balanced structure greatly reduced the beat-oscillator signal in the conjugate waveguide arm; but, as in any nonlinear converter, the harmonics and

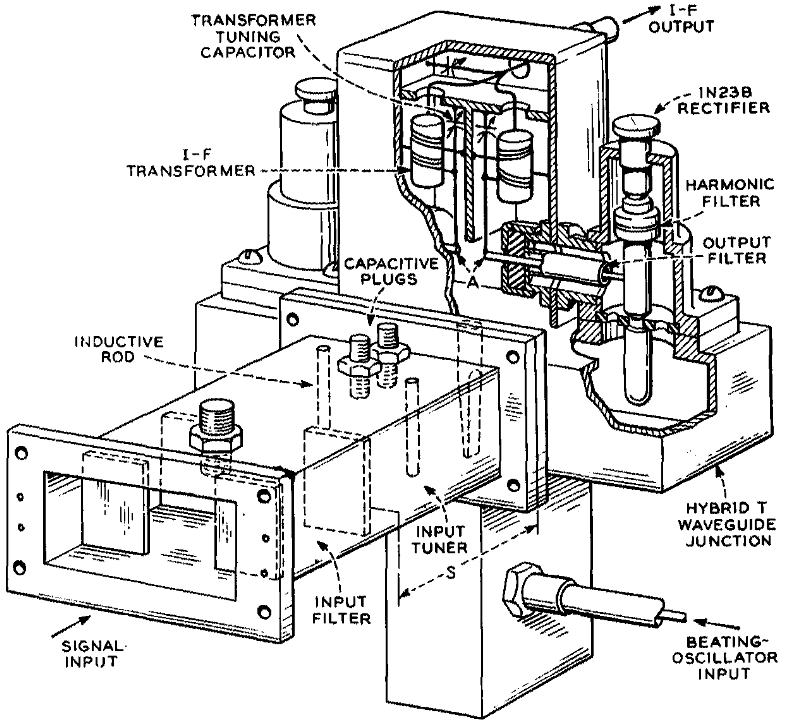
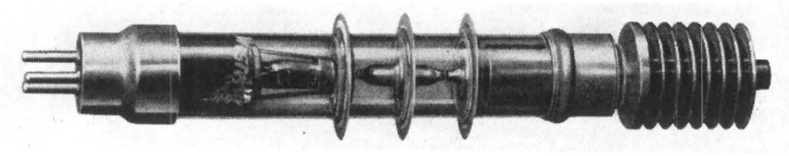


Fig. 7-29. Receiving converter for translating 4-GHz microwave signal to 65 MHz IF in experimental radio-relay system, 1947.

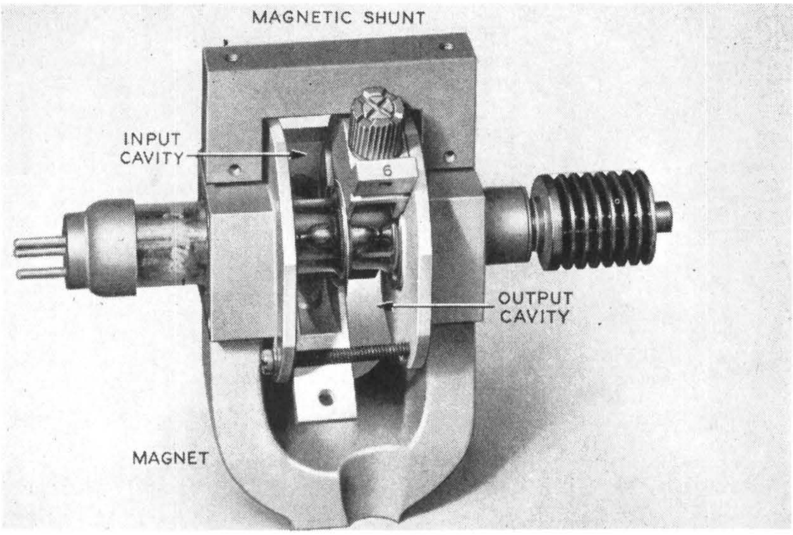
sums of the applied frequencies were generated, in addition to the desired difference product. The impedance presented to these frequencies affected the converter performance, as well as the impedance encountered by the desired frequency. The so-called image signal, a microwave frequency at the difference between the beat oscillator second harmonic and the received signal, in particular, had to be carefully handled for good performance. The last stage of microwave channel filtering was closely associated with the converter to reflect the image signal back into the converter in proper phase for optimum results. The welter of frequencies and the wide range to be considered, from 65 MHz to microwave, made the converter a decidedly intricate device.

4.1.5 Microwave Amplifier

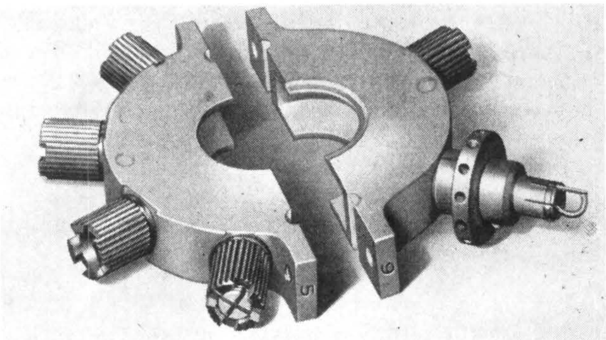
An up-converter, following the IF amplifier, shifted the amplified signal to the assigned microwave transmitted frequency. This was similar in principle to the down-converter, but the object at this point was to get maximum power out, rather than the lowest noise increment. This, too, used rectifying crystals in a balanced waveguide hybrid and, like the down-converter, unfortunately contributed about 10 dB of loss to the signal path. This loss did not impair the



(a)



(b)



(c)

Fig. 7-30. Experimental velocity-modulation microwave amplifier. (a) Disk-sealed tube. (b) Tube mounted in its focusing magnet. (c) External adjustable tuning cavity that clamped to the magnet assembly.

signal-to-noise ratio, since the signal was at a high level relative to the converter noise at this point, but it did add to the gain requirement for the following microwave amplifier. The maximum output obtainable from the up-converter was only a few milliwatts. An RF amplifier with a gain of more than 20 dB was therefore necessary.

Although much progress had been made in extending triode performance into the microwave range, laboratory tests showed that available triodes still fell considerably short of the required performance. An improved version of a klystron-type velocity-modulation amplifier remained the only realistic prospect for an RF amplifier within the time limits available. Numerous forms of disk-sealed tubes with adjustable resonant cavities outside the evacuated envelope and with magnetically focused beams were built and tested [Fig. 7-30]. Adjustable coupling loops picked up the energy within the cavities, and short coaxial links between stages were used to construct a four-stage amplifier with 0.5 to 1.0 w output to the antenna.

As the date for construction of the trial route neared, both new types of triodes and traveling-wave tubes were showing promise of becoming better broadband amplifiers than the velocity-variation amplifier. The velocity-variation amplifier, however, was the bird in hand and was not overtaken by the competitors in time to be displaced in the trial route. Considering the difficulties, the designers of the tube did remarkably well to produce a device that did the job at all. From the standpoint of the repeater designers, however, it must have been regarded with something less than affection.

Each tuned stage had an inherent bandwidth of only 5 MHz, so that it was necessary to stagger tune the stages to achieve the overall 10-MHz band needed. Broader tuning of the tubes within the microwave band was by means of screw studs into the cavity, but only a 250-MHz range was achievable in a single-cavity design. Two basic designs were therefore necessary to cover the entire band. A bulky external magnet was required for beam focusing and a 1500-v power supply was required for the collector, where 45 w was dissipated in each tube. The cavities were temperature sensitive. They required about 20 minutes to stabilize after being turned on and had to be enclosed in temperature-controlled chambers to maintain the tuning. This was accomplished on one side of a partition, while forced-air cooling was required by the 180 w dissipated in the collectors of four stages projecting through the other side. All in all, considerable hardware was required and heat generated for the 0.5 w or so of useful power realized.

4.1.6 Testing and the Complete Repeater

A variety of testing techniques were developed, both point by point and by oscilloscope displays of swept frequency response. By such means, for example, the difficult RF amplifier could be adjusted to flat gain, with less than 0.1-dB deviation over a 10-MHz band. Methods were developed to measure delay distortion over a wider (20-MHz) band to an accuracy of about ±1 ns, corresponding to a relative phase shift of ±0.35 degree, and equalizers were built at both RF and IF to equalize the delay of each component to about the limits

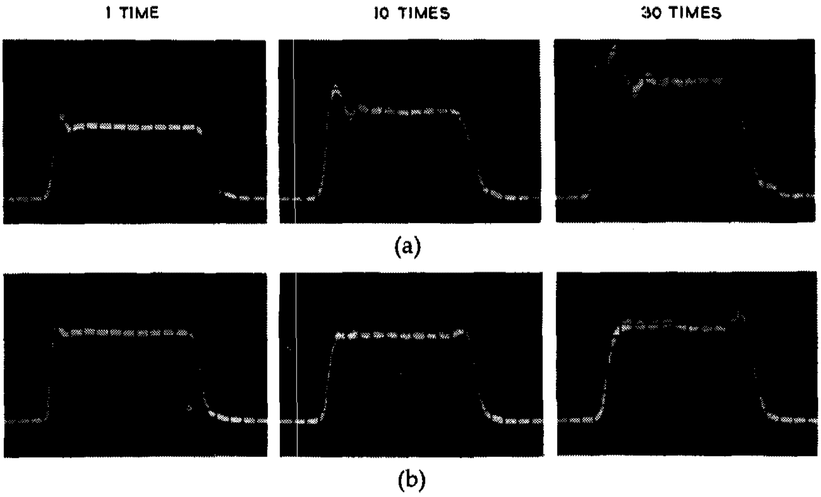


Fig. 7-31. Results of circulated pulse tests on IF amplifier. (a) Amplifier alone. (b) Amplifier with delay equalizer.

of the measuring ability. This was vital to broadband transmission of television by FM and to the as yet unknown capacity for multichannel telephony.

A particularly interesting technique was devised for testing the entire repeater with circulating pulses. A 2000-foot waveguide, with appropriate attenuation, was connected as a loop between the output and input. (For this test, the repeater was tuned to the same input and output frequency, permissible since the signal was not radiated.) It was thus possible to observe pulsed signals while they passed through the repeater many times, simulating what would happen in a multihop system [Fig. 7-31].

Some thought was also given in the research stage to fitting a repeater into an overall station. The nature of the terrain was such that the repeater stations on the first route had hilltop locations and the radio equipment could be located immediately below the antennas. For comparable hop lengths in flat country, it was recognized that towers of considerable height would be needed. The large amount of equipment in the repeater made it desirable for maintenance reasons to keep it on or near the ground. Long waveguide runs would then be required to the antennas at the top. There was grave concern that small imperfections in the waveguide or antennas could cause reflections that might prove to be the most serious source of distortion in the system.

4.1.7 System Planning

Finally, a great deal of thought was given to the total system layout and the frequency plans for future evolution. It was observed that as each new frequency range was opened, it was first used for services able to operate with relatively crude apparatus and methods (e.g., spark-gap transmitters). Later, as technology advanced and more intensive and efficient exploitation of the spectrum resource became possible, it could be impossible to bring any order out of the chaos because of established positions. Much thought was given to frequency bandwidths and assignments within the total band that were practicable within the limits of the technology in hand, but that would still permit a fuller exploitation and higher efficiency as the technology advanced. To quote the conclusion of an unpublished memo on the subject, written in early 1946:

> The possibilities of handicap to efficient development of radio relaying as an economical public utility are obvious. What should be done about it? The writers can at present suggest only some initial steps:
>
> 1. The Bell System should make up its mind how much risk it will take in the way of forward commitments in order to avoid the probability of later finding the situation compromised in the way outlined above.
> 2. In order to make up its mind it needs a comprehensive plan for Bell System use of radio relays, and the frequencies therefor, based upon reasonably optimistic assumptions as to prospective operating results.
> 3. The plan should contemplate an orderly evolution into which the necessarily sporadic initial experimental trials fit but which assumes rapid enough expansion to be impressive.
> 4. Means must then be sought to obtain some form of commitment or franchise to hold open the way for this evolution and growth. To date no way to do this has been available but to apply for construction permits and proceed actively with establishment of stations—in other words, to actually occupy and use.
> 5. After studying and understanding the matter as completely as we are able we will want to inform the FCC and its staff as soon as possible on the subject and lay a basis for their dealing with it in a sympathetic and enlightened manner.

It should be emphasized that the study leading to this conclusion by no means contemplated an exclusive franchise in these bands by the Bell System. It laid out an approach that all potential users could follow to their collective advantage. The intensive and highly efficient use that has been made of the common-carrier radio bands testifies to the wisdom of the approach advocated.

At the conclusion of their account of the development of each component and of the complete repeater and as they contemplated the antennas, filters, converters, oscillators, and amplifiers, the researchers were impelled to make the somewhat rueful comment: "Nevertheless, the equipment is very complicated. A straight-through radio frequency amplifier repeater is still to be desired and, no doubt, further research will eventually produce such an amplifier."[49] It was not to be. They had established a pattern that was admirable and adaptable to enormous improvement and extension and became successful beyond their most optimistic dreams.

4.2 The Experimental Line (1945–1947)

4.2.1 The Experimental System Plan

While the research was establishing the understanding and basic methods to be used, development was under way at the same time to firm up a specific plan and to build the equipment for the trial system, designated TDX. For the New York–Boston route, 2 two-way microwave channels were to be realized. This required the assignment of four radio-frequency bands, each about 10 MHz wide. After some vicissitudes, the assignments settled at 3930, 3970, 4130, and 4170 MHz. The bands were used in pairs, with one pair required for each two-way channel. In passing through a repeater, the frequency was always shifted by 40 MHz, so that transmitters and receivers at the same repeater station were never on the same frequency. At the next station, the inverse shift was made and the frequencies reused, so that the two RF assignments were sufficient for each two-way channel. The frequency plan was originally chosen to allow for later expansion to as many as 5 two-way radio channels in what was then expected to be a 400-MHz band, but there were to be some shifts and changes before the 4-GHz band and the channel assignments finally settled down. A repeater amplifier, with the intermediate frequency centered at 65 MHz, was designed with local oscillators for up- and down-converters at frequencies appropriate to the RF microwave assignments [Fig. 7-32].

Emphasis was on high-quality television transmission. In the transmitting terminal, a baseband, amplitude-modulated video signal in the range from 30

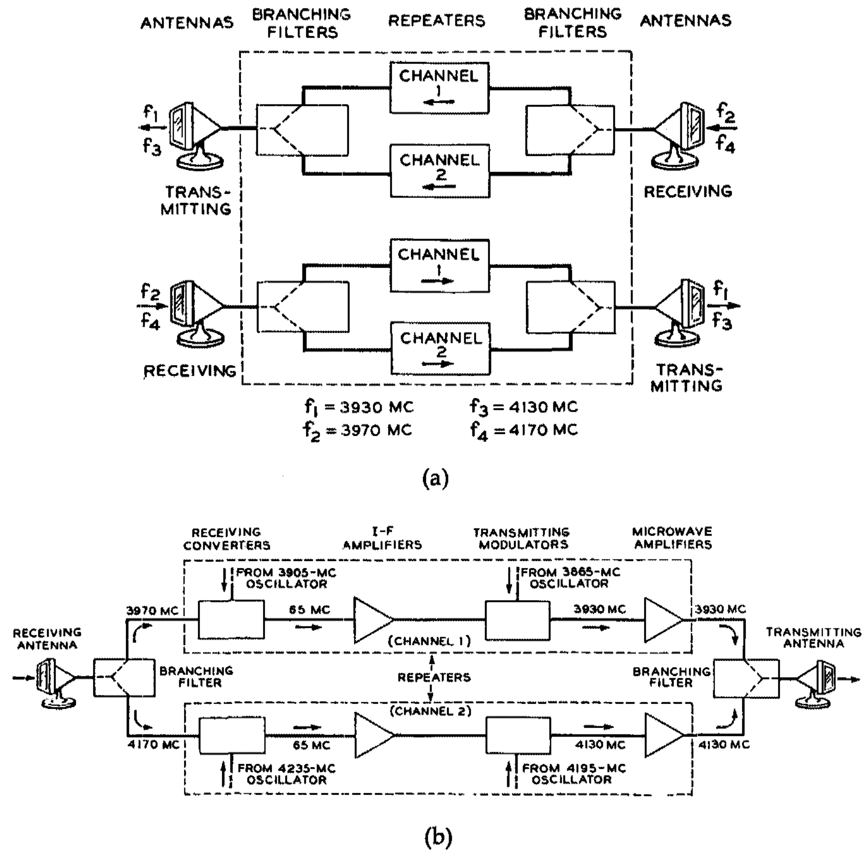


Fig. 7-32. Frequency plan and repeater schematic for experimental line. (a) A pair of back-to-back antennas was used for each direction of transmission. Frequencies were shifted at each repeater station to prevent having high-level transmitted and low-level received signals at the same frequency. (b) Repeater schematic for one direction.

Hz to 4.5 MHz was used to frequency-modulate a tone in the 65-MHz IF band. At the receiving terminal, the 65-MHz FM signal was passed through a discriminator, and the AM baseband video signal was recovered. The remainder of the terminal equipment, that is, from the FM signal at IF to microwave output in the transmitter and from microwave input to amplified IF in the receiver, was identical to the corresponding repeater circuits. Tests of FM television transmission by these methods were under way at the Deal laboratory shortly after the project was started in 1945.

Only a limited capability for multiplexed telephony was expected. No target was set for telephone capacity at first, but it was proposed that eight or possibly a few more channels might be transmitted by the pulse-position multiplex terminals used in the military AN/TRC-6 system. (For additional information, see *National Service in War and Peace (1925-1975)*, Chapter 5, Section VI.) Single-sideband AM transmission of multiplex voice was seen as very difficult because of transmitter linearity problems. Even before the war, FM for telephony as well as television was proposed and even strongly advocated, but this was viewed as exchanging the known difficulties of amplitude nonlinearity for the unknown problems of phase or delay nonlinearity. In 1945, pulse-code modulation (PCM) was regarded as probably the brightest prospect for multichannel telephony, and plans were made to transmit up to 48 calls by PCM. By mid-1945, however, FM was gaining in favor as understanding of multichannel intermodulation improved. Effort on PCM continued for a year or more, but the work was increasingly concentrated on FM. The hope was to transmit up to several hundred channels by frequency modulating the microwave carrier with the multichannel AM signal from a frequency-division multiplex terminal similar to that used for coaxial cable.

4.2.2 Building the Trial Route

An eight-hop route was selected with seven repeaters in addition to the two terminal stations [Fig. 7-33]. The overall route length was 220 miles, with hops averaging 27.5 miles and a longest hop of 35 miles. Repeaters were on hilltops to permit low towers or no tower at all, with the repeater equipment close to the antennas. The terminals were in tall downtown buildings with the antennas on the roof and the radio gear on the floor immediately below. Waveguide runs on the experimental route were not expected to exceed 30 feet. A plan was also formulated for a 21-mile link between the Bell Laboratories location at West Street on the Hudson River in New York and a tower at the suburban Murray Hill, New Jersey location. This link was to serve as a test bed for all system components and transmission methods. Path-propagation tests along proposed routes were planned using a portable, easily erected, 80-foot tower. A tower kit of this type was assembled, and the New York–Murray Hill path measured in August 1945. Later the same month, tests on the New York–Boston hops were started. The first leg, from the Long Lines building in lower Manhattan over the long 35-mile hop to Jackie Jones Mountain in New York State, was found to be within 0.5 dB of the calculated free-space loss. By November 1945, all links of the trial route had been path tested with satisfactory results [Fig. 7-34].

In the meantime, plans had been made for another early system installation from Chicago to Milwaukee, which was to be completed on essentially the

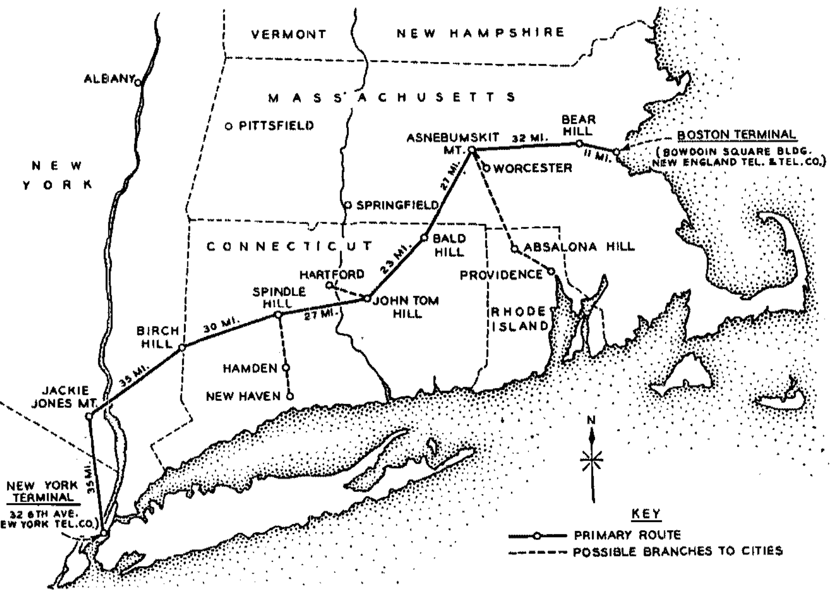
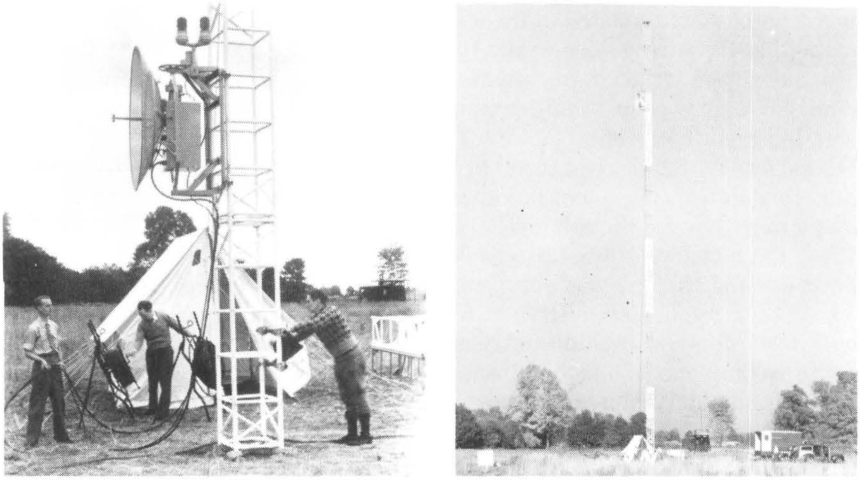


Fig. 7-33. Route of the experimental microwave relay system between New York and Boston.

same schedule as the New York–Boston route. On completion of path-loss tests on the first route, the path-test tower and equipment were sent to the Long Lines Department for tests on the second route.

There had been minor adjustments in the original frequency plan to conform to changed government allocations, but, in December 1945, the FCC dropped a bombshell. They announced that for nongovernment use, the 4-GHz band was to be split; frequencies from 3700 to 4000 MHz and from 4200 to 4400 MHz were to be available for commercial use; the band from 4000 to 4200 MHz was to be assigned to air navigation aids. AT&T took a calculated risk and continued work at the planned frequencies, which would have been ruled out under the new allocation, but petitioned immediately to restore a continuous band of at least 400 MHz for public communications. AT&T sought bands of either 3900 to 4400 MHz or, if only 400 MHz could be assigned, 4000 to 4400 MHz.

By February 1946, the FCC seemed disposed to assign 3700 to 4200 MHz for common-carrier use, with air navigation aids in the range 4200 to 4400 MHz. This wide and continuous band was eventually established. It was a generous assignment, and AT&T was presumably well satisfied, but all specific waveguide component-design work above 4200 MHz had been wasted, and the developers had a new problem. The waveguide components developed so far were all in one- by two-inch waveguide, which cut off below 3900 MHz. The Boston–New York system would have been confined to frequencies between 3900 and 4200 MHz or would have to shift to a larger waveguide. There was some hesitation, but since the first route was supposed to serve as a prototype for subsequent commercial designs, the decision was finally made to



(a) (b)

Fig. 7-34. Path propagation tests with a portable tower and antenna. (a) An antenna starts its ascent with cables connected. (b) The parabolic-dish antenna could be oriented in azimuth and elevation. Transmission was measured at various heights to determine a clear path.

convert to 1-1/4 by 2-1/2 inch waveguide, and the design conversions were made by late 1946.

Interest in microwaves by other companies and agencies continued to be intense. By 1946, requests for assignment in the microwave bands already indicated that this portion of the radio spectrum would be crowded. Requests from broadcasters, common carriers, private users, and government agencies, both military and nonmilitary, presented the FCC with an extremely difficult allocation problem. At an AT&T conference in May, a decision was made to follow the first systems, with whatever improvements could be made, as quickly as possible with a link from New York to Chicago. This would serve to test the system on a longer route as a prelude to nationwide use and demonstrate to the FCC a serious intent to utilize the spectrum in an efficient and effective way. In October 1946, a Bell Laboratories committee proposed two options for the follow-up system: a system with minimal changes from TDX, to be available in June 1949, or one incorporating a larger number of improvements by June 1950. As might have been expected, the management opted for the improvements, but pressed for the earlier date. (In the event, service to Chicago was not established until 1950, but with far more extensive changes than had been considered in 1946. A transcontinental line to San Francisco was completed only a month or so later.[50] [See Chapter 11, Section I.])

The decision to build a fairly long commercial system at an early date placed an additional burden on the development team, so considerable redesign was necessary, and effort had to be diverted from the work on the trial route. There were already problems to spare on that project. The early postwar years were difficult times in the construction industry, and buildings were repeatedly delayed by strikes and shortages. Experience at the Murray Hill site showed that water and ice accumulated in the waveguide antenna feeds. The decision was made that the waveguides would have to be airtight and filled with dry gas. The velocity-modulation microwave amplifier was an especially troublesome item. Not only was the tuning complex and the temperature sensitivity a problem, but the life was short as well. The objective of a 1000-hour life (only six weeks) proved extremely difficult to realize. Replacing the amplifier tubes with a more suitable component (an improved microwave triode) was one of the major objectives for the commercial design. Despite all difficulties, by early fall 1947, the installation was essentially complete [Fig. 7-35]. Three links were lined up, and the line was approaching the overall test stage. On November 13, 1947, a public demonstration was held of transmission over the line of both television and multiplexed telephony.[51] The radio-relay line was linked to the existing coaxial line to Philadelphia and Washington. Television cameras were on site at Washington, New York, and Boston, and the program was broadcast in the three cities. As part of the television tests, the 2 two-way radio channels were looped in tandem, and a test pattern transmitted over the 880-mile, double round-trip from New York to Boston. It was reported to be difficult to detect any impairment (presumably the observation was by a fresh observer—

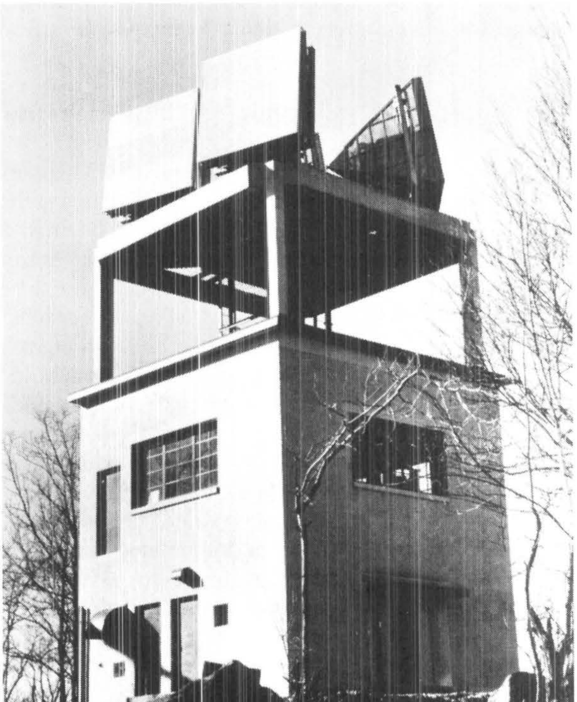


Fig. 7-35. An early repeater station of the TDX system. Hilltop locations permitted relatively low buildings, with short waveguide runs from the electronic equipment in the lower structure to the antennas.

some of the regulars must have been a little bleary-eyed at this stage). For the telephone demonstration, a standard Type K carrier group was transmitted on the FM carrier, permitting 12 test conversations. Somewhat later, noise and intermodulation measurements indicated that at least 240 single-sideband channels, multiplexed as in the L1 coaxial terminal, could be transmitted over the FM radio line without serious degradation.

The Long Lines Department took over the line in January 1948 and assumed all maintenance in April; Bell Laboratories was completely off the job by July. Starting on May 1, 1948, the New York–Boston line was kept in continuous operation, even though the microwave amplifier tubes were in short supply. The line was converted to the standard commercial design within a few years.

REFERENCES

1. L. Espenschied, "Radio Repeating System," U.S. Patent No. 1,397,093; filed September 27, 1919; issued November 15, 1921.
2. H. W. Nichols and J. C. Schelleng, "Propagation of Electric Waves Over the Earth," *Bell Syst. Tech. J.* **14** (April 1925), pp. 215-234.
3. See reference 2.
4. L. Espenschied and D. K. Martin, "Directive Radio Repeating System," U.S. Patent Nos. 1,668,637; 1,668,638; 1,668,674; and 1,669,123; filed December 28, 1921; issued May 8, 1928.
5. C. J. Young, "An Experimental Television System: Part IV—The Radio Relay Link for Television Signals," *Proc. IRE* **22** (November 1934), pp. 1286-1294.
6. C. R. Englund, "The Short Wave Limit of Vacuum Tube Oscillators," *Proc. IRE* **15** (November 1927), pp. 914-927.
7. W. E. Benham, "Theory of the Internal Action of Thermionic Systems at Moderately High Frequencies," *Phil. Mag.*, Series 7, **5** (March 1928), pp. 641-662.
8. F. B. Llewellyn, "Vacuum Tube Electronics at Ultra-high Frequencies," *Bell Syst. Tech. J.* **13** (January 1934), pp. 59-101.
9. F. B. Llewellyn, "Operation of Ultra-High-Frequency Vacuum Tubes," *Bell Syst. Tech. J.* **14** (October 1935), pp. 632-665.
10. B. J. Thompson and G. M. Rose, Jr., "Vacuum Tubes of Small Dimensions for Use at Extremely High Frequencies," *Proc. IRE* **21** (December 1933), pp. 1707-1721.
11. M. J. Kelly and A. L. Samuel, "Vacuum Tubes as High-Frequency Oscillators," *Bell Syst. Tech. J.* **14** (January 1935), pp. 97-134.
12. J. C. Schelleng, C. R. Burrows, and E. B. Ferrell, "Ultra-Short Wave Propagation," *Bell Syst. Tech. J.* **12** (April 1933), pp. 125-161.
13. C. R. Englund, A. B. Crawford, and W. W. Mumford, "Further Results of a Study of Ultra-Short-Wave Transmission Phenomena," *Bell Syst. Tech. J.* **14** (July 1935), pp. 369-387.
14. C. R. Burrows, L. E. Hunt, and A. Decino, "Ultra-Short-Wave Propagation: Mobile Urban Transmission Characteristics," *Bell Syst. Tech. J.* **14** (April 1935), pp. 253-272.
15. F. F. Merriam, "An Extension of Land Telephone Lines By Ultra-Short Wave Radio," *Bell Lab. Rec.* **13** (October 1934), pp. 34-38.
16. J. E. Smith, F. H. Kroger, and R. W. George, "Practical Application of an Ultra-High-Frequency Radio Relay Circuit," *Proc. IRE* **26** (November 1938), pp. 1311-1326.
17. F. H. Kroger, B. Trevor, and T. E. Smith, "A 500-Megacycle Radio-Relay Distribution System for Television," *RCA Rev.* **5** (July 1940), pp. 31-50.
18. G. Rodwin and L. M. Klenk, "High-Gain Amplifier for 150 Megacycles," *Proc. IRE* **28** (June 1940), pp. 257-261.
19. A. C. Peterson, Jr., "Multi-Channel Radio Telephone Spans the Chesapeake Entrance," *Bell Lab. Rec.* **22** (May 1944), pp. 387-390.

20. W. H. Wenstrom, "Historical Review of Ultra-Short-Wave Progress," *Proc. IRE* **20** (January 1932), pp. 95-112.
21. See reference 20.
22. H. Barkhausen and K. Kurz, "Die kürzesten mit Vakuumröhren herstellbaren Wellen," *Phys. Zeit.* **21** (January 1920), pp. 1-6.
23. E. W. B. Gill and J. H. Morrell, "Short Electric Waves obtained by Valves," *Phil. Mag.*, Series 6, **44** (July 1922), pp. 161-178.
24. See reference 20, p. 107.
25. S. Uda, "Radiotelegraphy and Radiotelephony on Half-Meter Waves," *Proc. IRE* **18** (June 1930), pp. 1047-1063.
26. See reference 20, p. 108.
27. "Micro-Ray Radio," *Elec. Commun. (ITT)* **10** (July 1931), pp. 20-21.
28. A. G. Clavier and L. C. Gallant, "The Anglo-French Micro-Ray Link Between Lympne and St. Inglevert," *Elec. Commun. (ITT)* **12** (January 1934), pp. 222-228.
29. F. B. Llewellyn, "The Barkhausen Oscillator," *Bell Lab. Rec.* **13** (August 1935), pp. 354-358.
30. R. H. Varian and S. F. Varian, "A High Frequency Oscillator and Amplifier," *J. Appl. Phys.* **10** (May 1939), pp. 321-327.
31. See reference 10, p. 1708.
32. Lord Rayleigh, "On the Passage of Electric Waves through Tubes, or the Vibrations of Dielectric Cylinders," *Phil. Mag.*, Series 5, **43** (February 1897), pp. 125-132.
33. O. Schriever, "Elektromagnetische Wellen an dielektrischen Drähten," *Ann. der Phys.*, Series 4, **63** (December 1, 1920), pp. 645-673.
34. G. C. Southworth, *Forty Years of Radio Research* (New York: Gordon and Breach, 1962).
35. J. R. Carson, S. P. Mead, and S. A. Schelkunoff, "Hyper-Frequency Wave Guides—Mathematical Theory," *Bell Syst. Tech. J.* **15** (April 1936), pp. 310-333.
36. See reference 34, p. 137.
37. G. C. Southworth, "Hyper-Frequency Wave Guides—General Considerations and Experimental Results," *Bell Syst. Tech. J.* **15** (April 1936), pp. 284-309.
38. W. L. Barrow, "Transmission of Electromagnetic Waves in Hollow Tubes of Metal," *Proc. IRE* **24** (October 1936), pp. 1298-1328.
39. G. C. Southworth, U.S. Patent No. 2,153,728; filed October 7, 1936; issued April 11, 1939.
40. See reference 30.
41. L. E. Thompson, "A Microwave Relay System," *Proc. IRE* **34** (December 1946), pp. 936-942.
42. H. T. Friis, "Microwave Repeater Research," *Bell Syst. Tech. J.* **27** (April 1948), pp. 183-246.
43. See reference 42, p. 187.
44. See reference 42, p. 191.
45. H. T. Friis and A. C. Beck, U.S. Patent No. 2,236,393; filed March 1, 1939; issued March 25, 1941.
46. J. H. Scaff and R. S. Ohl, "Development of Silicon Crystal Rectifiers for Microwave Radar Receivers," *Bell Syst. Tech. J.* **26** (January 1947), pp. 1-30.
47. C. F. Edwards, "Microwave Converters," *Proc. IRE* **35** (November 1947), pp. 1181-1191.
48. J. R. Pierce and W. G. Shepard, "Reflex Oscillators," *Bell Syst. Tech. J.* **26** (July 1947), pp. 460-681.
49. See reference 42, p. 243.
50. A. C. Dickieson, "The TD2 Story—From Research to Field Trial," *Bell Lab. Rec.* **45** (October 1967), pp. 283-289.
51. G. N. Thayer, A. A. Roetken, R. W. Friis, and A. L. Durkee, "A Broad-Band Microwave Relay System Between New York and Boston," *Proc. IRE* **37** (February 1949), pp. 183-188.

Chapter 8

Supporting Technology (1925–1950)

I. COMPONENTS

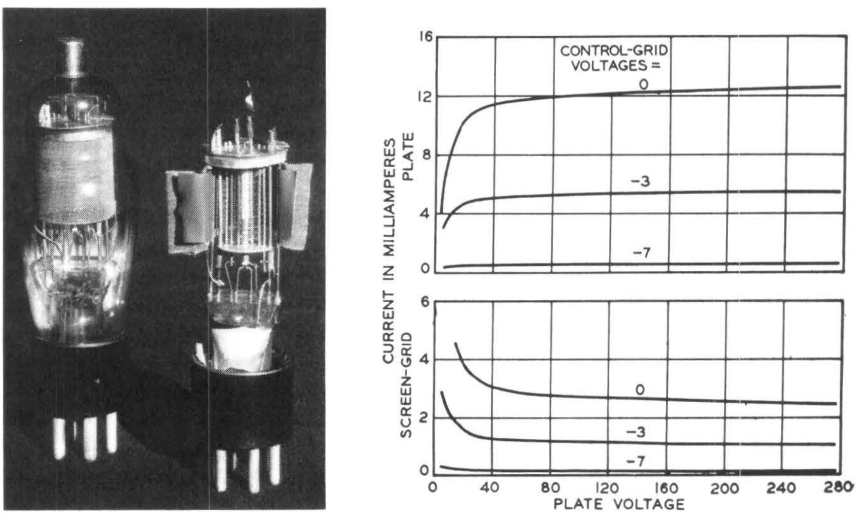
1.1 Vacuum Tubes

Early vacuum-tube repeaters all used triodes with emission from a directly heated filament. These were tubes of moderate gain, well suited to voice-frequency application and the fairly low frequencies of the early open-wire carrier systems. DC filament current and electrode voltages were obtained from office battery power. Tube development was directed toward the improvement of reliability, stability of characteristics, and long life. In the late 1920s and early 1930s, development was extended in two major new directions: the achievement of higher-frequency operation, especially for radio applications, and the realization of properties especially valuable in the emerging new carrier systems. Good linearity and low interelectrode capacitance were the qualities needed for higher-frequency systems on coaxial and paired cable.

A big advance for carrier applications came in the early 1930s with the development of tetrodes and pentodes with indirectly-heated equipotential cathodes. In the tetrode, the tube designers addressed the problem presented by the relatively large capacitance between the triode grid and plate by inserting a screen grid between them. This grid, at fixed potential equal to the DC plate voltage, provided the accelerating potential for the emitted electrons, but intercepted only a small fraction of the total current, the remainder passing to the plate. By this means, the grid-plate capacitance could be reduced by a factor of ten or more. Pentodes were a further elaboration in which another open-grid electrode, the suppressor, was placed between the screen and the plate. With the suppressor kept at cathode potential, close to ground voltage, secondary electrons emitted from the plate because of the electron bombardment were returned to the plate instead of reaching the screen. Without the suppression of the secondary electron current, the effectiveness of the screen grid as a grid-plate shield would have been reduced. These new multielement tubes had many other advantages. The indirectly-heated cathode could be better controlled in dimensions than the earlier incandescent filaments. This permitted more uniform and closer grid-cathode spacings with resulting higher performance. The uniform cathode potential also contributed to improved performance and could be made different from the battery ground, a property especially useful in many feedback circuits. Although these tubes had been developed in part because of the need for AC-operated heaters in home radio receivers, they were almost always operated from DC batteries in telephone applications.

Pentode characteristics were especially valuable to the circuit designer. Not only was the troublesome grid-plate capacitance greatly reduced, but the fact that the plate current was virtually independent of the plate voltage made the tube a nearly constant-current source. In vacuum-tube terminology, it had a large plate resistance. The gain of each stage, therefore, depended almost entirely on the passive elements of the interstage and output networks. It was no longer necessary in design to allow for the limiting and variable elements of the triode-tube characteristics [Fig. 8-1].

By 1937, a new family of pentodes had been developed for the 12-channel cable and open-wire carrier systems.[1] These tubes provided a transconductance of up to 2800 micromhos and a gain of as much as 40 dB per stage, compared to a maximum of only 20 to 25 dB available with triodes. (Transconductance, or mutual conductance, in micromhos [designated G_m], was the change of plate current in microamperes per volt change in grid voltage. It was a common measure of tube performance. A more elaborate figure of merit was the G_m divided by the sum of the plate and grid capacitance to ground. This determined the gain-band product available, the limiting factor in feedback-amplifier design.) Grid-plate capacitance was reduced by a factor of 14 and the power



(a) (b)

Fig. 8-1. The 310A pentode amplifier tube. (a) *Left,* the complete tube and, *right,* the tube dissected by folding back the plate and outer screen to expose the inner elements. (b) Static characteristics of the 310A. The plate current is a function of grid voltage but is almost entirely independent of the plate voltage.

output of about 2 w was almost 10 times that of the earlier triodes. It was these tubes, used in amplifiers highly stabilized and linearized with feedback, that made possible the cable-carrier system that so far exceeded in performance the first tentative designs.

Because of the closer repeater spacing and relatively low loss of the cable, coaxial repeaters did not require as much gain or power output as the paired-cable and open-wire systems. However, the extremely wide band, compared to the paired conductor systems, made other requirements on the vacuum tubes even more demanding for coaxial cable than for the wire systems. It was especially important to minimize capacitance and shorten the leads to make the tubes suitable for the coaxial-repeater feedback design, which extended to ten or more times the transmitted band. The result was a tube structure more radically redesigned than was necessary for the 12-channel systems. Capacitance and lead length were reduced by omitting the base and bringing the plate lead out the top of the envelope. In the original L1 coaxial repeaters, these tubes were soldered in place. With a grid-cathode spacing of only 0.0035 inch, these tubes had a transconductance of 5000 micromhos and represented a major advance in performance over anything else available. In a commercial version, designated the 6AK5 tube, the advanced tubes were made in large numbers by Western Electric and a number of other manufacturers for use in World War II radars and other military systems [Fig. 8-2].[2] In all the high-

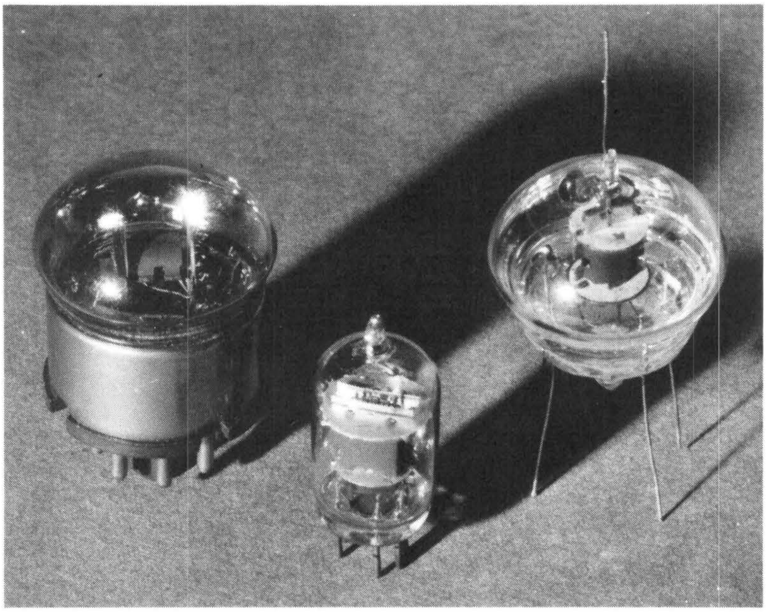
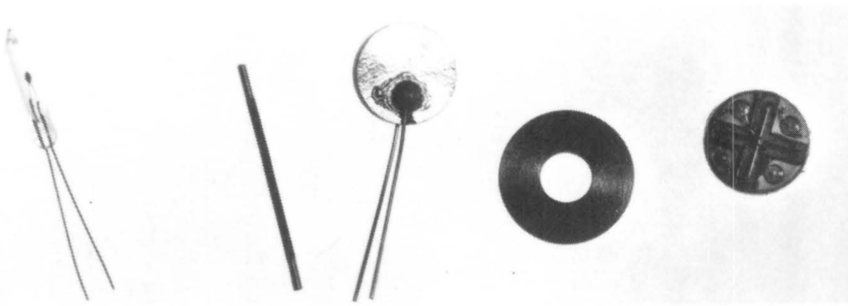


Fig. 8-2. Miniature pentode tubes designed between 1940 and 1944. *Right,* the 386A, was furnished in two commercial versions: *left,* with a base, as the 717A, and, *center,* as the unbased 6AK5. The 6AK5 was used in large numbers in military systems during World War II.

performance tubes, reliability, stability, linearity, and long life were no less important than formerly. In fact, these qualities became even more important as the tubes were used in larger numbers at the close repeater spacings of the new carrier systems and were usually located in unattended huts or manholes. (See another volume in this series, *Electronics Technology (1925–1975)*, Chapter 3.)

1.2 Passive Components

Besides vacuum tubes, a number of devices resulting from the pre–World War II research in semiconductors proved extremely useful in the carrier systems. Copper oxide rectifiers in bridge-type modems were a big improvement over balanced vacuum-tube modems because, unlike tubes, they required no maintenance adjustments after manufacture and installation. The thermistor was another device that found long and useful application.[3] A thermistor is a thermally sensitive resistor made up of any of a variety of solid semiconductor materials [Fig. 8-3]. At any given temperature, it behaves as a simple ohmic resistance for small signals, but its value can be changed over a wide range by changing its temperature. Thermistor elements were designed into the regulating networks of the L1 coaxial-cable system and of the later Type K2 cable carrier to compensate for the change in line loss with temperature. In these systems, a pilot tone was transmitted between the signal bands. At the output of regulating repeaters, the pilot was picked off via narrow band-pass filters and amplified and rectified to generate a current to control the thermistor temperature. Departures from the correct pilot level would bring about an increase or decrease in the heating current, and a corresponding change in thermistor resistance and amplifier gain, until the pilot was restored to the reference level. The pilot-controlled heating current was usually applied via a separate indirect heater, isolating the resistance-controlling currents from the network transmission path controlled by the thermistor resistance. The thermistor provided a sensitive, wide-range, electronically-controlled variable element and, with a few exceptions, made obsolete the cumbersome and high-maintenance electromechanical regulators.



(a) (b) (c) (d) (e)

Fig. 8-3. Thermistors made in the form of a (a) bead, (b) rod, (c) disk, (d) washer, and (e) flake.

In addition to the active and variable components, the new high-frequency circuits called for improvements in purely passive components as well. Resistors, capacitors, and inductors all became smaller, with much attention paid to unwanted stray capacitance and inductance. In critical circuits, the orderly mounting in arrays on equipment panels gave way to components mounted by their leads and arranged in ways to reduce unwanted capacitance and circuit lengths. Transmission transformers were especially critical items. They were needed since the line impedances were low, on the order of 75 to 135 Ω, while amplifiers were effective only at much higher impedance levels. At the input, the attenuated signal could be stepped up by a transformer to help override the vacuum-tube noise. At the output, the pentode tubes provided high gain only into a high impedance. The resulting relatively high-voltage signal had to be stepped down by an output transformer to a lower voltage and high current at the low line impedance. As these transformers were also in the feedback loop, they had to perform the in-band impedance-transformation function while presenting manageable circuit properties for the designer at frequencies many times the transmitted band.

II. MEASUREMENTS

In 1883, Lord Kelvin wrote: "When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of *science*."[4] The need for effective measurement dates back to the earliest days of telegraphy. Telegraph frequencies, however, were so low that practitioners thought of their circuits primarily in terms of the steady state and measurements were at DC, based on one form or another of the Wheatstone bridge. Telephony introduced voice frequencies, and telephone engineers were well aware that the higher frequencies placed new requirements on the circuits; but, for many years, measuring equipment remained crude and of limited accuracy. Around 1900, when the first long telephone cable circuits were established between Boston and West Newton, Massachusetts, the need for improved measurements of cable characteristics became evident. For this purpose, the first shielded AC bridges were developed by G. A. Campbell, then a research engineer at AT&T.[5] According to Campbell, one of these bridges was used for over five million measurements of capacitance and conductance between 1903 and 1922.

In 1921, a department was formed in Western Electric with the responsibility for furnishing measurement apparatus needed by those developing telephone circuits and for the control of production testing. In 1924, all work on measurement technology was consolidated and made the responsibility of Bell Laboratories at its founding in 1925. The measurements department continued as a distinct organization in Bell Laboratories over the entire period of this history from 1925 to 1975.

The measurements development group was concerned with two types of measurement: the impedance of individual components and transmission measurements—the gain or loss and phase shift experienced by a signal as it passed through a network, amplifier, or system. (In network terminology, the first type are two-terminal or driving-point impedance measurements, the second are four-terminal or two-port transfer measurements.) Impedance measurements were accomplished by various bridges, the AC equivalent of the DC Wheatstone bridge [Fig. 8-4], while transmission measurements were made with transmission measuring sets (TMSs). Impedance bridges were used to measure individual circuit components where accuracies in the range of 0.1 to 1.0 percent were usually adequate. Requirements for transmission measurements, typically on assembled networks and complete repeaters, also tended to be in the same accuracy range, but with increasing need to extend the accuracy available. The accuracy limits of bridges were usually expressed as percentages. Since TMSs measured ratios, their accuracy was usually given in decibels.

Measurements have always been used to provide a basis for system design and a means to confirm completed designs. The need for higher-quality circuits with more closely-controlled transmission and the greater frequency bands involved in the higher-capacity systems drove measurement needs to both higher accuracy and higher frequencies. The development of measuring techniques to meet the new needs was often regarded as an integral part of the system development. It was not unusual, especially in a new field, such as microwave radio, for the researchers or system developers to develop the required measuring capability as the system work evolved. Such techniques then often found wide application, and were standardized and further advanced for general use in the measurements group.

In a typical development sequence, concepts were first tested in a bench-top circuit (breadboard) and often on a metal chassis approximating the expected final form (brassboard). At each stage, measurements were made to confirm the performance. Obviously, it was always desirable to simulate and check out ideas on a laboratory basis as far as possible. If early results were promising, more complete equipment was built and measured, still on a laboratory basis. It is, however, the nature of transmission systems that they consist of long

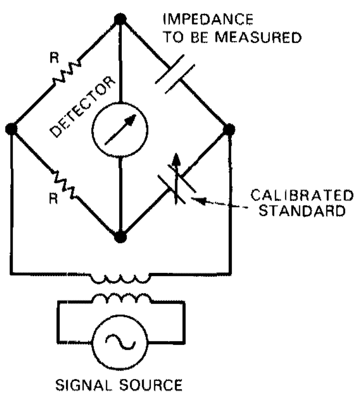


Fig. 8-4. Basic form of impedance bridge.

lines of equipment strung out over hundreds or thousands of miles. Only limited quantities could be assembled and tested in a laboratory. At some point, laboratory simulation was no longer adequate, and it became necessary to move the work to the real world of the field environment for a field trial. The purpose was to expose a larger array of the experimental equipment to what was hoped were representative field conditions. From the standpoint of measurements, this stage was an extension of laboratory measuring capabilities to the often difficult conditions of the field. From the measured results on the field-trial system, it was then necessary to extrapolate to the expected performance of systems up to 4000 miles in length.

The extrapolation placed extraordinary demands on measurement capability. The natural, measurable unit of a system was the repeater section. The repeater section was the basic building block from which long systems were constructed. In factory production, the single repeater was the unit manufactured and tested. Field trials usually involved multirepeater sections and were occasionally of considerable length to provide early information on how transmission variations accumulated, but the basic unit to be measured was the individual repeater. End-to-end circuit requirements were on the order of ±1 dB of the desired transmission loss for a 4000-mile system. On voice-frequency cable, repeaters were at about 50-mile intervals, but the extrapolation to long systems was hardly necessary since the line carried only a single voice circuit and the gain or loss was readily adjusted at repeaters and terminals to a reference level. In broadband systems, individual channels were not readily available at repeaters; and, as the number of channels expanded enormously, individual channel adjustment, even at terminals, became burdensome. Furthermore, appreciable departures from the intended levels along the way incurred a penalty in the signal-to-noise ratio. As repeater spacing dropped to 16 miles and, on coaxial cable, to 8 miles and ultimately to 1 mile, the extrapolation became awesome. In the years after World War II, it was barely possible under laboratory conditions to measure to ±0.001 dB reliably (with corresponding accuracy in phase), but the pressure to do so was intense.

2.1 Transmission Measurements

In the most elementary type of transmission measurement, a source of known frequency and power was connected to the input of the device or circuit under test. At the output, a calibrated indicator, usually a DC meter or galvanometer, to read the rectified current from the signal gave an indication of the received power. This arrangement provided no information on phase shift, which was hardly a consideration for purely voice transmission.

The first step to increasing accuracy was to measure by substituting an adjustable attenuator standard of known accuracy for the device under test, such that the detector had the same reading for each connection [Fig. 8-5]. This technique, which came to be called radio-frequency substitution, removed linearity and calibration requirements on the detector. The main remaining requirements were stability of the oscillator and detector and accurate calibration of the adjustable attenuator over the operating-frequency range. Measuring systems of this type were available with an upper frequency limit of 10 kHz

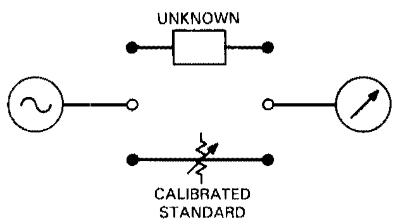


Fig. 8-5. Measurement by the radio-frequency substitution method, 1924. Accuracy of about ±1 percent was achieved at frequencies up to 10 kHz.

in 1925.[6] By 1927, to meet the needs of three-channel Type C carrier on open wire, the capability was extended to 50 kHz.[7] Specified accuracies were in the range of 0.10 to 0.25 dB.

In 1931, measurement capability was carefully reviewed because of its impact on the approach used to equalize the first cable-carrier system. Depending on the measurement accuracy obtainable, the system would use either extremely precise equalizers in each repeater or would use less precise repeater equalizers together with mop-up adjustable equalizers, inserted less frequently, to correct for the accumulated signal-level misalignment. Requirements for needed accuracy in the frequency range from 8 to 100 kHz were determined to be about ±0.001 dB for the more precise equalizer. Although careful procedures in the laboratory did achieve measurement accuracies of a few thousandths of a decibel, such accuracy was possible only on certain ideal networks and painstaking effort was required. The highly precise equalization approach was abandoned. Even if the precise equalizers could be designed, there was no really practical way to verify that they had been realized in production.

The coaxial-cable system under development in 1934 required precise measurements of loss and phase at frequencies up to about 5 MHz, more than a factor of ten higher than earlier systems. Phase measurements were of new importance because of the anticipated transmission of television signals over the line. In addition, measurement of phase shift, as well as of loss and gain, was required for the feedback-loop design for much higher frequencies, although at reduced accuracy requirements. Attenuation and phase standards of sufficient accuracy over a range of frequencies could not be made to frequencies much above 100 kHz, so a technique called intermediate-frequency (IF) substitution was employed [Fig. 8-6]. In IF substitution, the device under test was measured at the test frequency, and the output was applied to a frequency converter that converted the test frequency to an IF frequency under 100 kHz. There the loss and phase of the unknown was compared to that of loss- and phase-standards operating at the fixed IF frequency. As with other substitution methods, the final detector had only to be stable. The frequency converter was, however, required to be both stable and linear over the range of signal levels at the output of the circuit under test. Since the loss and phase standards had to be calibrated only at the IF frequency, potential errors could be compensated

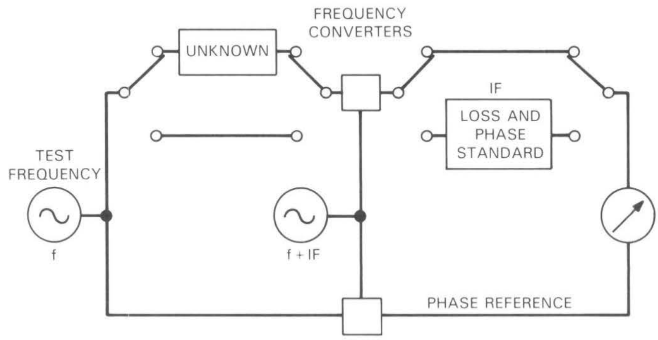


Fig. 8-6. Measurement by the intermediate-frequency substitution method, 1937. Loss and phase measurements to an accuracy of about ±1 percent were possible at frequencies up to 4.5 MHz.

for and minimized. Another benefit of IF substitution was that the detection bandwidth at IF could be very narrow relative to the top measurement frequency, with the resulting low noise and improvement in accuracy and loss

Fig. 8-7. Direct-reading phase and transmission measuring system, 1949.

range. With this type of circuit, accuracies in the range of ±0.1 dB and ±0.3 degree were achieved at frequencies up to 4.5 MHz by 1937. Measurements at this stage still required painstaking effort and measurement times were long. Loss and phase data at a single frequency required from 6 to 12 minutes, depending on operator skill. Measurements adequate to characterize a device over the complete frequency range would take several hours.

Near the end of World War II, as plans for expanded broadband carrier facilities were made and the increasing emphasis on the transmission of television required precise phase as well as loss measurements on larger numbers of transmission networks, the need for improved transmission measurements became urgent. By 1949, TMSs with improved accuracy and convenience were developed.[8] Accuracies of ±0.05 dB and 0.25 degree were achieved up to 3.6 MHz, and the speed of measurements was increased by at least a factor of ten. Some of the sets were used for production testing as well as laboratory evaluation [Fig. 8-7]. The sets used IF substitution and included such convenience features as detector self-tuning and direct reading of loss or gain and phase. The sets were stable enough so that the small deviations from linearity in calibration could be compensated by a movable zero index on the scales [Fig. 8-8]. This eliminated the tedious and time-consuming subtractions otherwise necessary.

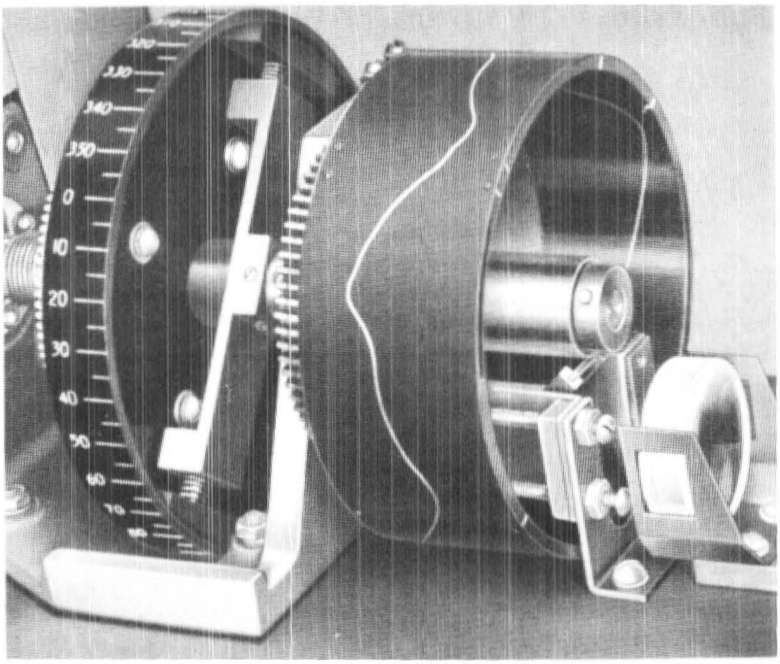


Fig. 8-8. Optical cam of phase shifter for zero index correction in direct-reading loss and phase measuring set, 1949.

2.2 Bridges

The development of impedance measurements at Bell Laboratories proceeded along lines parallel to, but separate from, the development of transmission measurements. Requirements for measurements associated with a new system development often resulted in a joint request for a new TMS and one or more new impedance-measuring sets, i.e., bridges. Early improvements in impedance measurement were in the areas of calibration methods and in bridge analysis and design. By 1928, a technique had been developed for standardizing capacitance to an accuracy of 0.003 percent in terms of the more exactly measured quantities of resistance and frequency.[9] This was an improvement by a factor of 30 over the best certified accuracies available from the National Bureau of Standards at that time. Bridge shielding was developed to a high degree of perfection to reduce the deleterious effect of parasitic capacitances and thus to obtain bridges that operated closer to the ideal.[10]

Unlike TMSs, bridges were often highly specialized. The impedances of the components to be measured ranged over many orders of magnitude, and their characteristics might be needed at any part of the frequency spectrum. The standard for measurement had to be calibrated at the same frequencies at which the component was measured and, for accuracy, had to be in the same general range of impedance. As a consequence, bridges generally operated only over a restricted frequency and impedance range, and the number of types proliferated. Over the period of this history, over 200 varieties of bridges were developed; and, even by the early 1930s, the number was so large that a paper was written classifying them into eight basic types, pointing out the advantages and trade-offs of each.[11] A big step in the convenience and accuracy of bridge measurements up to 150 kHz occurred with the development of the Type 12 admittance bridge in 1940.[12] This bridge used, for the first time, a true admittance standard (parallel capacitance and conductance standard). Bridge balances were easier to achieve, and the amount of computation required was reduced considerably, particularly in those cases where achieving the desired accuracy meant taking into account the conductance standard. Twenty-seven of these bridges were built, and most were still in operation in 1975 for standardization work.

The basic accuracy of the earlier bridges was ± 0.03 percent, improved in later ones to ± 0.01 percent. Some of these bridges were highly specialized to meet specific needs. For example, one bridge used a complex double-Y network of unusual form and was capable of measuring direct interelectrode capacitance smaller than 0.00001 pf at 465 kHz.[13] Another bridge, developed much later, operated at up to 10 MHz and used novel conductance standards and a phase-sensitive detector to achieve improved conductance resolution. It was capable of measuring the dissipation factor of polyethylene to an accuracy of one part in 10^6.[14]

With the large variety and large number of bridges used by Bell Laboratories and Western Electric, it was desirable to have capacitance and inductance standards to check the accuracy of individual bridges. The work of calibrating standards was done at Bell Laboratories because the National Bureau of Standards did not provide calibrations with sufficient accuracy over most of the range of impedance and frequency.

In time, the accumulation of bridges presented the local management with a housekeeping dilemma. They were kept in a large room; the space was needed for other purposes, and the instruments grew more archaic looking with each passing year. But they served a large clientele from all parts of Bell Laboratories and Western Electric; and, at unpredictable intervals, someone would show up who badly needed a measurement or calibration available in no other way. Eventually the technology changed, computer-operated transmission sets were adapted to make impedance measurements, the visits became less frequent, and the last defenders retired. It is hoped that the more magnificent specimens, with their highly finished hardwood cases and huge brass contacts, have found suitable last resting places in appropriate museums.

III. NETWORK DESIGN

By the early 1900s, the use of an inductor and capacitor to separate widely different frequencies and the resonant properties of such a pair were well known in telephony. The first vacuum-tube repeaters in the 1915 transcontinental line required the first electric networks designed for a more complex frequency characteristic. The successful operation of amplifiers in both directions over the single open-wire pair depended on the use of hybrid coils with networks to simulate the line impedance. In addition, filters were required to avoid instability in the circuits because of line-impedance variations outside of the voice-frequency range. In 1918, the first carrier system using four channels was installed. This required band-pass filters to separate the carrier band from the voice frequencies and the four channels from each other. It was the first use of channel filters. At about the same time, the first attenuation equalizer consisting of a resistance and reactance network across the line was used.

3.1 Filters

3.1.1 Campbell's Invention

About the turn of the century, G. A. Campbell was experimenting with loading coils on both actual cables and laboratory simulations (artificial lines). In his 1903 paper on loading he says, "I have made use of these results by employing artificial loaded lines for cutting out harmonics in generator currents. The harmonics may all be cut down as far as desired by the use of a sufficient number of sections, while the attenuation of the fundamental can be reduced at pleasure by decreasing the resistances. . . . Combining condensers and inductances, we may make a system which will not only cut out higher frequencies but also all frequencies below a certain limit."[15] Campbell's patent on filters, which was issued in 1917, covered low-pass, high-pass, and band-pass filters.[16] (At about the same time as Campbell's patent, K. W. Wagner in Germany was also contributing to the invention of the electric wave filter.)

The reader may well wonder why there was such a big gap in time between Campbell's apparent knowledge of filters in 1903 and his first filter patent of 1917. As reported in an unpublished letter, many years later, at a lunch set up to introduce some new Bell Laboratories employees to the sages of the business, a young engineer asked Campbell: "When you wrote your paper on loaded

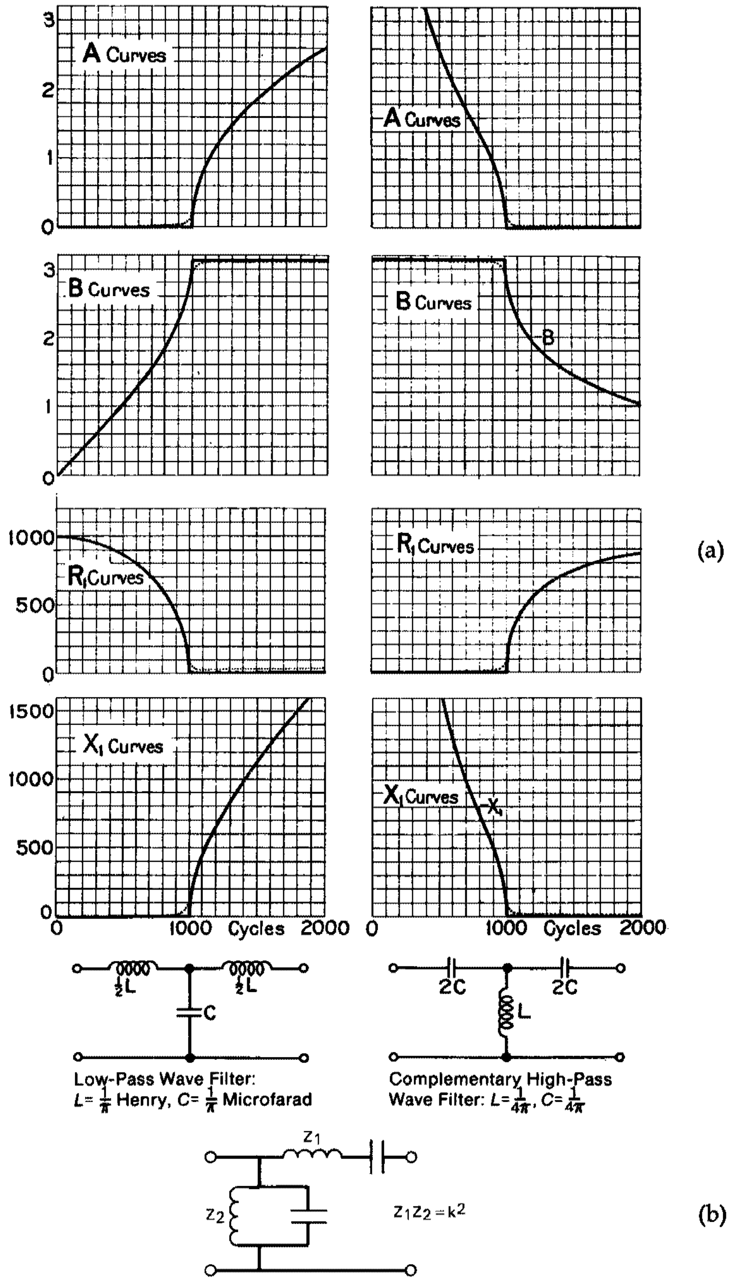


Fig. 8-9. Campbell electric wave filters. (a) Attenuation, phase, resistance, and reactance of low-pass and high-pass sections with iterative impedance terminations, as depicted by G. A. Campbell in 1922. (b) A constant *k* band-pass section.

lines in 1903, did you realize that the results contained the theory of wave filters?" Campbell said, "Yes, I knew that, but I forgot it and rediscovered it all at a later time."

Campbell showed that, with ideal lumped elements, the loss up to a preselected cutoff frequency could be made small, increasing above that frequency monotonically to provide a low-pass filter. Also, with the inductance and capacitance interchanged, the low-pass characteristic could be changed to high-pass. If a series-resonant circuit was substituted for the series arm and a parallel resonant (antiresonant) circuit for the shunt arm, a band-pass filter was realized [Fig. 8-9]. By making the product of the impedances of the two arms a constant ($Z_1Z_2 = k^2$), Campbell derived what he called a constant *k* section. Using sections with this constraint, he showed how to proportion the filter elements for any specified cutoff frequency and to size them to operate effectively in circuits of any given impedance.[17]

The first filters were formed by cascading the simple four-terminal networks to yield a ladder structure with the desired loss outside the passband. The early built-up filters, however, had two major shortcomings: the loss did not build up as rapidly as desired at frequencies beyond the cutoff and the design of the cascaded ladder was not a straightforward matter. The loss of the cascaded sections was equal to the sum of the individual sections only under certain ideal conditions, not easily achieved in practice. Furthermore, the impedance over the passband, while in the right general range, varied widely across the band, making the analysis of performance difficult and further complicating the design.

3.1.2 Design Advances and Systematization

In the early 1920s, O. J. Zobel showed that improved performance could be achieved by a modification of the basic filter-section design and a better design approach realized by what was called the image-parameter design technique. In the so-called *m*-derived section, Zobel maintained the ability to assign the cutoff frequency and impedance level as in the constant *k* section, but elaborated the filter elements to introduce a peak of attenuation (a pole) at a selected frequency outside the passband. (The factor *m* was the ratio of the cutoff frequency to the frequency of the attenuation peak.) By the proper location of the pole, he was able to steepen the cutoff and, at the same time, realize a more uniform impedance in the passband. Zobel also showed how to build up a more complex filter by the cascaded connection of individual sections on what was called the image-impedance basis. By definition, a network is image matched at a pair of terminals if the impedances presented in both directions are identical [Fig. 8-10]. If all the sections are matched, then the total loss of the network can be found by simply adding the losses of the individual sections, because there are no reflections. For there to be no reflections, however, the individual sections must be matched, not only where one section meets the next, but also at the terminations. Zobel's *m*-derived sections helped by providing a more uniform impedance in the transmitted band, but terminal matching remained a serious limitation because it still required complex image impedances to terminate the filter perfectly, whereas in practice, the filters were usually terminated in a pure resistance.

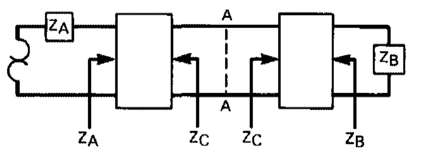


Fig. 8-10. Cascaded network sections matched on an image-impedance basis.

Design techniques based on image-parameter theory were developed even though the filter terminations could not provide exact image matching. Exact analysis methods were known, but the procedures were tedious and did not provide the insight required in design. To aid the designers, Zobel introduced charts that could be used to determine the loss of image-parameter filters under a variety of conditions. His charts greatly simplified the process of filter synthesis by letting the designer use a selection of building blocks. The charts could be used to determine the loss of each section, and the losses could then be added up to help determine the loss of the total filter. Other charts allowed for the effects of nonideal components and for the reflections at the imperfect input and output terminations of the structure.[18] About 1930, Zobel introduced another basic building-block section (the double-*m*-derived section), which could present a very nearly constant image impedance over the passband of a filter.[19] This helped to reduce the reflection loss that was associated with image-parameter filters and improved performance.

Perhaps as much as any branch of transmission research and development, network design lent itself to exact mathematical analysis. Much of the effort in the years from about 1925 to 1940 was devoted to building the theoretical base. R. M. Foster, in 1924, clarified some of the characteristics of purely reactive networks and showed that some things are possible and some impossible.[20] That is, Foster found necessary and sufficient conditions for a driving-point (two-terminal) impedance using only inductors and capacitors to be realizable. He showed when conditions were sufficient by presenting realization schemes that were based on partial-fraction expansions. (See also *Communications Sciences (1925–1980)*, Chapter 1, Section 3.2.) Networks realized in this manner have become known as Foster realizations. In 1926 in Germany, W. Cauer presented another realization scheme for impedances made up of pure reactances. This procedure was based on a continued division process instead of partial fractions, and the resulting networks are known as Cauer realizations. Both Foster and Cauer realizations are canonic, i.e., for a specified impedance function, they both use the minimal possible number of components.

Further efforts to reduce terminal reflection losses at the input and output ports of filter networks resulted in the Zobel *x*-termination structure for split-apart or directional filters. These filters, consisting of complementary structures (that is, low-pass and high-pass filters or band-pass and band-elimination filters), are connected in parallel at the input and are used to separate signal bands or individual carrier channels coming over the same transmission facility. The design technique was important for the application of the equivalent-four-

Fig. 8-11. H. W. Bode, who generalized filter design in the 1930s.

wire Type C carrier system and, later, for the application of Types C and J carriers to the same open-wire pair. In Zobel's approach, the combined susceptances of the paralleled networks were canceled over the passband. In 1930, H. W. Bode generalized this method of design [Fig. 8-11].[21]

By the early 1930s, network practitioners had shown that many network configurations, such as ladder, lattice, and bridged-T networks, had filtering properties. No general theory uniting all configurations, however, had been developed. Each structure was analyzed by a method adapted to that configuration alone. This situation was changed by a generalization developed by Bode in 1934.[22] As in Zobel's scheme, Bode's general filter was composed of simple structures in tandem, each section being obtained by transformations or derivations of elementary prototype sections. For all physically possible characteristics to be obtained, however, it was necessary to add two new sections to those described by Zobel: a complex m-derived section and an h-derived section, which was the converse of an m-derived section, since it changed the image impedance without affecting the transfer characteristics.

3.1.3 Approximation Theory

Throughout the work on filters, theoreticians and designers were concerned, not only with the analytical realization of a filter function, but also with the "best" realization. The measure of "goodness" was always within a framework of several constraints, not the least of which was the practical realizability of the filter when the design was achieved. Within practical limits, however, several criteria were used to measure quality, depending on the particular application in view. The art of approaching these measures as closely as possible was the concern of approximation theory.

The characteristics of transmission networks are expressed in polynomials. Maximally flat filters were derived from polynomials that, for a given degree, have the maximum number of derivatives equal to zero at a specified passband frequency. W. R. Bennett at Bell Laboratories had derived some expressions for a maximally flat ladder network in the late 1920s, but S. Butterworth is generally credited with applying these polynomials broadly to network synthesis; maximally flat filters are often referred to as Butterworth filters.[23] Maximally flat filters provide an excellent approximation at one particular frequency, or, stated another way, they concentrate their approximating power at one frequency. Equiripple filters, on the other hand, provide a uniform approximation in the passband and thus spread their approximating power over a wider frequency band than do maximally flat filters. A consequence of this is that equiripple filters have a steeper transition region than maximally flat filters and thus give more stop-band loss.

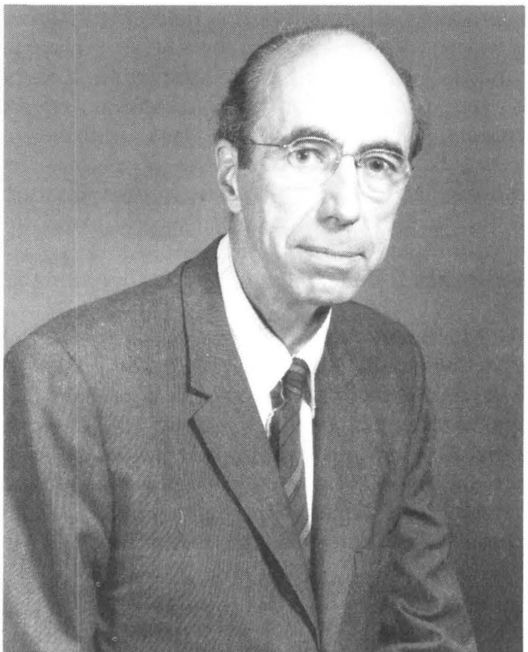


Fig. 8-12. S. Darlington, the originator of insertion-loss theory.

The mathematical functions that are used to solve the approximation problem for filters with equiripple passbands are known as Tchebycheff polynomials. They were probably first used for network design by Cauer, who also was probably the first to demonstrate theoretically how to obtain both an equiripple passband and equiripple stop-band (an equiripple stop-band has equal minimums of insertion loss). Cauer's initial work was for image-parameter filters. As pointed out by S. Darlington, Cauer's theory would lead to an insertion-loss characteristic with equal ripples in the passband and equiminimums in the attenuation band only if interaction and reflection effects were not present, which is not the case for image-parameter synthesis. Developments in approximation theory helped formalize the definition of ideal filter shapes such as the equiripple type. These filters, however, could not be synthesized by image-parameter sections; their exact synthesis had to wait the development of insertion-loss theory. Darlington was the initial advocate for insertion-loss filters, which could be designed to give equiripple passbands and stop bands [Fig. 8-12].[24]

3.1.4 Insertion-Loss Theory

Image-parameter theory was popular because the network designer could cascade image-parameter sections; and, by using enough of them, obtain large amounts of stop-band rejection. However, interaction and reflection effects degraded performance, especially in the passband. This could be improved by using special terminating networks, but it is theoretically impossible for such an approach to yield an ideal response such as an equiripple passband.

At Bell Laboratories, E. L. Norton* started the search for a better method.[25] Abandoning the image parameter concept, Norton proceeded to the problem of designing a complete filter directly from the specification of the required insertion-loss characteristic. He considered the problem of designing a two-port (or four-terminal) network terminated in an open or short circuit at one end, a special case of insertion-loss theory in which one of the specified terminations is zero or infinite. Norton's procedure, however, as with Zobel's x termination, was specifically concerned with filters operating in parallel at the input terminals.

The birthplace of insertion-loss theory can be considered to be the doctoral thesis of Darlington, later published in 1939 in the *Journal of Mathematics and Physics*.[26] This thesis and his subsequent contributions profoundly affected the synthesis of filters everywhere. Darlington attacked the general problem of designing a reactance network such that the loss caused by inserting it between a source resistance and a load resistance yielded a prescribed filtering shape. The reactance network was built up from elementary building blocks with poles of infinite attenuation at zero or infinite frequency and at finite frequencies

* Norton was something of a legendary figure in network theory work who turned out prodigious numbers of designs armed only with a slide rule and his intuition. Many anecdotes survive. On one occasion T. C. Fry called in his network theory group, which included at that time Bode, Darlington and R. L. Dietzold among others, and told them: "You fellows had better not sign up for any graduate courses or other outside work this coming year because you are going to take over the network design that Ed Norton has been doing single-handed."

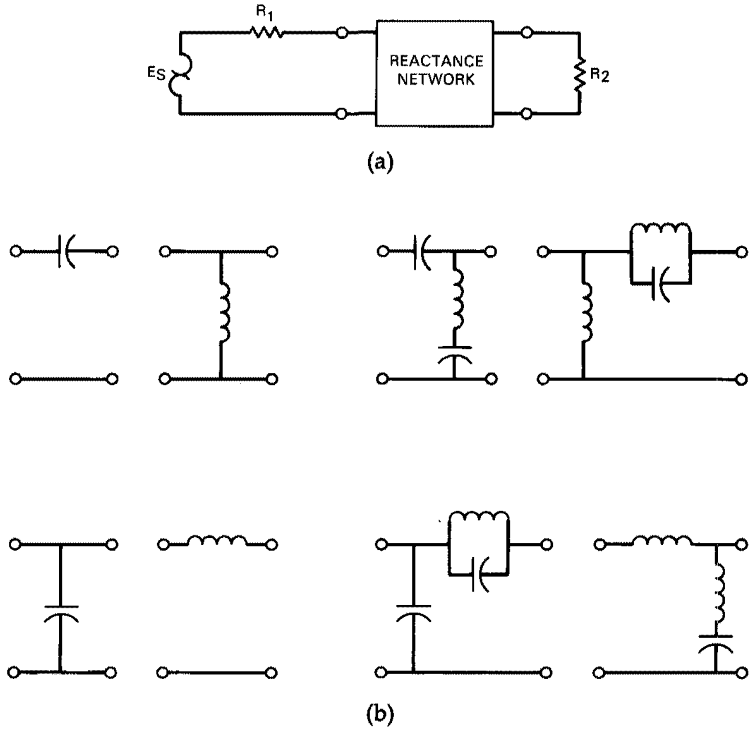


Fig. 8-13. Elements in S. Darlington's insertion-loss-theory approach to network synthesis, 1939. (a) A reactance network was synthesized with prescribed insertion loss between given terminations. (b) Common building blocks from which the reactance network was constructed.

below and above the passband [Fig. 8-13]. From these sections, one could create the same topologies as used in image-parameter filter design, but only the form would be the same. The element values obtained by insertion-loss theory would be different from those obtained by image-parameter theory, and the exact synthesis procedure offered by the insertion-loss theory led to better performance. The insertion-loss solution that Darlington presented was very general, but as he pointed out himself, the numerical calculations that it required limited its practical application. For the full benefits it promised, the Darlington technique had to await the next great advance in network technology—the application of computers to the process, not to occur until after 1950.

3.2 Attenuation and Phase Equalizers

As systems needs for frequency-selective networks grew, a corresponding need for attenuation and phase-correcting networks developed. Excessive amplitude and phase distortions in lines and repeaters and similar distortions introduced into the passband of filter networks by nonideal elements had to be corrected before they could be used in telephone circuits. For the transmission of television between New York and Washington, DC over an open-wire line in 1927, both amplitude and phase equalization of the line were required because the picture signal is sensitive to both amplitude and phase distortions. The equalizers used in the demonstration were made up of bridged-T structures to correct both types of distortion. When properly designed, such structures have a constant resistance input and output impedance over the frequency band of interest, and they can be cascaded to synthesize complicated loss or phase shapes by simply adding up the characteristics of each section [Fig. 8-14]. With these equalizers, circuits were made capable of carrying signals sensitive to both types of distortion, such as telephoto and television. Simple variable elements were also included in the networks on the New York–Washington line to correct for changing line characteristics in dry or wet weather. A paper in 1928 by Zobel laid a foundation for the understanding and design of networks to equalize the loss-versus-frequency characteristic of a filter or line.[27] Another 1928 paper by S. P. Mead established a mathematical basis for the design of delay-correcting networks.[28]

There was little need for more elaborate equalizers until the broadband paired-cable and coaxial systems were developed in the mid-1930s. In 1938, Bode published a classic paper on variable equalizers.[29] The Bode equalizer, in contrast to the Zobel type, depended on a variable resistor for control of a specified loss-versus-frequency shape. The loss shape in the Bode equalizer could be increased or decreased in magnitude, while retaining the same basic shape, or could be inverted to convert a loss shape into the same gain shape. Such variable equalizers were useful in compensating for daily or seasonal changes in the transmission loss of cables. They were especially useful in adjustable "mop-up" equalizers where, by using a number of "bump" shapes across the band that were adjustable in amplitude and sign, the finer-grained residues of distortion from the broader basic equalizers could be reduced.

It was while attempting to reformulate certain areas of network theory in connection with the design of equalizers that Bode showed that, because of the similarity between the general analytic conditions in such structures, equalizer theory is closely related to the theory of active networks and the stability

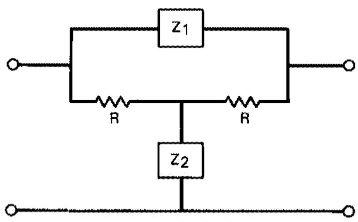


Fig. 8-14. Bridged-T constant-resistance equalizer network. If Z_2 is chosen as R^2/Z_1, the bridged T has constant input and output resistance equal to R over the frequency band of the design.

criteria for feedback amplifiers. Previous work by L. A. MacColl in the 1920s and early 1930s on the subject of signaling and its relation to the Fourier and Laplace transform techniques had given impetus to complex variable analysis as the natural basis for most linear circuit problems. Using MacColl's approach, Bode's systematic studies led to his formulation of the general relations between attenuation and phase in physical networks. Subsequent involvement in the equalizer problem, specifically directed toward inserting an equalizer in the feedback path of a feedback amplifier, led to a general attack on the feedback-design problem, as discussed in Bode's text on network analysis and feedback-amplifier design.[30] Speaking of this book, Bode said, "My book is essentially a text on equalizers and other dissipative networks, with incidental reference to amplifiers, rather than the other way around, and perhaps might have better been written with the equalizer point of view more obviously in the foreground."[31]

3.3 Network Realizations Circa 1925

Components available for network construction in 1925 included types that had been available, with only minor changes, for many years. Capacitors were made of mica and metal foil for applications requiring stability and low dissipation, and paper and foil for less demanding applications. (See another volume in this series, *Electronics Technology (1925–1975)*, Chapter 8.) Inductors were of two types: magnetic-core toroids for larger-inductance values and air-core solenoids for low-inductance, higher-frequency applications. Both types could be wound with stranded wire to reduce copper losses, and air-core inductors were sometimes wound with special winding geometries to reduce the distributed capacitance. In solenoids, small permalloy slugs were often provided to obtain fine adjustment. All resistors specified for precise network use were wire wound using materials of low temperature coefficient. Noninductive windings were widely used. Experts on each component type were available to supply custom designs on request and for consultation on the performance capabilities of their product. All components were available to close tolerances and were constructed for high reliability and high stability.

Measurements on networks and components could be made with a high degree of precision but with a very low degree of convenience and speed. Oscillators and detectors for laboratory use were large (several cubic feet in volume) and required the setting, by charts, of four to eight knobs and switches for each frequency. Impedance bridges and loss-measuring equipment were stable and precise, but were equally large and difficult to use. A transmission loss measurement, for example, might require different arrangements for balanced or unbalanced measurements and for a range of terminating resistances. A single measurement could require careful tuning of the oscillator and detector and several adjustments to balance a bridge or obtain a transmission-loss value.

In the specification of network components for manufacture, it was recognized that the actual element values in the product would not correspond exactly to the desired values. The effect of element variations could be calculated, but only very laboriously, using mechanical desk calculators. As a result, such computations were made only when necessary. A general practice was to specify

0.5-percent tolerances for sensitive network components, 1.0-percent tolerances for less sensitive elements, and 2.0-percent tolerances for insensitive elements in run-of-the-mill applications. For very demanding applications, slug tuning of solenoid inductors was specified, and the addition or removal of turns from magnetic-core toroids was required to meet specific inductance requirements. The resonant frequency adjustment of inductor-capacitor pairs was frequently specified, and the individual adjustment of inductor *Q*s was called for when required. (*Q* is the ratio of the reactance of an inductor or capacitor to the equivalent series resistance.) Production test sets capable of duplicating the accuracy of laboratory development measuring equipment were available for the measurement of both components and networks. Measurements of terminating impedances or reflection coefficient and insertion loss, all within given tolerances, were specified to guarantee that networks conformed to their design capabilities and that requirements imposed by system needs would be met.

3.3.1 Quartz-Crystal Filters

A new element of enormous importance, the piezoelectric quartz crystal, was first used in filters around 1930. In all filters, the sharpness of cutoff at the edge of the passband depends on the dissipative resistance associated with the reactive filter elements, that is, their *Q*. Capacitors were readily built with low resistance, but inductors invariably had at least the resistance of the often lengthy copper windings. The dissipation in the inductances in band-pass filters had an increasingly serious effect as the passband became a smaller percentage of the frequency at which the filter was used. It was thus easier to build a satisfactory 4-kHz channel-band filter for use at 20 kHz in the Type C carrier terminal than to build one at 100 kHz in the Type K carrier terminal. As much higher frequencies were considered for the planned coaxial system, the problem would have been acute even for the broader group-band filters. The solution to this and similar problems lay in the use of piezoelectric crystals as filter elements.

In a piezoelectric crystal, an applied electrical potential mechanically deforms the crystal, and, conversely, a mechanical deformation generates a voltage across the crystal. The energy stored in the elastic deformation of the crystal mass is analogous to the energy stored in the magnetic field of an inductor. By incorporating such crystals in an electrical circuit, a periodic exchange between the mechanical energy stored in the crystal and the electrical energy in the circuit can be effected in a way similar to the exchange between the magnetic and electric energy of circuits consisting of inductors and capacitors. It was known that the electrical impedance of a piezoelectric crystal near its resonant point could be closely approximated by a capacitance shunted by a series-tuned circuit [Fig. 8-15]. An important difference between the equivalent circuit of the crystal and a typical electrical circuit was that the dissipation of the crystal was extremely low. The *Q* was about two orders of magnitude higher than that available in wire-wound inductances.

Using Rochelle salt crystals, A. M. Nicolson at Western Electric made the initial applications of piezoelectric elements to loudspeakers, microphones, and phonograph pickups during the early 1920s.[32] He was also the first to control

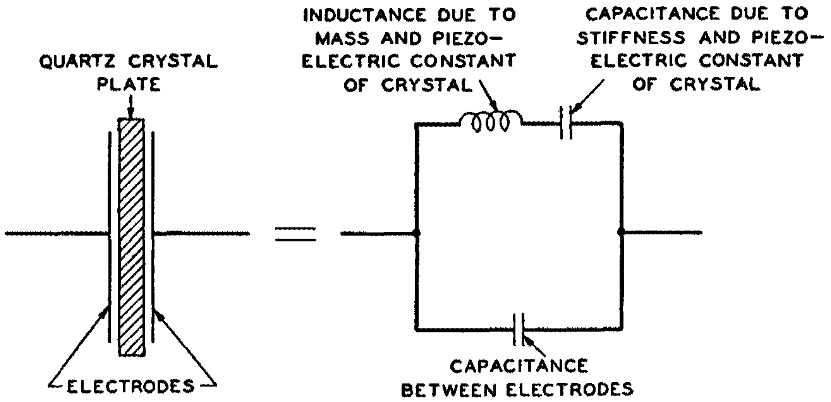


Fig. 8-15. Equivalent circuit of a piezoelectric quartz crystal.

an oscillator by means of a crystal.[33] Continuing efforts by W. A. Marrison to the problem of achieving improved frequency control resulted in a 100-kHz frequency standard using a doughnut-shaped quartz crystal having essentially zero temperature coefficient.[34,35] Of more interest to filter specialists, however, was his circuit in which a crystal was used to obtain transmission-frequency

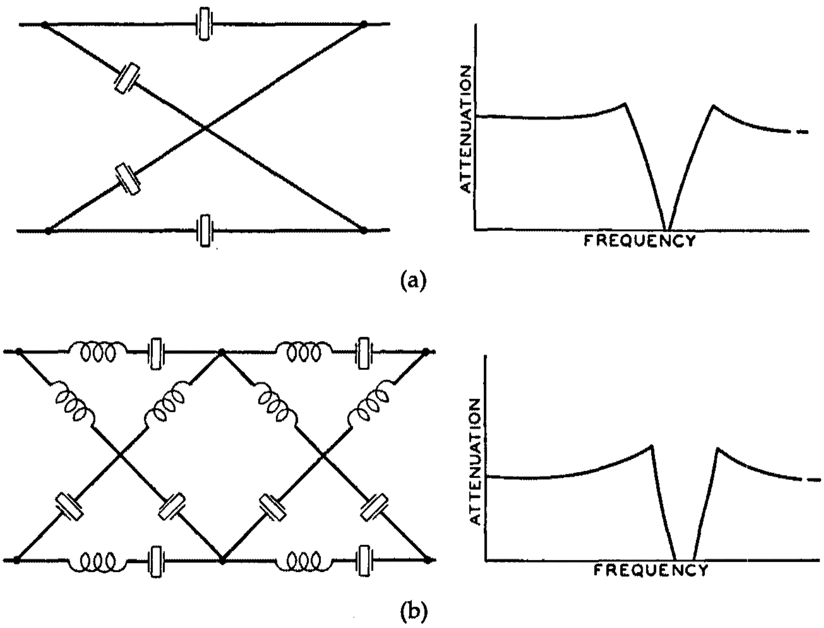


Fig. 8-16. Lattice-structure crystal filters. (a) With crystals alone, only a very narrow passband can be obtained. (b) Inductance in series with the crystal makes it possible to widen the passband.

selectivity. The basic patent covering the use of piezoelectric crystals in filters having application to broadband carrier systems was filed by L. Espenschied in 1927 and issued in 1933.[36] Espenschied's patent was for a ladder structure, but Campbell had shown earlier that lattice structures were the most general type of filter circuit and that they provided a more flexible design than ladders. They required a higher degree of precision in their element values, however, and were not in general use, except for some delay lines and phase equalizers. With crystal elements, they became the preferred configuration for filters.

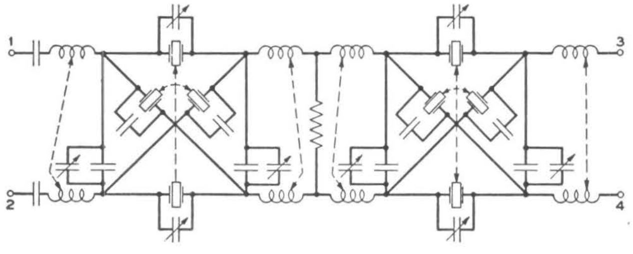
The reactance of the crystal has a series resonance and a parallel resonance (antiresonance) at a higher frequency. With four crystals in a lattice network, two of which have their series resonance at the same frequency as the parallel resonant point of the other two, a band-pass filter may be made. The theoretical bandwidth is the frequency separation between the series resonance of the lower-frequency crystal and the parallel resonance of the upper-frequency crystal. Due to the piezoelectric constant of quartz, this difference is very small, restricting the attainable bandwidth for this type of filter to less than one percent. With more than one crystal in each arm of the lattice, the bandwidth may be increased, but this leads to complexities in adjustment and increased cost. One way of increasing the frequency separation for crystals was to place an inductance in series, since this would decrease the frequency of the series resonance and have no effect on the parallel resonance. It would also add another series resonance to the equivalent circuit, further increasing the possible bandwidth [Fig. 8-16].

W. P. Mason suggested this method for designing lattice-network crystal filters of appreciable bandwidth in 1934, and, in addition, pointed out that if the added inductances were equal, they could be placed outside the lattice in series with the source and the load impedances. The resistance, due to coil dissipation, could then be considered part of the generator and load resistances.[37,38] This made it possible to consider the lattice itself as made up of the quartz crystals and the pure reactance of the inductances. This method increased the practical bandwidth for crystal filters from less than one percent to over ten percent. Since coil and condenser filters were useful down to about ten-percent bandwidth, the entire carrier range could be covered. Crystal filters were in commercial use by 1930, using Mason's lattice structures and were used in all the long-haul carrier systems designed after that date [Fig. 8-17].

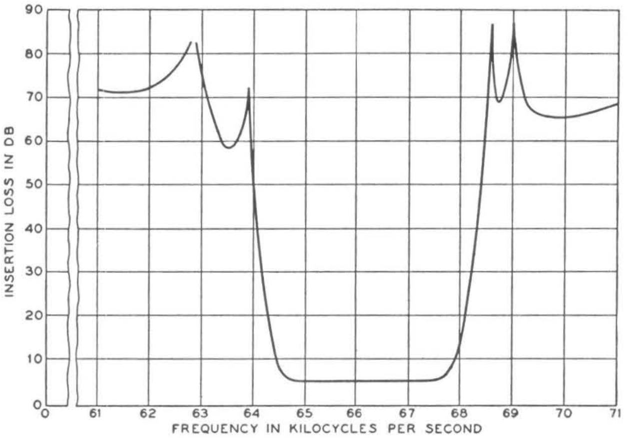
IV. REPEATER AND SYSTEM DESIGN

4.1 Evolution of the System-Design Concept

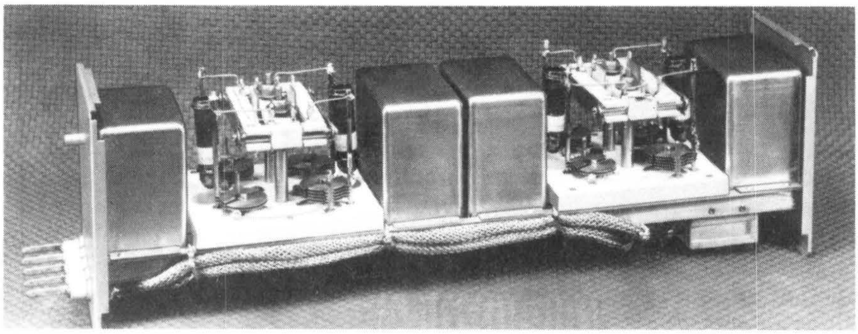
When vacuum tubes were first introduced in the early long-distance circuits, designers hardly thought in terms of systems or system design. Typically, there were one or two major limitations to be overcome, such as the high loss of the first transcontinental line or the even higher loss and crosstalk of the first carrier and long cable circuits. Most of the development effort was directed at solving the problems that were obviously the key to success or failure. As the transmission lines were improved and better devices and techniques became available, there were more options to consider, and many system-optimization decisions were made. An early example was the decision to dispense with loading



(a)



(b)



(c)

Fig. 8-17. Channel-band crystal filter, 1937. (a) Schematic showing inductors external to the lattice where their resistive dissipation can be considered part of the source and load impedance. (b) Typical attenuation-frequency characteristic. (c) Crystal-filter assembly as used in Types J, K, and L carrier terminals.

on open wire and to increase the number of repeaters well above the bare minimum required. With further advances, a still wider range of choices faced the developers. Decisions were needed on bandwidth, signal-to-noise ratio, and repeater spacing, in addition to the wire gauge to be used and the type of loading, if any.

In principle, all these could be traded off to arrive at an economic optimum for the service capability desired. In practice, there were many constraints on the range of choice. While the total bandwidth remained a system choice, the bandwidth objective for an individual voice channel was improved and standardized during the 1920s and early 1930s. The total spectrum of a voice signal contains components up to about 8 kHz, but experiments showed that little was gained in either intelligibility or naturalness for a transmitted band much above 3500 Hz. With the adoption of 4-kHz channel spacing on carrier systems, the voice band to be transmitted became stable at about that value. Voice-band signal-to-noise objectives were similarly improved and standardized. Signal-to-noise objectives were upgraded as late as the 1940s and 1950s, as the prospect for high-grade overseas circuits on submarine cable arose. Similar broadly applicable standards were adopted for radio-program channels and for television, as these services were provided. These basic service standards were accepted by all parties and were not sacrificed by the system designer in exchange for other system advantages.

The pattern of the existing plant itself imposed a number of constraints on what made sense in any new system. Thus, in the early 1930s, the widespread existing network of voice-frequency cables plus the depressed economy and the prospect of slow growth provided a strong incentive to design the cable carrier system for existing cables, rather than around a wholly new cable design. In addition, with attended voice-frequency repeater stations at 50-mile intervals, it was out of the question to consider carrier systems on cable requiring repeaters, inconsistent with this pattern. Terminals and repeaters had to make use of the existing sites as far as possible and intermediate repeaters had to be at spacings that were submultiples of the existing spans. When Type K carrier was designed, there was little hesitation to sacrifice an existing voice circuit to obtain 12 carrier-derived circuits. It was much less attractive, however, to remove four voice and one or more telegraph circuits on an open-wire pair to gain only eight new voice circuits. The existing voice and widely-installed three-channel carrier systems were strong incentives to place the new open-wire Type J carrier at frequencies above the existing circuits and to design it to coexist with them. In the coaxial-cable development, the territory was relatively virgin and there was a great deal more latitude. The development of the medium and repeaters went hand-in-hand, with few preexisting limitations. As the 4-kHz voice-channel spacing and basic group band at 60 to 108 kHz became established, however, all new systems were designed to these standards to provide for ready interconnection.

During the 1930s, with the success of the multichannel AM carrier systems, the concept of a system design and the optimization of that design became firmly established in Bell Laboratories. The first version of each system of a radically new type was, of necessity, quite experimental. Both the initial coaxial-cable development and the first microwave-radio system project in many ways resembled the struggle for sheer feasibility, characteristic of the earlier long systems. Both the medium and repeater techniques were new. The question was whether a system to transmit television satisfactorily over long distances could be built at all. Even if the radio system worked for television, there was great uncertainty as to its telephone capacity. It was not until long after the development was committed that any confidence was expressed in the system's ability to transmit several hundred telephone channels up to the current standards. In the first microwave-radio development, each component was pushed to the technical limit and there was very little discussion of optimization. Nevertheless, as soon as the first versions of coaxial and radio systems were achieved, the process of optimization and improvement started.

In concept, the optimized system was like the wonderful one-horse shay with each part in perfect balance with the others. But those were years of rapid technical progress. Designs were not expected to survive unaltered. There was always the hope, amounting in many cases to an expectation, that if a component could be made better than the system required, some way would be found to improve the remaining links. Designing in such a way as to leave the door open to future enhancements was an accepted objective.

It became characteristic that systems evolved in successive generations. Each generation achieved at least lower cost, required less space and power, and usually provided increased capacity and improved performance. Repeaters were built from the very best of available components. Designers pressed for vacuum tubes with higher gain, less distortion, and longer life. Coupling networks, interstages, and feedback networks made use of the latest findings in network theory and design techniques. The bandwidth was increased as far as the circuit performance and constraints of the medium permitted. Thus, in the course of the 1930s, coaxial amplifiers were successively redesigned and improved from the original 1-MHz design to 2 MHz, and eventually to 3 MHz as the technology permitted and the needs for high-quality television transmission became better understood.

4.2 Repeater Design

Terminals and terminal multiplexing were always important in system design, especially in the economics controlling the range of application, but repeaters usually established the system characteristics, setting the technical limit to what could be accomplished. The repeater section was the basic building block around which the high-frequency line was designed. A great deal of the system design effort, therefore, focused on the bandwidth achievable and the spacing permissible between repeaters. While bandwidth and spacing could be (and were) traded, the band tended to be set by the signal to be carried, such as television on coaxial cable. It was, of course, desirable to have as few repeaters as possible within the other limits set by system objectives. With the bandwidth more or less set, the maximum repeater spacing became a question of how high a signal could be transmitted and to how low a level it could be attenuated before amplification in the following repeater.

The low-level limit was set by the thermal noise of the conductors, the noise of the vacuum tubes, and the end-to-end quality-of-service standard. The physics of thermal noise was well understood by the early 1930s, and vacuum tubes were available that approached that absolute limit within a few decibels.

For a known number of repeaters, it was a simple matter to calculate how closely a signal could be exposed to the noise for satisfactory end-to-end performance. But the number of exposures depended on the repeater spacing, and the spacing also depended on the power level transmitted from a repeater. The higher the transmitted level, the wider the repeater spacings and the fewer repeaters required. Like many optimization problems, the best solution, taking into account all the variables, was not susceptible to an exact analytic solution; but the range of values was usually quite limited, and a few iterations brought the design quite close to the best that could be done.

With the vacuum tubes available, gain was no problem, even with large amounts sacrificed to feedback. Vacuum tubes with the other characteristics needed could be provided with an output power in the range from about 0.1 to 1.0 w. The design difficulties, power consumption, and expense ruled out multiwatt tubes. The feedback amplifiers, especially for the relatively narrow bands of the 12-channel systems, were highly linear. So much so that, when designers finally learned to build feedback amplifiers on the basis of Nyquist's and Bode's theoretical work, there was a wave of enthusiasm, and terms like *perfect linearity* and *infinite gain stability* began to crop up. But the linearity, however perfect, could exist only up to an upper limit in signal level. However, to determine the maximum power that could be transmitted involved surprisingly complex analysis.

4.2.1 Overload

Two general areas of amplifier operation were recognized. At low signal levels, there was little interaction between the signals of the highly linear feedback amplifiers; but, at some point, as the levels were raised, the amplifier would overload, and all the signals would be seriously degraded. The latter condition occurred when the peaks of the highly complex multichannel speech signal drove the control grids of the amplifier output stages into the positive conducting region. Under these conditions, the input impedance, the gain of the stage, and the loop feedback were sharply reduced, and the amplifier performance was abruptly degraded. In effect, the amplifier cracked. The question became: What portion of the time could this be permitted to happen on a high-quality circuit? The answer obviously was very little, but a quantitative answer was needed.

There were basically three aspects of the problem. It was first necessary to obtain data and determine the amplitude characteristics of individual speech signals and of the combined signal on the multichannel line. This was by far the most difficult part of the problem. The second question was then how much of the time the peak of this signal could be permitted to exceed a predetermined maximum level; and, finally, it was necessary to relate this to a simple load-capacity measurement that could be made in the laboratory with sine-wave signals.

The electrical analog of speech is a highly complex and highly variable signal. J. T. Dixon of the Toll Transmission Development Department and B. D. Holbrook in the Circuit Research Department attacked the first two problems on a statistical basis.[39] Data were first obtained on the instantaneous voltage distribution of the speech signal of a single talker at constant volume. A number of subjects, both male and female, were then measured at different volumes and with various telephone subsets. Talker-volume distributions were obtained by measurements on circuits in service. Finally, by further consideration of the probability that a circuit would be occupied and, if occupied, that a talker would be active, a composite picture of the line-signal amplitude distribution was constructed. (Talker activity was defined in these studies. It was found that on a busy [i.e., occupied] four-wire circuit, because of the talk-and-listen nature of conversation and the pauses in speech, activity was about 25 percent. These factors, much refined, were used later in the design of speech interpolation terminals.)

Since the nature of the basic speech signal, even on a statistical basis, was difficult to characterize analytically, much of the work was done experimentally. This was done, in part, by making phonograph records, carefully recording and playing back for measurement the speech signals for, first 1, then 4, 16, and finally 64 superimposed talkers. As might be expected, while the signal from a single talker is extremely peaky, with frequent amplitude excursions many times the average value, as more and more signals are combined the composite distribution approaches more closely a normal probability curve [Fig. 8-18].

The remaining steps in establishing the permissible load were relatively straightforward. The amount of time that the repeater could be permitted to be overloaded was set at about 0.1 percent. There was little room for judgment in setting this figure. The overload characteristic was something of a stone wall. Levels only slightly higher rapidly degraded performance to an intolerable degree, and levels a little lower rapidly improved it. It was also a simple matter, with the composite signal well characterized, to relate the capacity needed for the complex signal to the maximum sine-wave load capability. It was convenient to specify these requirements in terms of signal levels at the transmitting test board. By knowing the gain or loss from that point to the output stage of the

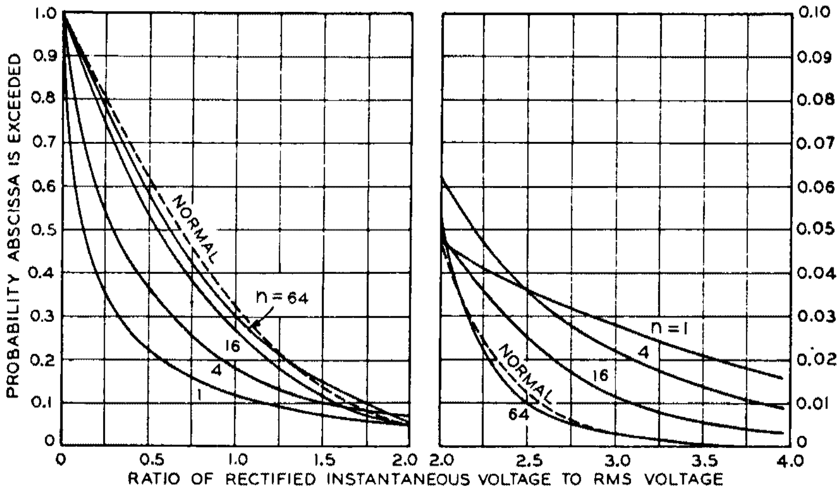


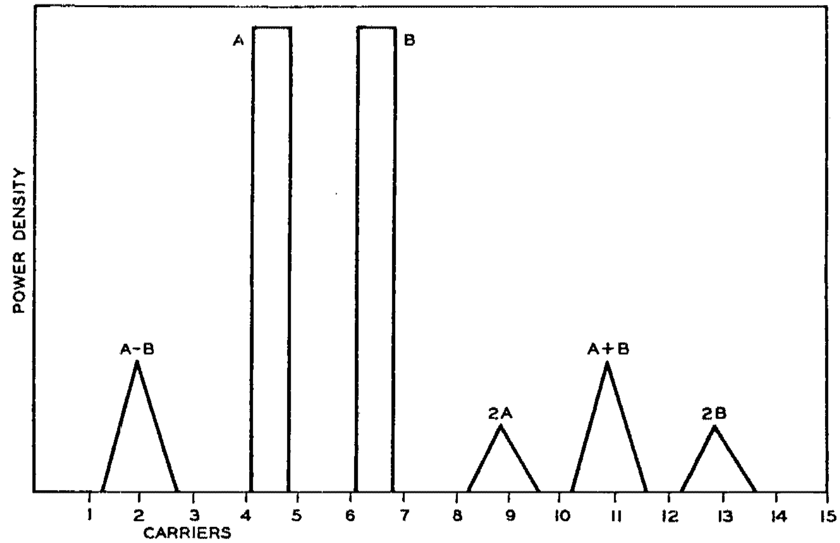
Fig. 8-18. Instantaneous voltage distributions measured for the combined signal from a number of talkers, 1939. For 64 talkers, the amplitude distribution is close to normal.

amplifier, the designer could determine an overload requirement in terms of a sine-wave voltage on the vacuum tubes.

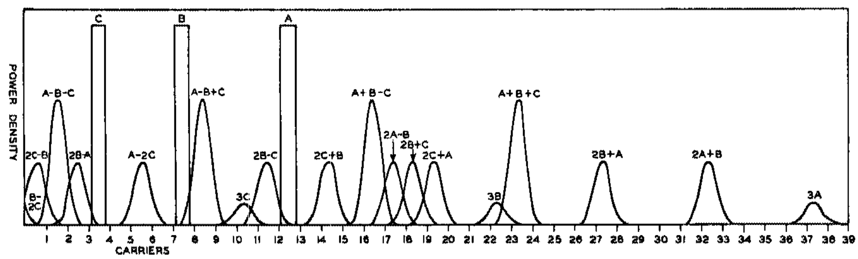
As the systems grew larger with the advent of coaxial cable, the statistics smoothed out more and more, and the line signal became closer in its statistics to random noise. In later coaxial systems design, it was usual to treat the composite signal for overload calculations as a random-noise signal of nearly constant average amplitude and with an effective peak-to-rms average ratio of about 13 dB.

4.2.2 Limits Due to Nonlinearity

But feedback amplifiers were not perfectly linear, and, in the wider-band coaxial amplifiers, with many more signals and lower feedback, nonlinear dis-



(a)



(b)

Fig. 8-19. Intermodulation products in a multichannel system. (a) Second-order products from signals *A* and *B* in two channels. (b) Third-order products from signals *A*, *B*, and *C* in three channels.

tortion rather than overload set the upper permissible signal levels. The analysis for determining the intermodulation-limited maximum signal level was published by Bennett in 1940.[40]

From the early days of vacuum-tube amplifiers, it was known that the nonlinear relation between the output current and input voltage could be approximated by means of a polynomial with linear, square, and cubic terms. If I_p is the output current and E_g the input voltage, then $I_p = a_1E_g + a_2E_g{}^2 + a_3E_g{}^3$. The first coefficient, a_1, is the linear gain factor. If the input is a sine wave, the second term leads to a second harmonic in the output, and the third term to a third harmonic. If the input consists of multiple sine waves, the harmonics of each are generated; but, in addition, various sum and difference intermodulation products are also generated, involving two or three of the signals. Bennett's analysis gave the magnitudes of the various products in terms of the nonlinearity coefficients, a_2 and a_3, and the volume distribution of them, taking into account the fluctuating nature of speech [Fig. 8-19]. He showed that the sum of the intermodulation products from a large number of speech signals could be treated as noise, similar in its interfering effect to thermal and tube noise. And he derived methods for counting and summing the intermodulation products for any number of channels to determine the total intermodulation noise falling in a disturbed channel. He further considered how the products generated in

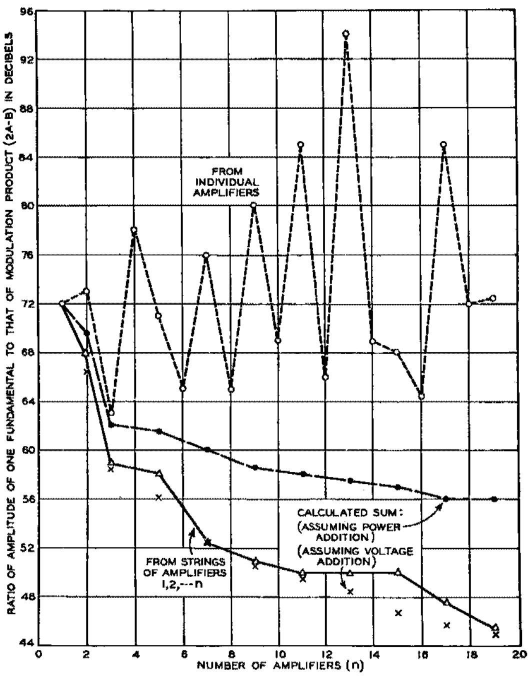


Fig. 8-20. Experimental data on the addition of third-order modulation in a multirepeater line. Fundamental test tones, $A = 920$ kHz, $B = 840$ kHz. Modulation product $2A - B = 1000$ kHz.

each repeater accumulated along a multirepeatered line. In this connection, he showed experimental confirmation of the bad news that R. V. L. Hartley had given to H. S. Black more than a decade earlier. Certain troublesome third-order products did indeed accumulate systematically in phase [Fig. 8-20]. Finally, he showed how to derive the modulation coefficients, from which the modulation noise could be computed, from simple harmonic measurements with sine waves.

The analysis of maximum permissible signal level, as limited by overload or excessive intermodulation, was the method by which these levels were established for Types J and K carrier and for several generations of coaxial systems. In practical application, many additional factors had to be taken into account. (Some of the most disturbing intermodulation products were caused by the many tones transmitted, some of which by cross modulating with speech could produce an intelligible signal in another channel. The television signal was also tonelike and troublesome.) The aging of components and imperfect line equalization and regulation, which caused signals to be at different levels at successive repeaters, also had to be considered. These factors were generally taken care of by allowing a suitable margin in the design, but the work done by Holbrook, Dixon, and Bennett was always the starting point.

REFERENCES

1. J. O. McNally, "New Tubes for Carrier Systems," *Bell Lab. Rec.* **16** (September 1937), pp. 17-20.
2. G. T. Ford, "Midget Tubes for High Frequencies," *Bell Lab. Rec.* **22** (November 1944), pp. 605-608.
3. J. A. Becker, "Varistors: Their Characteristics and Uses," *Bell Lab. Rec.* **18** (July 1940), pp. 322-327.
4. W. Thomson, Lord Kelvin, "Electrical Units of Measurement," in *Popular Lectures and Addresses* (London: Macmillan and Co., 1889), p. 73.
5. G. A. Campbell, "The Shielded Balance," *Elec. World Eng.* **43** (April 2, 1904), pp. 647-649.
6. S. J. Zammataro, "A Universal Laboratory Transmission-Measuring Set," *Bell Lab. Rec.* **2** (May 1926), pp. 98-102.
7. H. A. Affel and J. T. O'Leary, "High Frequency Measurement of Communication Lines," *Trans. Amer. Inst. Elec. Eng.* **46** (May 1927), pp. 504-513.
8. D. A. Alsberg and D. Leed, "Precise Direct Reading Phase and Transmission Measuring System for Video Frequencies," *Bell Syst. Tech. J.* **28** (April 1949), pp. 221-238.
9. J. G. Ferguson and B. W. Bartlett, "The Measurement of Capacitance in Terms of Resistance and Frequency," *Bell Syst. Tech. J.* **7** (July 1928), pp. 420-437.
10. J. G. Ferguson, "Shielding In High-Frequency Measurements," *Bell Syst. Tech. J.* **8** (July 1929), pp. 560-575.
11. J. G. Ferguson, "Classification of Bridge Methods of Measuring Impedance," *Bell Syst. Tech J.* **12** (October 1933), pp. 452-468.
12. W. D. Voelker, "An Improved Capacitance Bridge for Precision Measurements," *Bell Lab. Rec.* **20** (January 1942), pp. 133-137.
13. C. H. Young, "Measuring Inter-Electrode Capacitances," *Bell Lab. Rec.* **24** (December 1946), pp. 433-438.
14. L. D. White and H. T. Wilhelm, "A 0.1 to 10 MHz Dielectric Specimen Bridge with Dissipation Factor Accuracy of $\pm 10^{-6}$," *IEEE Trans. Instrum. Meas.* **IM-15** (December 1966), pp. 293-298.

15. G. A. Campbell, "On Loaded Lines in Telephonic Transmission," *Phil. Mag.*, Series 6, **5** (March 1903), pp. 313-330.
16. G. A. Campbell, U.S. Patent No. 1,227,113; filed July 15, 1915; issued May 22, 1917.
17. G. A. Campbell, "Physical Theory of the Electric Wave Filter," *Bell Syst. Tech. J.* **1** (November 1922), pp. 1-32.
18. O. J. Zobel, "Transmission Characteristics of Electric Wave-Filters," *Bell Syst. Tech. J.* **3** (October 1924), pp. 567-620.
19. O. J. Zobel, "Extensions to the Theory and Design of Electric Wave-Filters," *Bell Syst. Tech. J.* **10** (April 1931), pp. 284-341.
20. R. M. Foster, "Reactance Theorem," *Bell Syst. Tech. J.* **9** (October 1930), pp. 794-835.
21. H. W. Bode, "A Method of Impedance Correction," *Bell Syst. Tech. J.* **9** (October 1930), pp. 794-835.
22. H. W. Bode, "A General Theory of Electric Wave Filters," *J. Math Phys.* **13** (November 1934), pp. 275-362.
23. S. Butterworth, "On the Theory of Filter Amplifiers," *Experimental Wireless Eng.* **7** (October 1930), pp. 536-541.
24. S. Darlington, "Synthesis of Reactance 4-Poles which Produce Prescribed Insertion Loss Characteristics," *J. Math. and Phys.* **18** (September 1939), pp. 257-353.
25. E. L. Norton, "Constant Resistance Networks with Applications to Filter Groups," *Bell Syst. Tech. J.* **16** (April 1937), pp. 178-193.
26. See reference 24.
27. O. J. Zobel, "Distortion Correction in Electrical Circuits with Constant Resistance Recurrent Networks," *Bell Syst. Tech. J.* **7** (July 1928), pp. 438-534.
28. S. P. Mead, "Phase Distortion and Phase Distortion Correction," *Bell Syst. Tech. J.* **7** (April 1928), pp. 192-224.
29. H. W. Bode, "Variable Equalizers," *Bell Syst. Tech. J.* **17** (April 1938), pp. 229-244.
30. H. W. Bode, *Network Analysis and Feedback Amplifier Design* (Princeton, NJ: D. Van Nostrand Co., 1945).
31. H. W. Bode, "Feedback—The History of an Idea," *Proceedings of the Symposium on Active Networks and Feedback Systems*, New York (April 19-21, 1960), p. 11.
32. A. M. Nicolson, "The Piezo-electric Effectin the Composite Rochelle Salt Crystal," *Trans. Amer. Inst. Elec. Eng.* **38** (November 1919), pp. 1467-1485.
33. A. M. Nicolson, U.S. Patent No. 2,212,845; filed April 10, 1938; issued August 27, 1940.
34. W. A. Marrison, "A High Precision Standard of Frequency," *Bell Syst. Tech. J.* **8** (July 1929), pp. 493-514.
35. W. A. Marrison, U.S. Patent No. 1,899,163; filed December 19, 1928; issued February 28, 1933.
36. L. Espenschied, U.S. Patent No. 1,795,204; filed January 3, 1927; issued March 3, 1931.
37. W. P. Mason, "Electric Wave Filters Employing Quartz Crystals as Elements," *Bell Syst. Tech. J.* **13** (July 1934), pp. 405-452.
38. W. P. Mason, "Electromechanical Transducers and Wave Filters," (Princeton, NJ: D. Van Nostrand Co., 1942).
39. B. D. Holbrook and J. T. Dixon, "Load Rating Theory for Multi-Channel Amplifiers," *Bell Syst. Tech. J.* **18** (October 1939), pp. 624-644.
40. W. R. Bennett, "Cross-Modulation Requirements on Multichannel Amplifiers Below Overload," *Bell Syst. Tech. J.* **19** (October 1940), pp. 587-610.

PART II
THE EXPANSION OF THE ANALOG NETWORK (1950–1975)

The success in providing less expensive circuits in the long-haul plant by analog carrier methods led to attempts in the late 1940s to extend the method to exchange area trunks in the Type N systems. This effort was only partially successful. Terminals with compandors reduced the effect of crosstalk and made single cable operation possible, but Type N systems, while widely installed in the 1950s and 1960s, were economical only for short toll and long exchange trunks. The first Type N designs used vacuum tubes and double-sideband transmission; later versions were solid state and used single sideband. Similar approaches were used for Type O carrier on open wire.

The 25 years from 1950 to 1975 were marked by an enormous expansion of the analog long-haul transmission plant. The 4-GHz TD2 microwave FM system, first installed in 1950, rapidly spread to a nationwide network carrying both telephone and television signals. In 1960, TD2 route capacity was doubled by using horn-reflector antennas and cross-polarized transmission in alternate radio frequency slots. Successive improvements in power output and front-end noise figure increased capacity from 480 circuits per radio channel in 1952 to 1500 circuits in the 1973 design. By 1979, all vacuum tubes, including the microwave transmitter, had been replaced by solid-state devices. Developments at 6 GHz in the TH systems paralleled those at 4 GHz. The 6-GHz work culminated in a return to single-sideband AM transmission in the AR6A system developed in the 1970s. AR6A had a capacity of 6000 channels per 30-MHz frequency slot in the 6-GHz band. Short-haul radio systems were designed first at 11 Hz in the late 1950s and later at 6 GHz. In 1964, improved designs were linked in the 6/11 GHz crossband diversity TL/TM system. All solid-state short-haul systems were in service by 1973.

A second generation design, the 8-MHz L3 coaxial system, was in use by 1953. After a pause in the late 1950s when microwave radio accounted for most of the long-haul circuit growth, development of coaxial systems resumed in 1960. New solid-state components made it economically attractive to derive additional circuits on the extensive network of cables already installed. The L4 system, with 3600 circuits per coaxial pair, was first installed in 1967, and the L5 system, with 10,800 circuits per pair, in 1974. An expanded version of L5, with 13,200 circuits per coaxial pair, and a total capacity of 132,000 circuits in a 22-coaxial-tube sheath, was developed in the late 1970s.

The installation in 1950 of the first telephone cable with submerged repeaters to Cuba extended coaxial techniques to submarine systems. A 36-channel twin cable system, the SB design, with flexible repeaters spaced 37 nautical miles apart, was installed from Newfoundland to Britain in 1956. Speech interpolation terminals doubled the capacity of this cable in 1960. The SD 1-MHz system, using armorless cable, rigid repeaters, and a new cable ship and machinery for installation, provided 128 circuits in a single cable in 1963. The SB and SD systems used vacuum tubes; later submarine repeaters were all solid state, with the SF system (1968) providing 800 circuits in a single cable 1.5 inches in diameter and the SG system (1976) 4000 circuits in a 1.7-inch diameter cable. Considerable development was devoted to protecting submarine cables by plowing and burial in the 1960s and 1970s.

The Echo experiment of 1960 demonstrated communication via a passive satellite. In 1962, an active satellite, *Telstar,* first transmitted television and the equivalent of several hundred voice circuits to Europe. *Telstar* was made possible by the device breakthroughs of transistors, solar cells, and the traveling-wave tube, along with low-noise antennas and ultra-low-noise maser receivers, all of which were developed in Bell Laboratories. By the terms of the Communications Satellite Act of 1962, AT&T could not own and operate satellites for overseas communications but reentered the field, jointly with General Telephone and Electronics (GTE) and Communications Satellite Corporation (Comsat), in the domestic Comstar system in 1976. Comstar satellites were built by the Hughes Aircraft Co.

Two-way mobile radio telephone circuits to ships and for police and fire vehicles were in use by the 1930s. General-use systems to automobiles at 150 and 450 MHz were developed after World War II, but growth was handicapped by extremely limited frequency assignments. Developments starting in 1963 showed the virtues of cellular FM systems with frequent reuse of limited frequencies. Broader band assignments in 1974 opened the prospect for rapid deployment of high-capacity systems working in a band around 850 MHz.

Developments in multiplexing terminals kept pace with expanding system capacities. New solid-state devices reduced size, cost, and power consumption while improving circuit stability and reducing maintenance. A continuous and growing effort was devoted to the subscriber-loop plant, which accounted for over one-third of all telephone plant investment. Improvements were made in cable and wire materials and in methods for planning, handling growth, and changes in an area where circuits were necessarily handled one at a time. In the 1970s solid-state electronics made long loops on fine-gauge wires possible. Loop carrier for feeder plant was in service by 1978, and digital loop carrier systems started a rapid growth in the 1980s.

Chapter 9

Short-Haul Carrier on Pairs

I. TYPE N CARRIER

At the end of World War II, the Bell System had a coaxial-cable system in production and was well started on microwave radio as a second technique for long-haul telephony and television. There was also widespread application of the 12-channel Types J and K carrier systems on open-wire and cable pairs, but all these systems were economical only for long haul use. At the same time, there were far more shorter links, of moderate circuit cross sections, for which voice-frequency (VF) pairs were the only available facility. Because of the rapid growth of circuit needs, there was an urgent need for a cable carrier system that could be operated over either exchange-area or short toll-grade cable. In late 1945, studies were started to determine if a low-cost transmission vehicle could be developed for short-haul application. In addition to low cost, it was important to improve transmission characteristics compared to existing facilities, which were predominantly inductively loaded VF cable pairs. Compatibility with the embryonic nationwide toll-dialing plan was essential. On May 17, 1946, a conference of transmission engineering and development people addressed the requirements for this new system. A broad proposal resulted from this meeting, which served as the basis for early development.[1]

The proposed system was to carry 12 circuits on No. 16, 19, or 22 gauge unloaded cable pairs, with both directions in a single cable. The transmitted band of each circuit was to be at least 200 to 3200 Hz, with loss not to exceed 6 dB between terminals (typical VF tandem trunks ranged up to 10 to 15 dB). The system was to be amenable to use with either two- or four-wire switching, and means were to be provided to transmit dialing and supervisory signals. Application was to be economical for distances as short as 15 miles, and the system was to be useful up to 100 miles at somewhat reduced signal-to-noise objectives compared to long-haul systems.

All participants recognized the importance of terminal costs in establishing the minimum prove-in distance, and attention was focused in that direction. Component costs in the existing Type A2 channel bank design were so high that a radical departure from earlier practice was obviously necessary to achieve the cost objective. The crystal channel filters of the single-sideband, suppressed-carrier design were identified as one of the most significant cost items; double-sideband transmission with lower-cost inductor-capacitor filters was considered as an alternative. This was a mixed blessing, because the double-sideband signal required twice the transmitted bandwidth, resulting in a higher top frequency and an aggravated crosstalk problem. With the recognition that signaling costs were comparable to transmission costs, built-in signaling (the means for transmitting dial pulse and on-hook/off-hook indications as an integral part of the terminal) was proposed, as well as built-in four-wire/two-wire termination sets. In the initial design, a tone at 3700 Hz was used as the signaling vehicle. Finally, power costs were also significant, especially when generated locally at each repeater. It was proposed that power for the electron-tube amplifiers be transmitted from a primary source over the conductors to the adjacent repeaters. Considerable development was thus required on a number of fronts to achieve the cost objectives for the new N1 carrier system.

Aside from costs, the most challenging technical problem was crosstalk control. This was especially acute because of the objective that both directions of transmission be in a single cable. For crosstalk to be kept at manageable levels, different frequencies were necessary for the opposite directions of transmission. But the higher frequencies would always suffer greater loss and crosstalk. To average out impairments, an ingenious frequency-frogging plan was devised [Fig. 9-1]. Terminals were designed to transmit a high group (164 to 260 kHz) and receive a low group (44 to 140 kHz) or vice versa. Frequency frogging

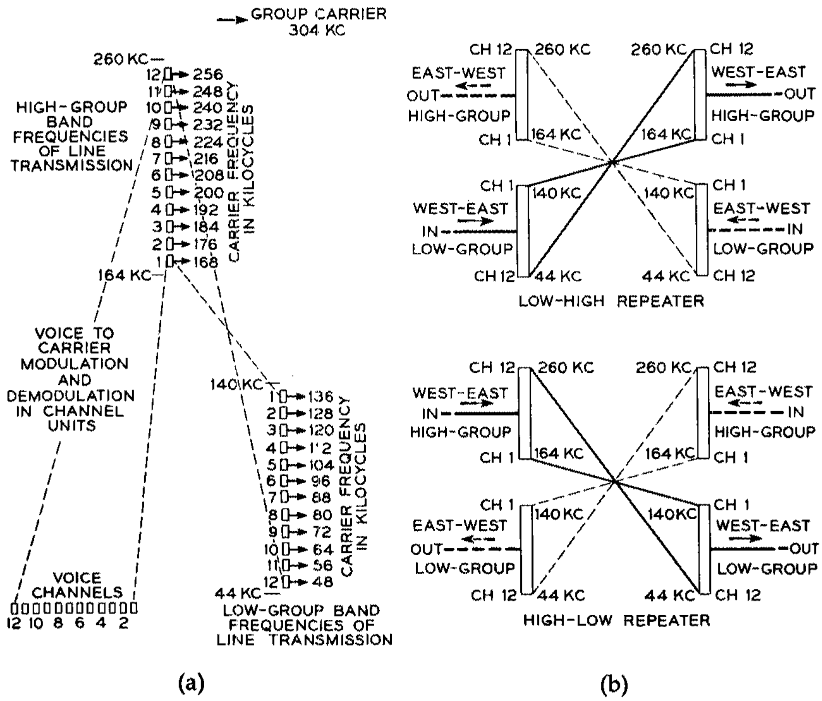


Fig. 9-1. Frequency allocations for Type N1 carrier. (a) Terminal-modulation plan for forming 12 double-sideband channels in a high group and a low group. (b) Frequency-frogging plan for inverting high-group and low-group frequencies at each repeater.

involved the interchange of the high and low transmission bands at each repeater point. All systems in one direction in a cable section would have about the same signal level at the same location, as would all those in the opposite direction. The widely different levels between received and transmitted signals at a repeater were in different frequency bands. The use of different frequencies and frequency frogging was to accomplish a result similar to that achieved by cable frogging in the Type K carrier [Fig. 9-2]. Through inversion of the channel order at each frogging repeater, a first-order equalization of the cable attenuation slope was also achieved. This permitted the use of simple, low-cost, flat-gain amplifiers at the repeater. The frogging repeater, however, required an oscillator, a pair of modulators, and a set of filters.[2]

Double-sideband transmission and the two-band approach made the top frequency over four times that of the Type K carrier cable system—260 versus 60 kHz. At these frequencies, the separate bands and frogging were not sufficient to reduce crosstalk interference to within the desired limits in the confines of a single cable. The needed additional reduction was achieved by the use of a compandor (compressor-expandor) in each channel at the terminals.

It had always been obvious that crosstalk would be much less troublesome

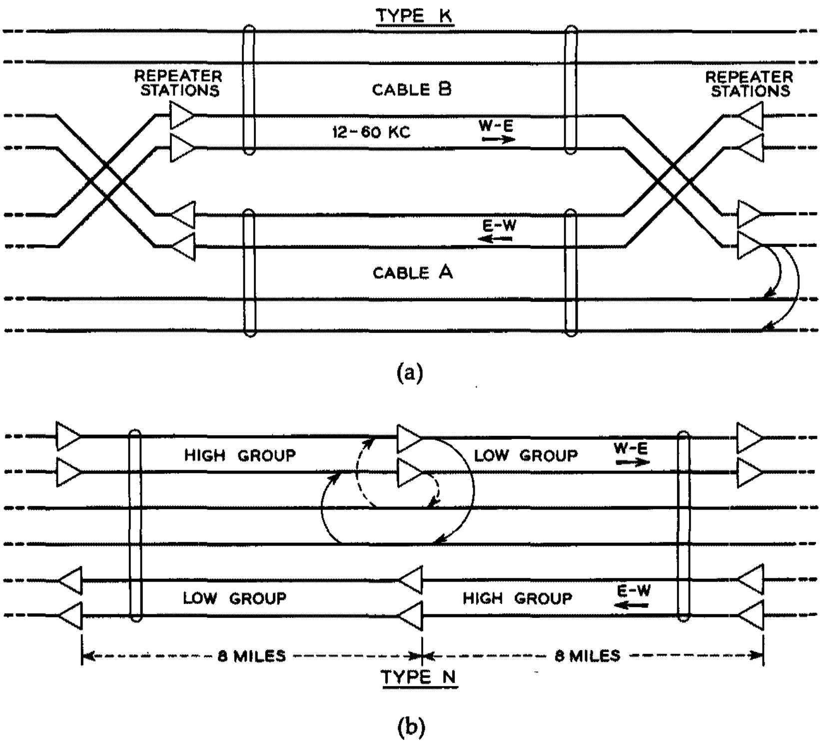
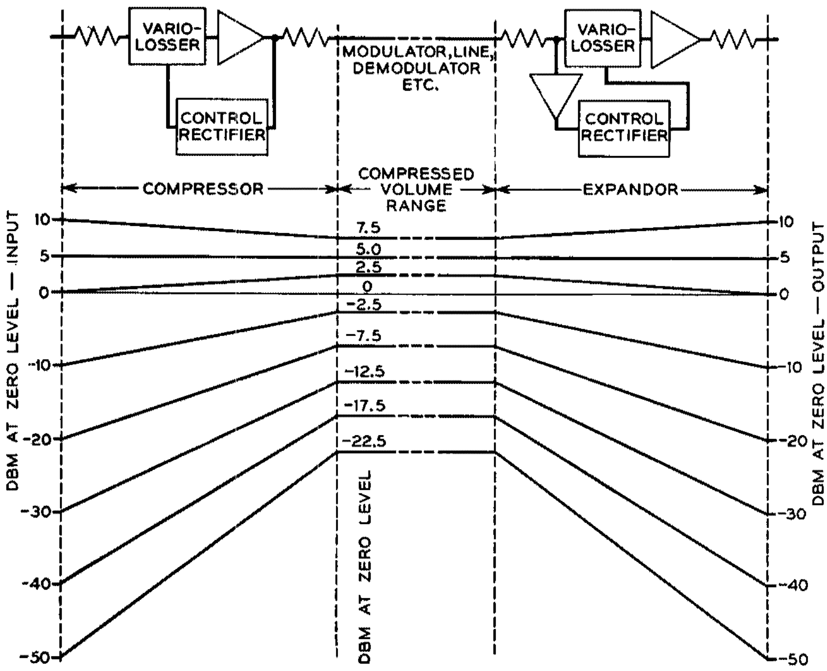
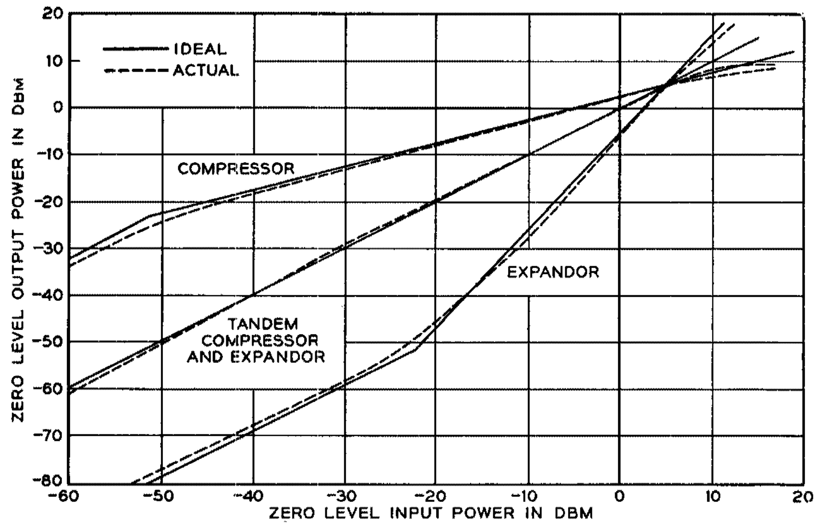


Fig. 9-2. Control of crosstalk by frogging. (a) Cable frogging of Type K carrier. (b) Frequency frogging of Type N carrier.



(a)



(b)

Fig. 9-3. Reducing the volume range and improving the signal-to-interference ratio by compandor action. (a) Compandor action on steady tones of different levels. (b) End-to-end transmission as a function of signal level for the Type N1 carrier compandor.

if all talkers spoke and were transmitted at the same volume. But talkers' habits varied widely, and the range of circuit losses for different connections further aggravated the problem. Crosstalk requirements had to be set so that the interference from the highest-level talker into the lowest-level disturbed channel was tolerable. The function of the compressor part of a compandor was to sense the talker signal level dynamically at a syllabic rate and, primarily, to raise low-volume talkers so that crosstalk and line noise were less interfering in their circuits. The compression reduced the volume range of low-level talkers by a factor of two in decibels. At the receiving terminal, an expandor sensed the reduced, but still appreciable, differencein levels and restored each channel to its original range [Fig. 9-3]. Compandors had been used for some time in overseas radio to reduce the effects of static and on open-wire carrier channels as an aid against crosstalk; but, for the Type N carrier system, a much more compact and cheaper unit was needed and developed. Compandors were of critical importance to N carrier, providing more than a 20-dB reduction in the subjective effect of crosstalk and other line-noise impairments. With their use, far-end crosstalk balancing was eliminated (it was probably not practical at 260 kHz in any case), repeater spacing was increased, and repeater linearity and terminal-filter requirements were eased.[3]

The advantages of compandors were so many and so substantial that it is not likely that any short-haul cable-carrier system with characteristics similar

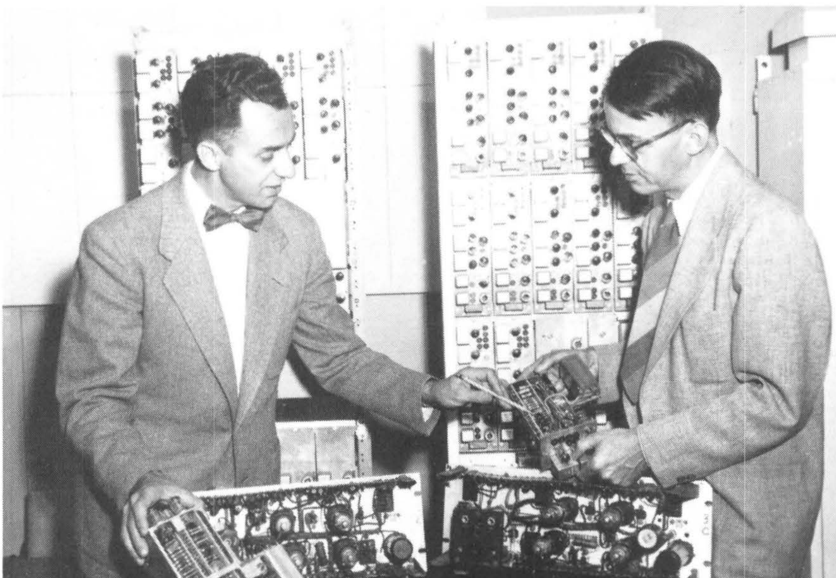
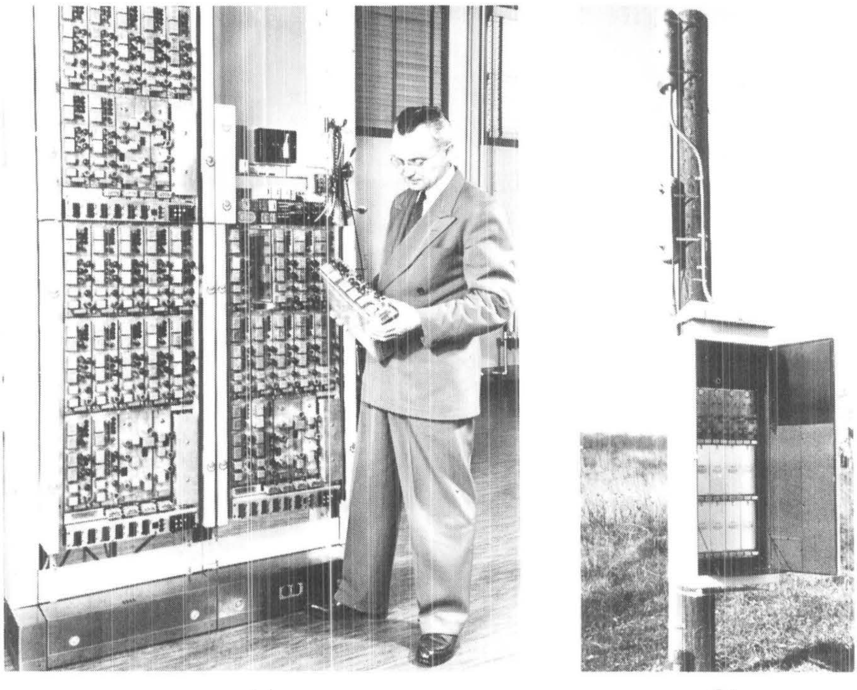


Fig. 9-4. Type N1 miniaturized compandor components of 1952. The older A1 compressor and expandor panels used in earlier systems are on the table.

to the Type N system could have been realized without them. However, their function was more complex than indicated in this simplified discussion, and there were disadvantages as well as advantages in their use. Some impairment could be introduced by imperfect volume tracking, and they amplified any line-loss variation between terminals. The time constants of the compressor and expandor were important design parameters. At least some listeners were somewhat aware of the variolosser action, which resulted in what was known as the hush-hush effect. As the design of N carrier became more focused on toll applications, there was apprehension about the behavior of several compandored links in tandem. Despite all this, when properly designed, compandors worked quite well and were a feature of several later generations of short-haul analog carrier. They remained a widespread and valuable, if not always loved, feature of the short-haul plant well into the 1980s [Fig. 9-4].

As the development work progressed, it was possible to make more accurate estimates of the cost of the N1 carrier system. It soon became evident that, despite the best efforts, the prove-in distance would be too large for most exchange-area applications. At the same time, it was confirmed that large savings were still possible in short-haul toll applications. This field of use had become



(a) (b)

Fig. 9-5. (a) Production N1 terminal equipment mounted on standard bays. (b) A pole-mounted Type N1 repeater cabinet. Four repeaters are at the top. Facilities were available for mounting eight additional repeaters in the cabinet.

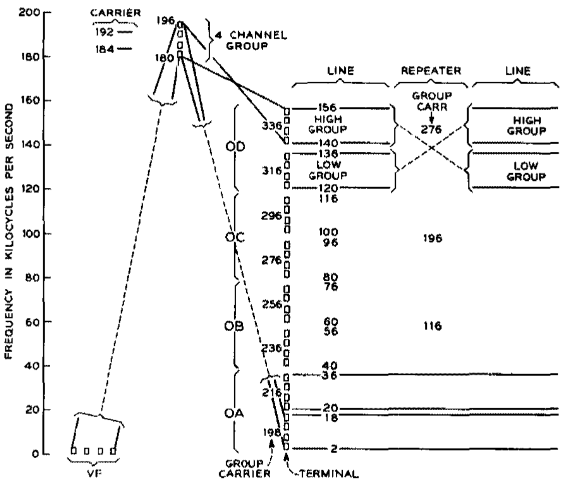
especially attractive as there was a need for improved short-haul links in the nationwide dialing plan. But the studies also indicated the need for further performance improvements in N1 carrier, implying greater cost. The system design was reviewed, circuit by circuit, component by component, and several design revisions were made. Line frequencies were increased slightly to ease filter and transformer realizations. Germanium varistors were selected as the compandor variolosser elements to improve stability. A per-channel regulator, much like the automatic volume control in radio, was added to the terminal to achieve the required net-loss stability. Overall, this effort resulted in a system better matched to the field of application, but the cost had been increased and the savings reduced.

Field trials were conducted over 29 miles of cable with three repeaters between New York City and Mt. Kisco, New York and over 80 miles with ten repeaters between Madison and Milwaukee, Wisconsin. Noise was essentially as predicted, verifying that the Type N1 system could not be operated in cables carrying Type K or M carrier systems (Type M carrier was a single channel system designed primarily for application to power lines) and that certain types of cable could not be used. The first production systems were installed in 1950 [Fig. 9-5]. The final design incorporated most of the basic ideas originally put forth as necessary for a low-cost system, though details were different. Originally planned for the exchange plant, where rather broad transmission limits would suffice, the system development evolved toward short-haul toll applications, with stricter requirements, resulting in increased cost and complexity. Nevertheless, the system was successful and widely used. By the end of 1953, there were more N1 channels in service than in either K or L carrier.

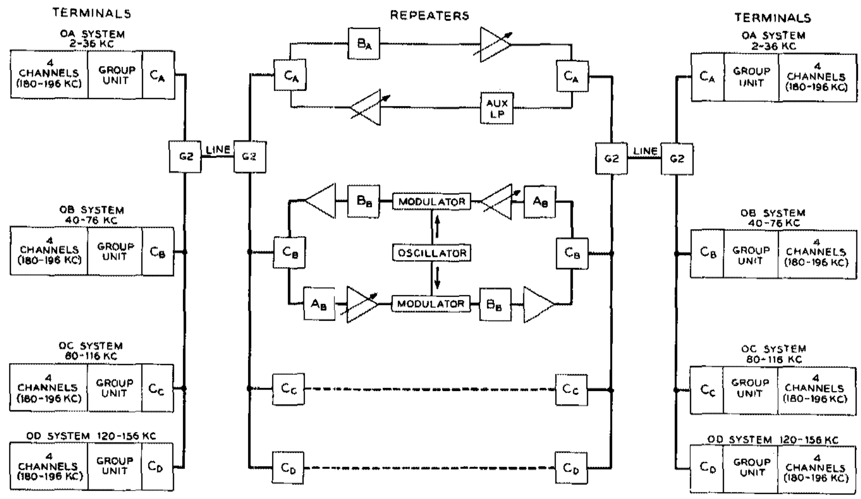
II. TYPE O CARRIER

The Type O carrier system was the open-wire counterpart of the N1 cable system. When O carrier was designed in the early 1950s, there were still about 1.5 million pair-miles of open wire in use for toll service. Types C and J carrier were used for the longer open-wire circuits, but, as more offices were converted for toll dialing, the need for short-toll trunks increased rapidly. The role of O carrier was to extend the economical use of carrier to open-wire circuits in the range of 15 to 150 miles.

The basic system module was a four-channel group, in effect, a four-channel stand-alone system. Four such groups, designated OA, OB, OC, and OD, each with a different frequency assignment, were installed as needed up to a total of 16 channels on a pair. Circuit cross sections in the areas of application were small, and the ability to install a few channels at a time at low cost was of paramount importance. Each four-channel system module had a low-group and a high-group frequency assignment. These were used to separate the two directions of transmission on a single open-wire pair. (In N carrier, opposite directions were on separate pairs.) Except in the lowest-frequency block, the two frequency bands of each module were interchanged, or frogged, at every repeater, as in the N1 system, to break up crosstalk paths via unequipped pairs and provide substantial equalization of the slope of line attenuation. The Type O repeaters included regulators that operated on the total four-channel group



(a)



(b)

Fig. 9-6. (a) Type O carrier modulation plan. (b) Type O carrier modulation assembly. The equipment was designed so that up to four independent four-channel modules, each with a different line-frequency assignment, could be assembled as needed. Filters designated *C* are directional filters; the *G2* filters separate and combine the four-channel groups at the line connection. *A* and *B* filters furnish additional discrimination for the four-channel band in each direction.

signal, including sidebands and carriers. This regulator handled a wide range of input-signal levels, as required by the severe variations of attenuation on open wire. A similar regulator was employed in the terminal group-receiving unit. Channel units performed the final, closer regulating function, operating on the separated carrier signal alone [Fig. 9-6].

The maximum frequency usable for O carrier was limited by crosstalk and by the extremely high loss of open wire at high frequencies under extreme weather conditions. A modulation scheme was adopted in which the highest line frequency was only 156 kHz, compared to 240 kHz in N carrier. The scarcity of pairs and the relatively narrow total band available caused the O system designers, after detouring to double-sideband transmission in N carrier, to return to single-sideband transmission. The sharp cutoff required to suppress the carrier and the unwanted sideband was achieved in small, low-cost filters by the use of newly available, low-dissipation ferrites in the inductors.

Translation to and from line frequency was a two-step process, involving an intermediate frequency band of 180 to 196 kHz. This band, which contained four channels, was common to all four system modules. Modulation to the common band and demodulation to voice frequency were accomplished by a novel twin-channel technique similar to that used in later overseas high-frequency (HF) radio. The channels were processed in pairs, using only one carrier-frequency oscillator per pair. Two balanced modulators first formed a double-sideband, suppressed carrier signal for each of two channels; then the lower sideband was selected for one channel and the upper sideband for the other. This resulted in a spectrum that looked like a conventional double-sideband signal, but each sideband carried a different channel. The carrier was reintroduced at an enhanced level for use in controlling the transmission regulators in repeaters and terminals. Two such twin-channel signals, each with a different carrier frequency (184 and 192 kHz), were used to form the basic four-channel group. The group-modulating oscillator unit furnished a 3700-Hz signaling tone as well. Equipment was mounted in plug-in modules, and pole-mounted repeaters were housed in a cabinet similar to that used for Type N1 carrier.[4,5,6]

III. FURTHER TYPES N AND O CARRIER DEVELOPMENTS

The Type O carrier development was quickly followed in the mid-1950s by an allied system: Type ON. The first Type ON equipment was developed to interconnect open wire and cable carrier systems at junction points more economically than by employing complete N1 and O terminals back-to-back. This equipment, known as an ON junction, consisted of Type O plug-in units with only minor modifications.

Twenty channel-frequency positions of O carrier (counting both directions of transmission) occupied a band from 40 to 136 kHz. This signal spectrum fell within the Type N carrier low-group band. In a connection from open wire to cable, the ON junction translated five four-channel O carrier groups so that they appeared in the ON band from 40 to 136 kHz [Fig. 9-7]. The five inputs could be any combination of O carrier groups and could be from different open-wire pairs.

Each four-channel group was first translated to the common band at 180 to 196 kHz, then to its assigned position in the ON band (and conversely for the other direction). The resulting signal went onto the N carrier pair, either without further modulation or after translation to the N carrier high-group range (168

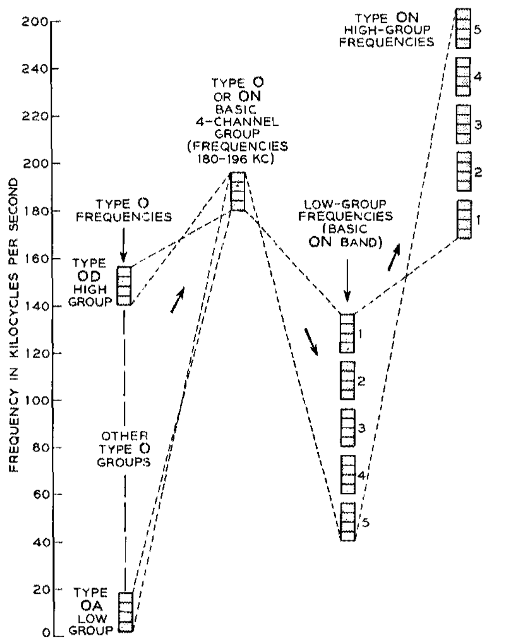


Fig. 9-7. Frequency translations between Type O and Type ON systems. *Left*, Type O frequencies are translated to high- and low-group frequencies that can be handled by Type N carrier line repeaters.

to 264 kHz), depending on which frequency band was required as an input to the N carrier line at that point. The frequency translation or straight-through connection was provided by an ON repeater, which was like an N1 repeater but the frequency shift was included or omitted, as required. The N carrier line was equipped with standard N repeaters, but the far-end terminal had to be either another ON junction or an ON terminal—a terminal like a Type O terminal, but with frequency assignments selected appropriately and without directional filters. These filters, used to combine the two directions of transmission on a single pair for O carrier, were not used in the ON application because the N carrier line operated on a physical four-wire (two-pair) basis.

It did not, of course, take telephone people long to realize that the ON single-sideband terminals could be used at both ends for pure-cable applications to increase the capacity of an N carrier line from 12 to 20 channels. This arrangement, dubbed ON1 carrier, was immediately economical on longer N carrier routes. The capacity of the all-cable ON system was later increased to 24 channels by closer packing of the four-channel frequency assignments and adding another four-channel group in the ON2 system design. (This was not practical in open-wire applications because the filters needed to separate the opposite directions of transmission required wider frequency spacing.) Both

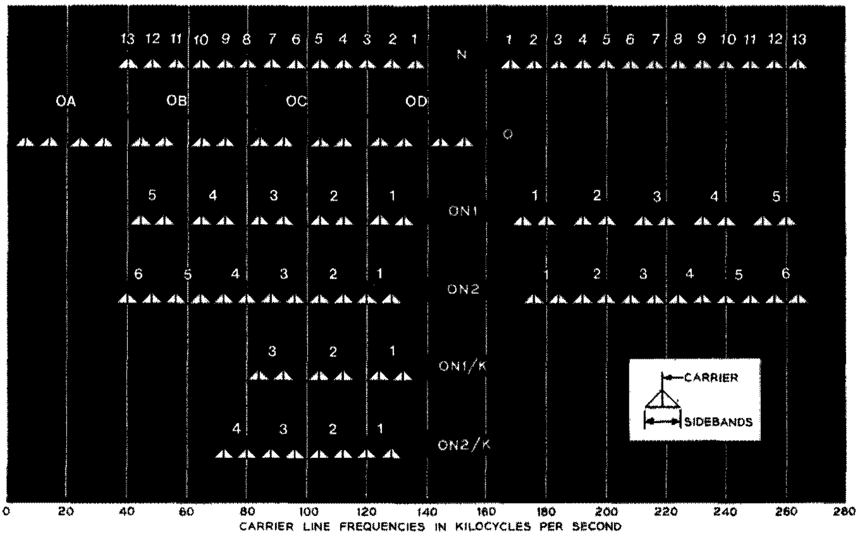


Fig. 9-8. Line frequencies of Types N, O, and ON short-haul carrier terminals.

ON1 and ON2 terminals were also applied to K carrier lines in 12-and 16-circuit packages [Fig. 9-8], and later, with terminals stacking as many as four ON2 systems for 96 channels, to short-haul radio.

The Types O, N, and ON carriers were developed in a typical sequence. The need for short-haul carrier initiated a project with difficult objectives. Its realization on one medium (paired cable) led to a design for another (open wire), which, coming later, embodied several advances in technology. The systems on different lines created a pressing need for economical interconnection. Meeting this need finally led to a transfer and further improvement of the techniques, resulting in higher-capacity systems on cable than originally envisioned or thought possible.[7,8]

IV. THE SOLID-STATE ERA

4.1 Type P1 Carrier

In the years immediately after World War II, the demand for rural service also increased, and the cost of providing it became very high as the price of copper and other materials rose rapidly. The transistor, invented in 1947, seemed to hold an answer. The Type P1 carrier system, intended for use on long open-wire subscriber lines, was the first attempt at an all-solid-state transmission system.

The P1 design used a four-channel, double-sideband, AM carrier-transmitted plan, with signals at 56 to 100 kHz from the central office to a remote terminal and signals at 12 to 56 kHz in the reverse direction. A field trial in 1954 and 1955 in Americus, Georgia was noteworthy for the first effort to use a solar battery as a commercial power source. In service, however, of the few P1 carrier systems installed, almost all used a lead-acid battery floating on rectified local AC power.

The P carrier system made use of the latest available, but in many cases still embryonic, technology. It marked the first application of germanium transistors in a carrier system. Tantalum capacitors, adjustable ferrite-core inductors, and other solid-state components were also used for the first time. Mounting these components and miniature relays raised new problems, as did the plug-in interconnection of printed-wiring boards. The major problem, however, was economic. In the 1950s, competitive options were posed by pole-mounted switching concentrators and less expensive pairs in plastic-sheathed cable. The use of cable became so general that P carrier capabilities had to be extended to work over routes served in part or entirely by cable. Only 700 P1 channels were installed. Subscriber line carrier was to remain a difficult hurdle for developers for many years to come. The lessons learned, however, were soon to be applied to the short-haul toll systems.[9]

4.2 The N1A Repeater

Direct distance dialing (DDD) was implemented very widely in the 1950s. An important feature of DDD was automatic alternate routing, which increased the number of transmission links likely to be switched in tandem. The increase in the number of short-haul carrier channels in the tandem links imposed net-loss stability and transmission-quality requirements considerably more stringent than the original N carrier design requirements. In the late 1950s and early 1960s, a broad plan for the development of a new short-haul carrier family was drawn up, based on the changed requirements. By 1960, transistors were competitive with vacuum tubes in gain, frequency range, and other characteristics. The strategy was to offset cost increases due to the required performance improvements, with savings from the lower power, smaller size, and lower maintenance costs of solid-state designs. The order of development in short-haul systems was a "transistorized" repeater (N1A), redesigned terminals (N2 terminals for N1 and N3 terminals for ON2), and, finally, the N2 repeater, a wholly redesigned version to replace both N1 and N1A.[10]

In the N1A design, the circuits of the N1 vacuum-tube repeater were changed as little as possible—just enough to accommodate the new element. This so-called transistorization of N1 was criticized at the time as an unimaginative use of the new device, a stubborn insistence on treating transistors as solid-state vacuum tubes. But the N1A design brought many of the advantages of solid-state technology to the rapidly expanding short-haul field without the considerable delay that would have been involved in completely repackaging and redesigning the entire system. The N1A repeater was developed by the end of 1960, and the next year saw the initial production used largely to satisfy intersite communication-facility needs at Air Force ballistic missile bases. The ability to have as many as seven cable repeater sections between power sources easily offset the somewhat higher repeater cost. (The original vacuum-tube design allowed only three sections.)

4.3 N2 and N3 Terminals

Initial shipments of the completely redesigned N2 terminal were made in 1962, with significant improvements in size, frequency response, stability, compandor tracking accuracy, intermodulation and thermal noise, and crosstalk [Table 9-1].[11,12] N3 terminals began to be deployed in 1964, providing a solid-state, 24-channel, single-sideband terminal to operate with Type N lines. These were economical for system lengths greater than about 35 miles, where the higher cost of a single-sideband terminal could be offset by transmitting twice the number of channels over the repeatered line. N3 terminals included many of the innovations and improvements introduced with the N2 design and also featured a common carrier supply and automatic frequency correction of frequency shift caused by the repeatered line.[13,14,15] N3 terminals were not only a considerable improvement over the vacuum-tube ON2 design they superseded, but outperformed the objectives set for them as well [Table 9-2].

TABLE 9-1: Comparison of N1 and N2 Terminals

		N1	N2
SIZE (Channels in one 11′6″ bay)		36 (19″-wide bay)	96 (23″-wide bay)
POWER (Watts per channel end)		28.8	4.8
PERFORMANCE (terminals back-to-back)	Channel gain-frequency response: Frequency in cycles for 3 dB points	220 and 3040	200 and 3400
	Net loss stability	—	4:1 improvement over N1
	Compandor tracking error: Average excess gain in dB at worst input power	1.4	0.07
	Peak overload power: dBm at 0 dB system level for 0.3 dB increase in circuit net loss	+5	+10
	Intermodulation distortion in one channel: dB below fundamental		
	2nd order (A = B)	31	30
	3rd order (2A = B)	23	34
	Noise: dBrnC message weighting at 0 dB system level		
	Background	<12	7-14
	Contribution due to speech loading in adjacent channels	—	13
	Crosstalk (in dB)		
	Worst equal-level coupling loss in dB	objective = 70	
	Far end	—	74
	Near end	—	95

TABLE 9-2: Performance of N3 Compandored Message Channels on Back-to-Back Terminals

	Objective	Expected Average Performance
Channel bandwidth	200	180
3-dB band edge-cycles	3450	3500
Idle channel noise—dBrnC0	16	8
Loaded channel noise—40 percent loading with 0 VU at 0 TL—dBrnC0	20	12
Compandor tracking—over 60 dB range, −52 to +8, error with respect to 0 level—dB	<±0.5	+0.1, −0.3
Channel distortion—2*A*-*B*; 0 dBm fundamentals at 0 TL—dB below fundamental	30	36
Net loss stability—6 months		
distribution grade (Std. Dev.)—dB	<0.5	<0.4
bias (mean)—dB	<0.25	<0.25
Crosstalk—0 dBm at 0 TL, 1—kc disturber		
NEXT equal level coupling loss—dB	74	88
FEXT equal level coupling loss—dB	74	>100

[Bleisch, *IEEE Trans. Commun. Technol.* **13** (1965): 361.]

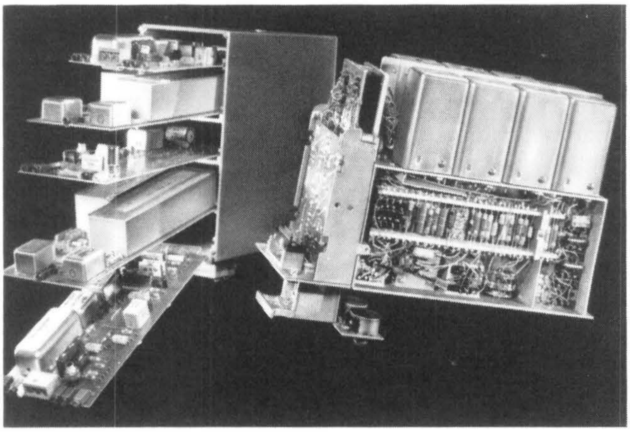
4.4 The N2 Repeater

The N2 line repeater, successor to the N1 vacuum-tube and N1A transistorized repeaters, was designed by 1966 to exploit the improved intertoll quality of N2 and N3 terminals. In addition to the latest available silicon transistors, the design used printed wiring, ferrite inductors, silicon-diode modulators, and other improved technologies. The repeater was designed to replace the older repeaters without reengineering the line.[16]

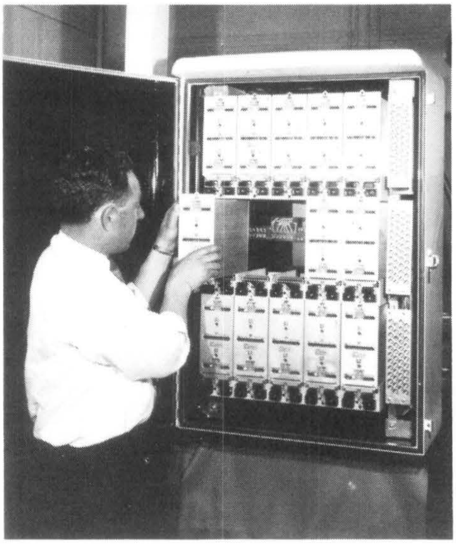
The major circuits of the repeater were on five plug-in printed-wiring boards contained in an aluminum housing that, in turn, plugged into a shop-wired cabinet that held up to 15 repeaters [Fig. 9-9]. The modular assembly provided a flexible arrangement that could be tailored to any specific location and eased manufacture, installation, and maintenance. New silicon units in place of the older copper oxide modulators and temperature-compensated filters gave transmission meeting requirements from −40 degrees F to +120 degrees F. New outdoor mounting arrangements were designed, but an adapter permitted placing the new repeater in one of the older N1 or N1A mountings, if desired.

4.5 The N4 Carrier System

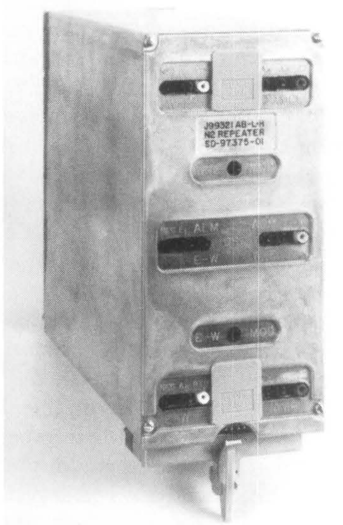
In 1973, a study was undertaken to evaluate the future need for carrier systems in rural areas and to interconnect smaller communities not on major carrier routes—the so-called out-state environment. The study was occasioned by the relatively slow spread of T1 digital carrier into this area and the need to determine the future of short-haul analog carrier systems. Everyone was



(a)



(b)



(c)

Fig. 9-9. (a) Solid-state Type N repeaters. *Left,* the completely redesigned N2 repeater of 1966. *Right,* the 1960 N1A design in which vacuum tubes were replaced with transistors with a minimum of circuit changes. (b) Type N2 repeater modules and cabinet. (c) Type N2 plug-in assembly.

eager to "go digital," but it became clear that there would be a continuing need for Type N carrier systems for some time. Many areas, generally the more sparsely populated ones, had sufficient cable plant in place, partially developed with N carrier, to last for a substantial period of time; and it was not possible to operate N and T carrier in the same sheath. The lack of compatibility of N and T carrier systems placed an economic barrier to the introduction of T carrier in those areas having existing N carrier. For the T systems, it was necessary either to place new cables or to remove the N systems and replace them with T systems while maintaining service. The latter option was generally not feasible without installing new cable plant. The procedure that naturally developed was to continue installing N carrier systems until the capacity of the existing cable plant was exhausted and then, when new cable went in, to provide digital systems on the new plant.

At about this same time, a number of technical developments took place in integrated circuits and filter design that had the potential of drastically changing the way in which N carrier systems could be designed. Among these new developments were the design of a compandor on a single silicon chip for a loop carrier system, a complete modulator on a chip for a coaxial system multiplex, and new small crystal channel filters for connectors between the L type analog-multiplexed signals and digital T carrier. By 1975, the continued need for short-haul analog systems and the opportunity for large size and cost reductions had been established. A proposal for the development of a new N4 carrier terminal was written and accepted in late 1975.

The new N4 carrier terminal was to be a replacement for the existing N3 carrier terminal, that is, it would furnish 24 single-sideband channels in the N line frequency band with two separate channels on each carrier. It was to be compatible with N3, that is, it would be possible to use an N3 terminal at one end and one of the new terminals at the other. However, the new N4 terminal was to have a different internal structure in order to be able to take advantage of the new technology. The major difference was to provide a two-step modulation scheme using, as a first step, the same frequency translation that was being used in the channel banks to form a standard group at 60 to 108 kHz. This allowed the use of channel filters and carrier frequency supplies nearly identical to those already in large-scale manufacture for coaxial and radio systems. The design had as its objectives a substantial reduction in cost over N3, a dramatic size reduction, the elimination of periodic maintenance, and a substantial reduction in the number of items that a user was required to order and assemble to make up a working system.

For N4, two identical sets of double-channel modems were used, each of which produced a 12-channel signal group in the range from 60 to 108 kHz [Fig. 9-10]. The modems each contained two crystal filters, each of which could select one of the sidebands from the modulation process. The 12-channel groups were then further modulated into the upper and lower halves of the N carrier high band and transmitted over an N carrier line. This signal could be applied directly to the line, or further modulated to place it in the low band, depending upon the needs of the office. The signals generated in the carrier supply were used both for the modulation process and as precisely regulated pilots to control the level of the speech signals as they passed through the terminals and line. The new smaller channel filters, integrated-circuit modulator, and compandor resulted in a terminal requiring only one-sixth the space and power of the N3 terminal it replaced.[17]

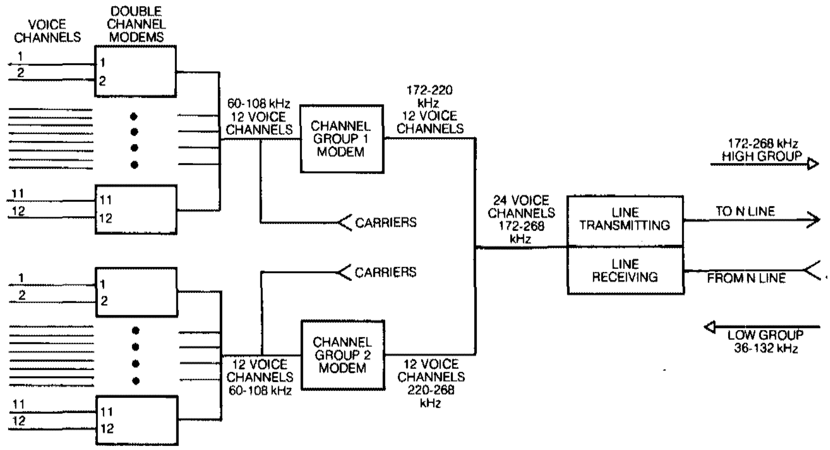


Fig. 9-10. The Type N4 terminal used to form 24 single-sideband signals in the N carrier high group (172-268 kHz). Double channel modems multiplexed two voice channels as opposite sidebands on a single carrier. The first modulation of 12 voice channels to the 60–108-kHz band permitted the use of low-cost filters.

The Type N carrier systems, which had a difficult time getting off the ground, ultimately proved extremely successful and were very widely installed. Over half a million channels and more than 40 million voice-circuit miles of N carrier of all vintages were installed by 1975. Because of the shrinking open-wire plant, O carrier was much less widely used, amounting at its peak in 1963 to about two million voice-circuit miles. But even this modest mileage, amounting to only five percent of the N carrier, was considerably greater than the maximum mileage carried on the earlier C and J open-wire carrier systems.

REFERENCES

1. O. H. Coolidge and D. K. Gannett, "The Birth of a New System," in *Transmission Aspects of Wire Communications Systems,* Communications Development Training Program Text, Vol. 1, Chap. 18 (New York: Bell Telephone Laboratories, 1951), pp. 1-36.
2. R. S. Caruthers, "The Type N-1 Carrier Telephone System: Objectives and Transmission Features," *Bell Syst. Tech. J.* **30** (January 1951), pp. 1-32.
3. R. D. Fracassi, "The Compandor in N1 Carrier," *Bell Lab. Rec.* **31** (November 1953), pp. 452-457.
4. P. G. Edwards and L. R. Montfort, "The Type-O Carrier System," *Bell Syst. Tech. J.* **31** (July 1952), pp. 688-723.
5. B. C. Griffith and H. Kahl, "Type-O Carrier: System Objectives," *Bell Lab. Rec.* **32** (June 1954), pp. 209-214.
6. R. D. Fracassi, "Type-O Carrier: System Description," *Bell Lab. Rec.* **32** (June 1954), pp. 215-220.
7. M. Aruck, "Type-ON: New Short-Haul Carrier," *Bell Lab. Rec.* **34** (November 1956), pp. 427-431.
8. J. B. Evans, Jr. and E. Homan, "The 'ON-over-K' Carrier System," *Bell Lab. Rec.* **38** (November 1960), pp. 413-416.

9. R. F. Clark, "Transistorized Rural Carrier System," *Bell Lab. Rec.* **34** (August 1956), pp. 281-284.
10. R. C. Boyd, "Redesign of Short-Haul Carrier Systems," *Bell Lab. Rec.* **41** (July/August 1963), pp. 275-279.
11. R. C. Boyd and F. J. Herr, "The N2 Carrier Terminal—Objectives and Analysis," *Bell Syst. Tech. J.* **44** (May/June 1965), pp. 731-759.
12. W. R. Lundry and L. F. Willey, "The N2 Carrier Terminal—Circuit Design," *Bell Syst. Tech. J.* **44** (May/June 1965), pp. 761-785.
13. D. T. Bell, L. H. Steiff, and E. R. Taylor, "Equipment Aspects of Packaged N3 Carrier Terminals," *Bell Syst. Tech. J.* **44** (May/June 1965), pp. 787-822.
14. G. W. Bleisch, "The N3 Carrier System: Transmission Aspects and Objectives," *IEEE Trans. Commun. Technol.* **13** (September 1965), pp. 355-361.
15. G. W. Bleisch and C. W. Irby, "The N3 Carrier System: Objectives and Transmission Features," *Bell Syst. Tech. J.* **45** (July/August 1966), pp. 767-799;
R. L. Haner and I. E. Wood, "Circuit Design of the N3-Carrier Terminal," *Bell Syst. Tech. J.* **45** (July/August 1966), pp. 801-844; and
W. J. Giguere, L. A. Bowie, and J. A. Watters, "Equipment Design of N-3 Carrier Terminals and Carrier Supply," *Bell Syst. Tech. J.* **45** (July/August 1966), pp. 845-894.
16. T. B. Merrick and W. P. Michaud, "A New Carrier Repeater," *Bell Lab. Rec.* **44** (July/August 1966), pp. 232-235.
17. T. R. Hull and R. E. LeBlanc, "New Life for Short-Haul Analog Routes," *Bell Lab. Rec.* **57** (November 1979), pp. 270-274.

Chapter 10

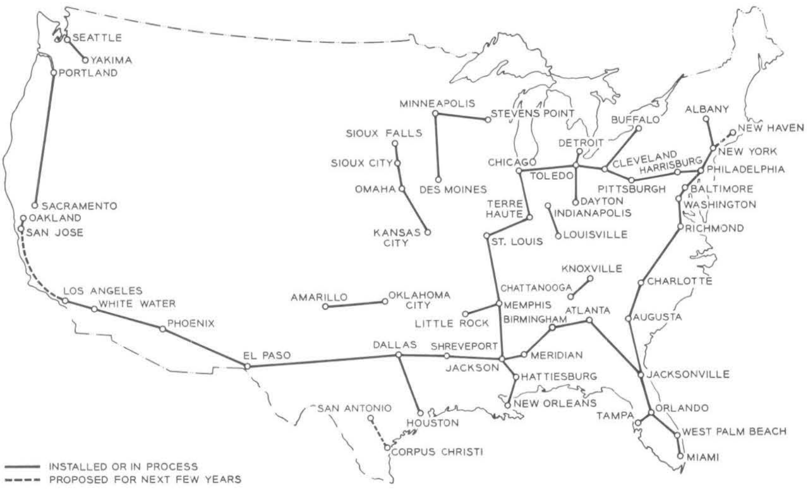
Coaxial Systems

I. THE L3 COAXIAL SYSTEM

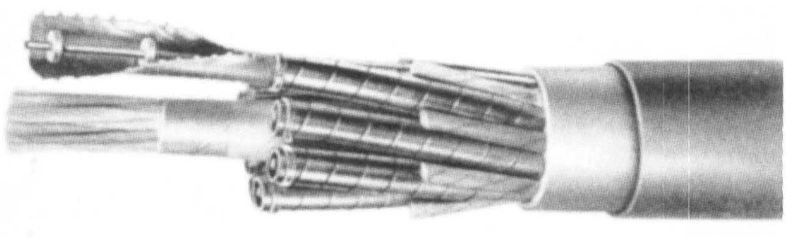
1.1 The System Plan

In early 1944, AT&T announced an ambitious program of postwar coaxial-cable installation. In December of that year, a few months after the announcement of the plan to develop and install an extensive microwave network, AT&T also announced a plan to develop a new coaxial system of about 7-MHz bandwidth. The new system was to be capable of carrying two to three times the L1 telephone channel load, or several hundred telephone channels simultaneously with a high-quality television signal. Late in 1945, as soon as the reduction in military work permitted, planning and exploratory development was started on the new system, designated the L3 coaxial system. (L2 was a small-scale effort to adapt coaxial to light routes and short distances. It was abandoned at an early stage.) The L3 system was to obtain the maximum channel capacity on existing cables, taking advantage of all the advances in components and design techniques since the prewar L1.

Since a major field of use of L3 was to replace L1 on existing routes, the L1 design, the cable, and the cable-route layouts presented the L3 system with a definite plant framework. There were four broad objectives of the L3 system design. First, the existing in-place cable was to be reused. Thus, the loss and its variation with temperature, the irregularities due to manufacturing and splicing, and the power-transmission capabilities of the cable became basic restrictions on the L3 system design. By 1953, the year of first commercial service of L3, there were about 8,000 miles of cable installed, of which about 70 percent consisted of eight-coaxial-unit cables; the remainder consisted of six- or four-unit cables [Fig. 10-1]. About 70 percent of the installed cable used coaxials with an outer diameter of 0.375 inch; the remainder used the older coaxial tubes with a diameter of 0.27 inch. Second, as much as possible of the L1 telephone terminal equipment was to be reused, including channel banks, group and supergroup equipment, and carrier supplies. Third, the existing L1 repeater locations and buildings were also to be used as far as possible. The loss at the higher L3 frequencies was much greater than at the top L1 frequency (3 MHz). Closer repeater spacings were required. Since the L1 line repeaters were spaced at eight-mile intervals, this virtually dictated a four-mile spacing for L3. The L1 main stations from which power was supplied to the line repeaters over the coaxial cable were spaced at 40- to 60-mile intervals because of geo-



(a)



(b)

Fig. 10-1. (a) Coaxial-cable routes, 1953. (b) An eight-coaxial-unit cable, 1953.

graphic and power transmission considerations. Fourth, as experience was gained with the L1 system, it was anticipated that its capacity would be increased to 600 channels. L3 was to provide about three times as many channels, all meeting long-haul objectives. Alternatively, 600 channels, the full complement of L1, were to be transmitted simultaneously with a full-bandwidth television signal of at least 4 MHz bandwidth. The proposed frequency allocation, allowing for 1860 telephone channels, required a band of about 8.3 MHz [Fig. 10-2].[1,2]

The L3 system required not only the most advanced components in hand, but also considerable extension of the existing art. New electron tubes, transformers, inductors, and other circuit elements were designed for improved precision, stability, and other performance factors. New techniques of network synthesis were developed to provide precise transmission characteristics over the wider frequency band. Techniques were devised for achieving hitherto unattainable accuracies in the measurement of impedance, loss, phase, and other transmission properties.

While much new technology was needed, the L3 coaxial system was a second-generation system. There were problems enough, but no major component's basic feasibility was in doubt. Issues of that type had absorbed much of

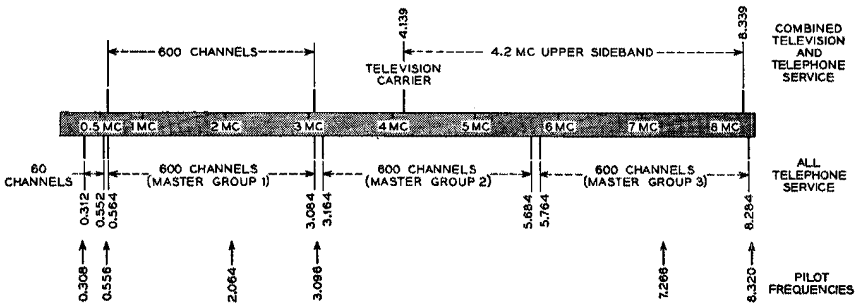


Fig. 10-2. Frequency allocations in the L3 system.

the effort on the first coaxial system and were to do so on the first microwave project. With L3, there was little doubt that a system of the general type projected could be developed. The questions were how best to go about it and how to squeeze the maximum capacity and performance out of the medium and components. Under the project leadership of C. H. Elmendorf* the system design was approached, for the first time in a major transmission development, with all factors systematically accounted and weighed from the outset. A strict budget of performance results for the various components and parts of the system was kept, and, as the pieces fell short of, or (more rarely) exceeded, expectations, compensations and trade-offs were made to optimize the system as a whole. It was, in many ways, the first total transmission *system* development and established a pattern that was to be followed in many later system projects.

A basic tool in this total-system approach was the use of statistical quality-control techniques in manufacture and testing. System components were made and delivered, not only within specified limits, but meeting requirements on the mean value and distribution of characteristics as well. This approach made it possible to design the system on the same basis, without having to allow for a combination of worst-case limits in the overall design. With a design for the mean values and margins allowed for the sum of the known distributions, the system could be realized with a low probability that it would be out of limits. A good deal of this sort of approach had been applied on a seat-of-the-pants basis in the past. Many systems would not have worked if a large portion of the components were near or systematically biased toward the test limits. In the L3 system the statistical method was applied uniformly on a system-wide basis for the first time.

1.2 Repeatered Line

With the prospect of 1000 repeaters in tandem in a 4000-mile system, the requirements on each repeater became very tight indeed. The L3 repeater was the basic unit of the system, and its design determined to a great extent the

---

* Elmendorf, before advancing to executive positions at AT&T, was "Mr. Coaxial Cable" at Bell Laboratories. He took part in the development of every system from working as a recruit on L1 to directing the development of L3 and L4 and initiating the work on L5.

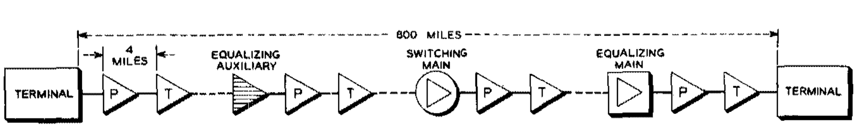


Fig. 10-3. Typical arrangement of repeaters along an L3 route. *P* is the pilot-controlled auxiliary; *T* is the thermometer-controlled auxiliary.

performance and economics of the system. There were two principal requ'rements to be met. The repeater had to offset the loss of four miles of cable under all temperature conditions while contributing low enough noise to meet overall signal-to-noise-ratio limits; and the repeater, plus addition equalizers, had to provide a transmission characteristic with gain and delay equalization adequate for the broadband signal [Fig. 10-3].

1.2.1 *Noise*

The random-noise design was fairly straightforward. The line noise was the thermal noise of the conductors and was accurately known. The tube noise was only a few decibels higher and was reproducible from tube to tube. The major problem was with interference from intermodulation products as a result of repeater nonlinearity; and, of the two system configurations, the combined television and telephone signal was by far the most difficult to analyze and control. Intermodulation between telephone channels produced modulation products whose sum was similar to thermal noise and could be treated the same way. Cross modulation between the television signal and the telephone channels resulted in a host of specific products that were potentially less tolerable than random noise and that had to be examined in detail. Of 400 types of products that were analyzed, six types were found to be most limiting. All intermodulation products were reduced by feedback; but, because of the wider band, the feedback in L3 was low compared to earlier systems. The cross products between television and telephone signals were further reduced in their interfering effect by locating the television carrier so that the worst products fell at frequencies that were attenuated by the cutoff of the telephone channel filters. Despite these measures, the combined-signal intermodulation was a nagging problem throughout the development of L3. The problem became academic in time as in-service L3 was almost never used in the combined-signal mode.

1.2.2 *Regulation and Equalization*

Equalization is the compensation of a band in gain and phase to a characteristic adequate for the signal to be carried. Regulation is automatically adjustable equalization, intended to keep the transmission characteristic within limits as changes occur, most typically because of temperature changes. In the L3 system, two main classes of equalizers were employed, fixed and variable. The latter was further divided into those intended to take care of changes the nature of which could be predicted, i.e., "cause-associated shapes" and those of which the shape could not be predicted [Fig. 10-4].[3]

The L3 line repeater was designed to compensate for the loss of four miles of three-eighths inch coaxial cable and thus was the first step of line equalization

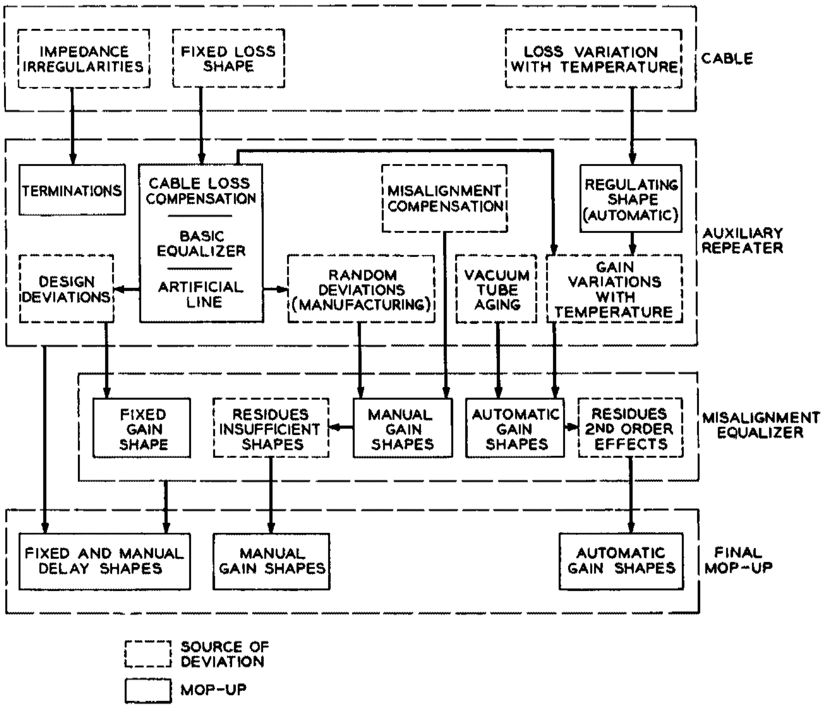
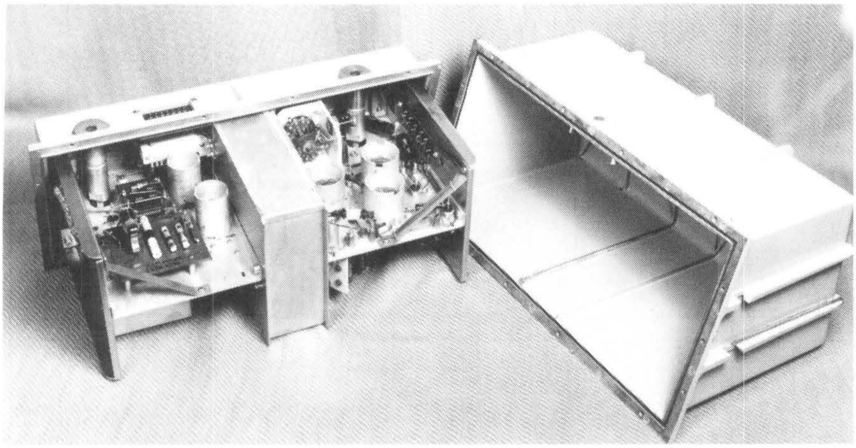


Fig. 10-4. The L3 coaxial-system equalization plan. A gain and delay equalization budget was developed allowing for each source and equalizer.

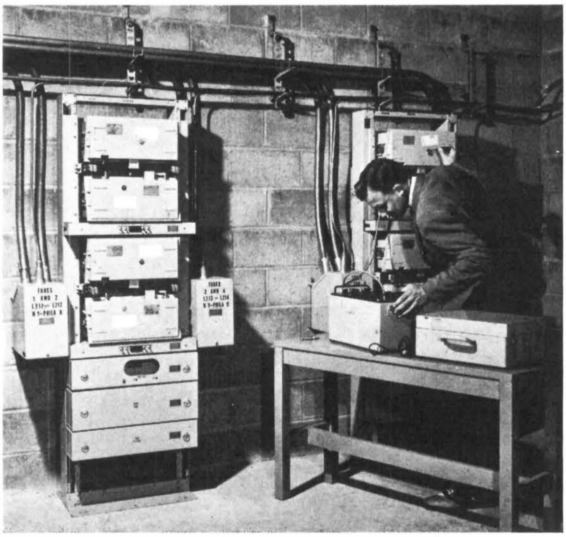
[Fig. 10-5].[4] In addition, line repeaters were equipped with artificial lines that were used to extend (build out) the loss of short sections to the equivalent of four miles of cable and with basic equalizer options that provided for differences in loss characteristics of the different types of cable. A second step of fixed-gain equalization was known as a design-deviation equalizer. Its function was to correct the accumulation of known deviations caused by the failure of the average line repeater to compensate exactly for cable loss. These equalizers were used at 40- to 120-mile intervals. When television was transmitted, fixed delay equalizers were also used at approximately 150-mile intervals. These equalizers compensated for the known delay distortion introduced by the line repeater sections.

Corrections for most changes in transmission were performed automatically by pilot-controlled regulators. Pilot tones were transmitted over the line at a fixed reference level at the transmitting terminal. At intervals along the line, regulators tuned to the pilots measured the deviation in pilot levels from the reference values. Depending on the deviation, the regulator increased or decreased the current to a thermistor-controlled regulating network to restore the pilots to the correct value.

The first step of automatic gain equalization was provided at each line repeater. The nominal gain characteristic of the repeater was designed to match the loss characteristic of the coaxial at 55 degrees F. As temperature changed,



(a)



(b)

Fig. 10-5. (a) L3 line repeater amplifier and housing. (b) L3 repeaters mounted in a repeater hut. The engineer is checking transmission by measuring the level of pilot tones.

the cable loss changed, but the change was predictable. It was the first major cause-associated equalization shape, being very closely proportional in decibels to the square-root of frequency. To compensate for these changes, two types of regulating circuits were used. The first adjusted the gain-frequency shape of the line repeater in accordance with the magnitude of a 7266-kHz pilot. The second type, a thermometer regulator, adjusted the same gain-frequency shape under control of a thermistor buried in the ground near the cable, thus measuring its temperature. Such control was not as accurate as pilot-controlled feedback regulation but was adequate for use in one-half of the line repeaters. Its simplicity resulted in considerable savings in first-cost and power requirements.

Additional steps of automatic equalization were provided to offset gain variations caused by vacuum-tube aging and changes in repeater-hut temperatures. These equalizers were used every 40 to 120 miles and were controlled by 308- and 2064-kHz line pilots. In the final step of automatic gain equalization, pilot-controlled networks were provided to compensate for second-order deviations remaining after the first-order corrections. The transmission characteristics of these networks were under the control of 556-, 3096- and 8320-kHz pilots. The second-order equalizers were located at approximately 150-mile intervals.

The cause-associated shapes were broad effects that covered the entire band. No one pilot was an adequate measure of a specific shape. Because of these broad effects, the thermistors controlling the loss frequency characteristics of the automatic equalizers were driven by the regulators through a simple form of analog computer. By a process equivalent to the solution of simultaneous equations, all the pilots were used to determine the amounts of an equal number of cause-associated shapes needed to restore the pilots to normal. The computer translated pilot errors into shape errors and drove the appropriate regulating networks the proper amount to obtain the corrections.

For the inevitable errors in correction and unpredictable residues from the automatic equalizers, manually adjustable mop-up equalizers were provided. When television signals were transmitted, manually adjustable delay equalizers were also provided at approximately 150-mile intervals. These equalizers were needed to trim the delay characteristic in finer detail than was possible with the fixed equalizers. The manually adjustable mop-up gain equalizer was based on an ingenious concept. It was observed that any gain shape (i.e., gain plotted against frequency) could be analyzed into a series of cosine terms by a Fourier analysis of the band transmission characteristic. The equalizer provided a number of harmonically related cosine shapes that could be adjusted with the appropriate sign and to the magnitude needed to flatten the band [Fig. 10-6].[5] The analysis was not actually carried out. Instead, on an out-of-service basis, a test set swept a tone across the band, skipping the pilots. The amplitude *variations* of the swept tone were then rectified and the resulting DC minimized by adjusting the cosine shapes [Fig. 10-7].

The various fixed and adjustable equalizers of the L3 system were combined in packages appropriate to the length of the system to be equalized. Each switching section had to be adjusted within limits so that it could be used in combination with others in a long system without re-equalization. The entire plan was applied to sections with a maximum length of 800 miles at which point the line signal was demultiplexed to the group or supergroup level and reassembled in a different order to break up the systematic accumulation of modulation products.

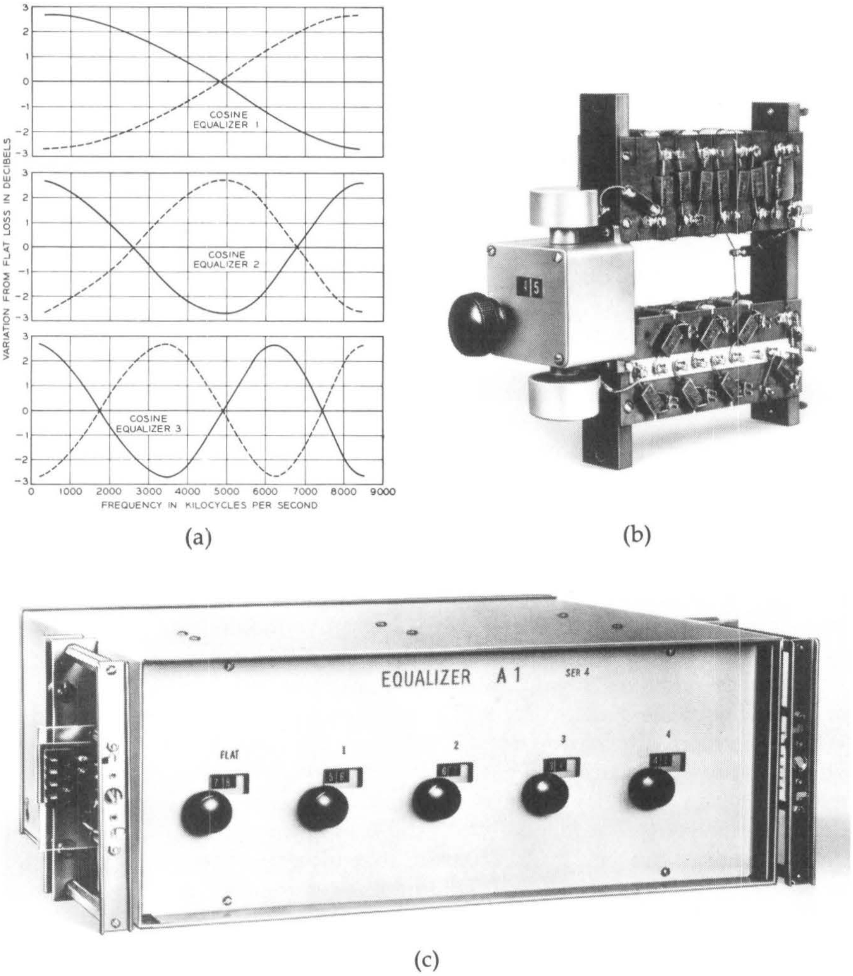


Fig. 10-6. (a) Shapes introduced by the first three L3 manually adjustable equalizers. Each equalizer could be adjusted between the limits of the solid and dashed curves. (b) L3 coaxial-system adjustable-cosine equalizers. Single-cosine-term network. (c) Assembly of five cosine terms. Up to 25 cosine-term networks were used for final equalization of a long L3 line.

1.3 Terminals

Additional stages of modulation beyond what was provided in L1 terminals were required to assemble the L3 telephone line signal.[6] Ten supergroups were stacked into a 600 channel mastergroup in the frequency range from 564 to 3084 kHz by modulation techniques similar to those described in Chapter 6, Section 4.1. One of these (mastergroup 1) was applied directly to the line. The remaining two were shifted by additional stages of modulation to the assigned line-frequency positions to achieve the desired signal spectrum. The television

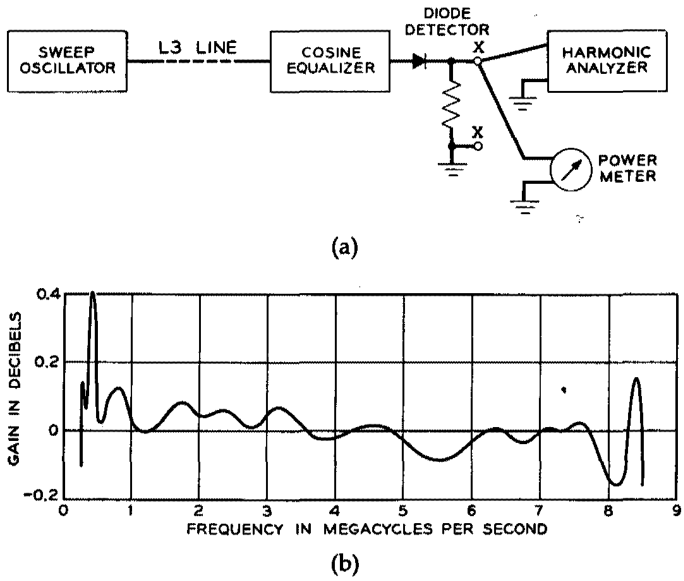


Fig. 10-7. (a) Equalization of an L3 line by a cosine equalizer. Each cosine term was adjusted to minimize the voltage variation across the terminals *x-x*. (b) Final gain characteristic of a 100-mile L3 line after equalization with 15 cosine shapes.

signal was translated to the line by techniques similar to those used in the L1 terminal. The major difference was in the use of a television line-carrier at 4.139 MHz with an upper sideband extending to 8.3 MHz instead of the lower frequencies of L1.

1.4 Trial and Service

The L3 system was installed for a field trial between New York and Philadelphia in 1951 and 1952 and was opened for service in February 1953. A link from Philadelphia to Chicago was completed by mid-1954, and, by September of that year, 1500 channels were being carried on the heaviest sections.

The L3 system provided good telephone transmission service for many years. The system provided a substantial first-cost savings over the cost of transmitting a large number of channels on L1; and, in a few years, it was carrying more circuit miles than the earlier system. But it was plagued with a number of problems that were only partially solved. This was due in part to the fact that the technology had perhaps been stretched too close to its limit but also to the reassignment of a part of the development staff to military work, which was expanding at the time. There was no follow-up coaxial project; and, with the reassigned staff and expanding radio projects, there was no natural home for the clean-up work essential to a smoothly working system. In L3, vacuum-tube life was unsatisfactorily short, and it was difficult to localize marginally functioning repeaters. Equalization for television in the combined telephone-television mode was never satisfactory for the full 4000-mile objective, although programs were started to improve the equalization to permit television transmission up to 1000 miles. Microwave radio far outstripped coaxial cable in television-channel mileage, and very little television was carried over L3.

II. THE L4 COAXIAL SYSTEM

The decade from 1955 to 1965 was one of extraordinary growth in the transmission plant. Microwave radio, with its low start-up cost, short construction intervals, and superior ability to carry signals for the rapidly expanding national television network captured the lion's share of the growth: it increased in capacity by a factor of ten. Coaxial systems, in contrast, often involved time-consuming acquisition of a continuous right-of-way, considerable construction, and higher initial costs. Nevertheless, coaxial systems provided low-cost circuits on routes with very heavy telephone cross sections and rapid growth to their full capacity. In the same decade, telephone circuit mileage on coaxial cable more than tripled.

The advantages of radio caused a long hiatus in new coaxial development following the start of service on the first L3 link in 1953. However, by the early 1960s, additional incentives for further coaxial development appeared. Once a cable was in the ground and much of the plant at auxiliary and main stations built, additional circuits could be obtained at relatively low costs by upgrading a route with a broader-band system. (This "mining" of circuits was an extremely important part of the network-capacity expansion.) Coaxial systems also offered a needed diversity to overcome system vulnerabilities. They could be built entirely underground and hardened to make them less likely to be disrupted by natural disasters or nuclear attack. By the early 1960s, progress in transistor technology made it possible to consider a system with capacity much higher than L3 and with the added attraction of the high stability and reliability promised by an all-solid-state design.

As always, the new system had to conform to the pattern of the existing plant. The widespread existing L3 installations, with repeaters spaced at four miles, required a spacing for the new system at a submultiple of that distance. Two or even three intermediate repeaters might have offered very broadband possibilities, but the maximum frequency attainable in devices with the required life and reliability quickly resolved the spacing in favor of one intermediate repeater and a two-mile interval. In 1963, work on the new system, designated L4, got under way with the target of commercial operation in 1967, an extremely short schedule for a complex system.

2.1 The L4 System Plan

Exploratory studies were conducted simultaneously in different areas to determine what could be accomplished in the new system. At the system level, developers concentrated on optimizing line-repeater amplifier characteristics such as linearity, noise figure, bandwidth, and output power. At the device level, other designers sought to optimize these same characteristics by advances in transistor design and processing. From these studies, the L4 system capacity was established as six 600-channel mastergroups or 3600 two-way voice circuits per coaxial-tube pair in the band from about 0.5 MHz to 17.5 MHz—approximately double the capacity and bandwidth of L3 [Fig. 10-8].[7,8]

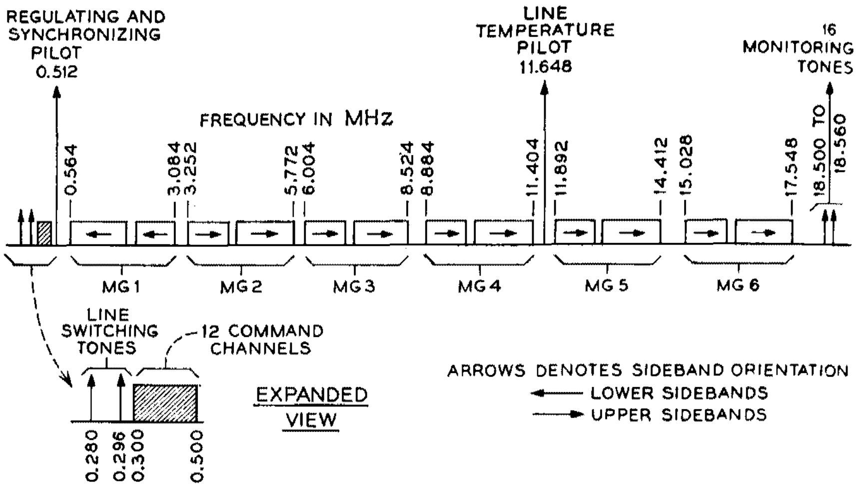
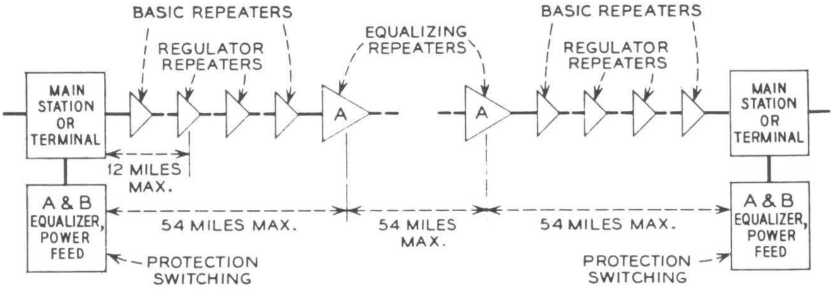


Fig. 10-8. L4 frequency allocations. The line-signal spectrum.

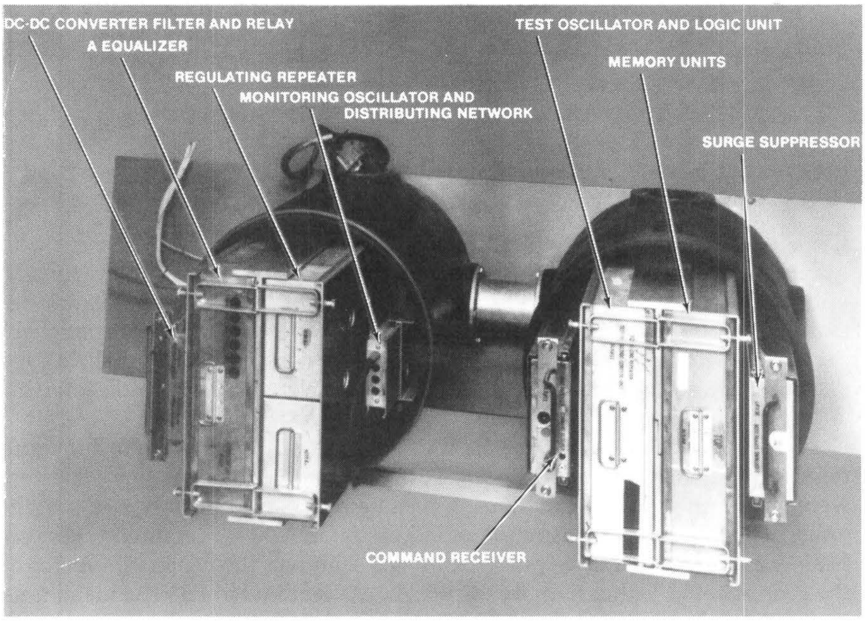
Fixed-gain basic repeaters were placed at two-mile intervals to compensate only for the nominal loss of the coaxial cable. Pilot-controlled regulating repeaters were inserted at 12-mile intervals to compensate for the variation of cable loss with seasonal temperature changes. More complex equalizing repeaters to correct for accumulated gain deviations were used at 50-mile intervals and at main stations, spaced about 150 miles apart [Fig. 10-9].

The L4 repeaters were smaller, required less power, and were far more reliable than their L3 predecessors. Both the basic and regulating repeaters were single plug-in units; equalizing repeaters consisted of only two plug-in units.[9] The low power consumption and moderate voltage required for L4 repeaters allowed them to be powered by DC sent over the center conductor of the coaxial cable, rather than by the 60-Hz AC required in L3. The mean time to failure of basic repeaters was projected as 20 years, so visits to repeater stations by maintenance personnel were expected to be infrequent. The characteristics of small size, low power consumption, and high reliability simplified the physical design of repeater stations. Some coaxial repeaters had previously been placed in special manholes, but most L3 repeaters were housed in aboveground huts. In L4, basic repeaters to serve four coaxial cables were mounted in a cylindrical, waterproof apparatus case about 16 inches in diameter. Five apparatus cases, adequate to serve 20 coaxial tubes, could be contained in a simple six- by ten-foot underground enclosure. The larger regulating and equalizing repeaters also plugged into the waterproof case and were mounted in manholes. These arrangements allowed standard L4 equipment to be constructed as a fully hardened system resistant to natural disaster and enemy attack.

Main-station buildings, at 150-mile intervals along L4 routes, were also built as hardened underground buildings, but, in contrast to repeater stations, the main stations were large structures similar in internal appearance to conventional



(a)



(b)

Fig. 10-9. (a) Typical L4 coaxial-system repeatered line layout. Fixed-gain basic repeaters were at two-mile intervals. Regulating repeaters to offset cable-loss changes due to temperature were spaced 12 miles apart. More elaborate *A* and *B* equalizers were at nominal spacings of 50 and 150 miles, respectively. (b) L4 equalizing repeater apparatus case. The left cabinet houses the transmission circuits; the right cabinet houses remote-control circuits.

telephone offices [Fig 10-10]. These buildings were designed and equipped to operate for several weeks completely sealed off from the outside world in the event of a nuclear attack. The functions performed at main stations varied. At a minimum, each station contained repeaters similar in design to the line-equalizing repeaters. In addition, each station provided power to the L4 lines

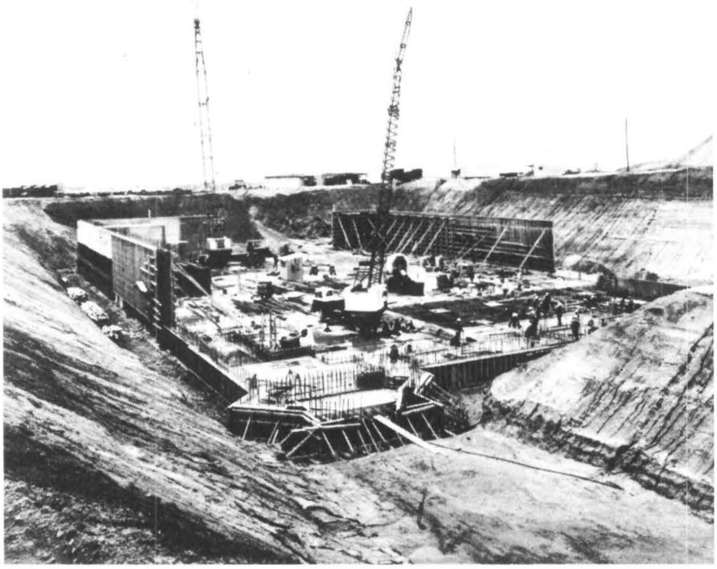


Fig. 10-10. One of the bomb-resistant main stations for the L4 coaxial system under construction on the Washington–Miami route, 1966.

and contained line-protection switching equipment. This equipment monitored the performance of all coaxial lines and automatically switched service to a protection coaxial in the event of failure of one of the working tubes. Larger main stations were equipped with remote control consoles that allowed maintenance personnel to monitor and control the condition of coaxial lines extending up to 300 miles from the station. The larger stations were also the site of multiplex equipment for constructing the line-signal spectrum or for adding or dropping major blocks of circuits.

Sixteen test tones spaced throughout the transmission band were generated at equalizing repeaters and used to measure deviations in the transmission response. Although these tones were located in guard spaces between the message circuits so they could be applied with the line in service, they were not continuous but were activated from the control consoles only when periodic equalization measurements and corrections were needed. Several system-control signals were clustered at the band edges. Equalizing repeater and line-protection switching-control signals were placed at the lower end of the band from 280 to 500 kHz, and repeater fault-location tones were grouped at the high end of the band, above 18 MHz. In addition to the temperature-regulation pilot at 11.6 MHz, pilots were transmitted at 512 kHz and 20.4 MHz.

2.2 Repeatered Line

Major advances in transistor technology took place concurrently with the design and development of the L4 system, and they were introduced to provide signal amplification with minimal addition of noise and distortion. Although great strides had been made in increasing the life of vacuum tubes, transistor life was expected to be orders of magnitude greater at considerably lower cost. Two types of transistors were developed with specific characteristics to meet the stringent requirements placed on the repeaters. One type was tailored to provide low noise and distortion with maximum gain-bandwidth product at low power. The second type, a medium-power device, was developed to minimize distortion at a higher signal level and also to provide a tightly controlled gain characteristic. Both types were n-p-n silicon transistors using planar epitaxial technology and were used in all repeaters.

2.2.1 Basic Repeater

The loss of two miles of a coaxial unit over the frequency range from 0.564 to 17.548 MHz varied from 6 to 33 dB. The L4 basic repeater was designed with a gain shape to offset the loss at 55 degrees F to within a few percent. Repeater sites, however, could not always be placed at exactly two-mile intervals because of roads, river crossings, or other placement restrictions. When restrictions were encountered, the site was selected to make the distance less than two miles, and an artificial line (a line build-out network) was used to compensate for the difference. One of a set of networks, which inserted cable-shaped loss corresponding to one-tenth mile increments, was combined with the amplifier on a plug-in basis to match the shorter distance [Fig. 10-11].

Amplification of the signal was in two separate circuit packages, a preamplifier designed for low noise and a power amplifier, where signal distortion was minimized. Each circuit configuration was simple but built to exacting standards [Fig. 10-12]. The line build-out network (LBO) was located between the two amplifiers, a better place than the input (for signal-to-thermal noise) or the output (for signal-to-intermodulation noise). Gain shape, as a function of frequency, was established by circuits in the feedback path of each amplifier. The noise of the repeater was only 6 dB above thermal noise. About 20 dB of feedback was realized at the top edge of the band, and, with this, the distortion products from transistor and transformer nonlinearities were at such a low level that new test arrangements had to be designed to measure them. The

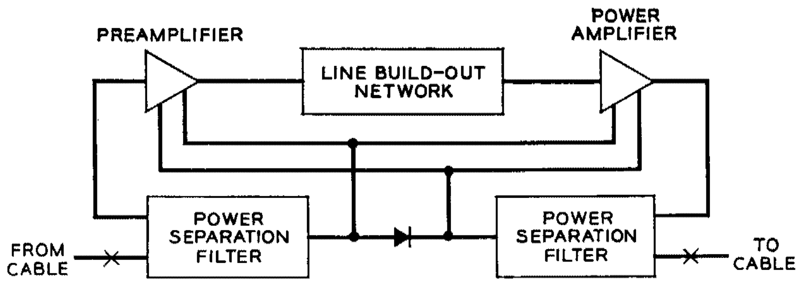


Fig. 10-11. Simplified block diagram of L4 basic repeater, showing location of line build-out network.

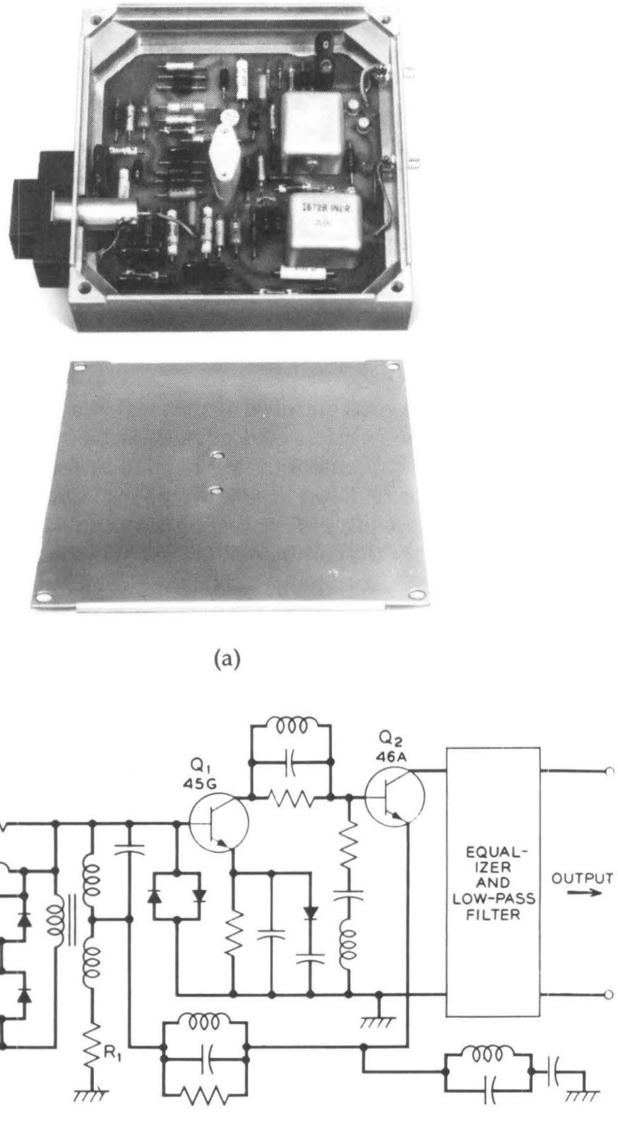


Fig. 10-12. L4 preamplifier. (a) Top view with cover removed. (b) Schematic diagram of preamplifier (less bias circuitry).

resulting projected total system noise met the long-haul objective of slightly less than 40 dBrnC0. (Telephone-band circuit noise is measured with respect to an arbitrary reference level of 10^{-12} [1 pw or −90 dBm]. To allow for the difference in perception of different frequencies within the band, it is "weighted;" the weighting used in the Bell System was the so-called C weighting. Finally, since gains and losses along a line change levels but not the signal-to-noise ratio, it was customary to refer levels to the toll test board—making it, by definition, the zero-level point. Forty dBrnC0 is, therefore, noise 40 dB above reference noise with C weighting at the zero-level point.)

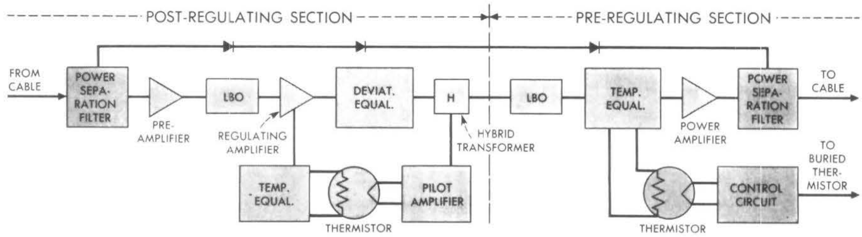
The power necessary for the repeater was taken from the center conductors of the coaxial units and was separated from the message signal at the repeaters unit by filters. Diodes regulated the voltage drop for each repeater, and additional diodes provided protection against lightning or other high-voltage surges. Line power was furnished from the main stations at about 1800 v at a constant DC of 0.5 ampere.

2.2.2 *Regulation and Equalization*

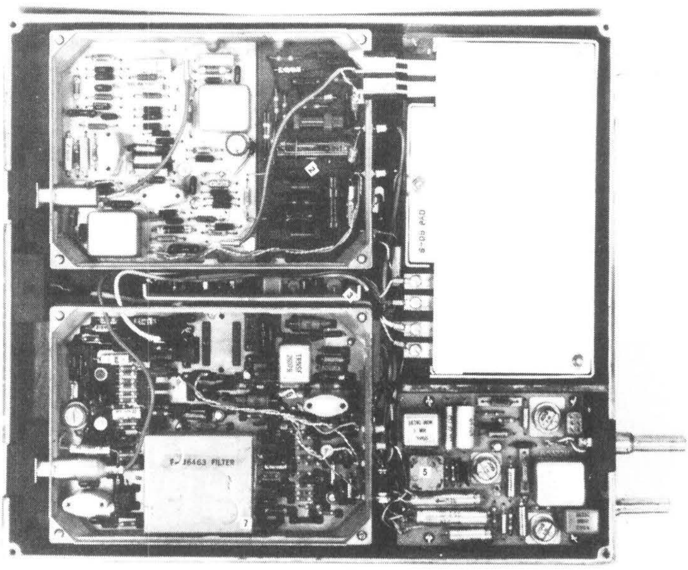
As in L3, the L4 line was controlled against temperature variations (regulated) and equalized for transmission-gain deviations by a hierarchy of fixed and adjustable equalizers.[10] Pilot-controlled regulation for temperature was necessary, but it was not required at every repeater. On a hardened route, the cable was buried at a depth of four feet in the earth, where the mean annual temperature for most of the continental United States is 55 degrees F. The deep burial reduced temperature changes, but they could still amount to as much as a 20-degree variation above and below the mean in the course of a year. For the total range at the top of the band, 12 miles of cable would change in loss by more than 12 dB. Regulating repeaters were installed at these intervals to compensate for this change in loss, in addition to providing the function of a basic repeater.

The regulating function, like the line build-out network and other equalizing networks of L4, was placed between amplifiers to minimize signal-to-noise penalties. The regulator consisted of two sections, a post- and preregulator [Fig.10-13]. The change in cable loss due to temperature varied as the square root of frequency and was very similar in effect to a lengthening (warmer temperature) or shortening (colder temperature) of the cable. The networks in each section were designed to match this shape; each network offset half the total change in a 12-mile section. The postregulator was controlled by the 11.6 MHz pilot, and compensated for half the change in the preceding section, including any error in the correction inserted by the preceding preregulator. The preregulator adjusted for loss change in the next half-section by sensing variations in ground temperature with a buried thermistor near the regulator site. By dividing the total correction in this way, the average signal levels on the line were kept closer to the optimum for the best signal-to-noise ratio, i.e., the misalignment penalty was reduced. The result of this approach to regulation was that temperature effects were offset to within ± 0.3 dB.

An additional function of the regulating repeater was to house the deviation equalizer. Predictable deviations from an absolute match of cable loss were known for the basic repeater. Although they were only a few percent of the total gain provided, the deviations were systematically additive and would



(a)



(b)

Fig. 10-13. L4 regulating repeater. (a) Simplified schematic showing the division into post- and preregulating sections. (b) The equipment package, about 14 inches long by 12 inches wide. Coaxial plug-in connectors are on the right.

have contributed significantly to system misalignment if not corrected. A number of deviation equalizers were made available, since regulating-line sections could vary from as few as two to as many as five basic repeaters.

2.2.3 *Equalizing Repeaters*

The adjustable stages of equalization were to compensate for the unpredictable effects in matching the cable loss, such as small differences in manufacturing batches, the failure of the deviation equalizer precisely to offset the known residual deviations, and changes with age. Equalizing repeaters (A

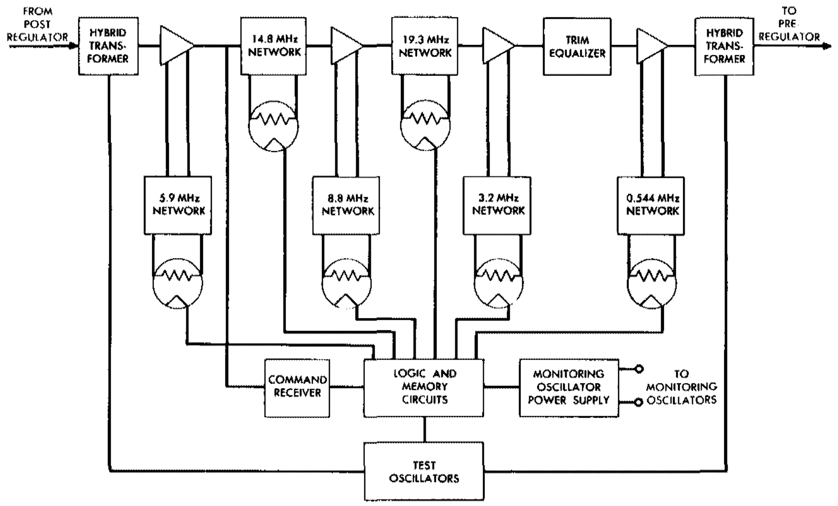


Fig. 10-14. L4 equalizing repeater. Test oscillators were activated during remote measurement and adjustment of the equalizers. Memory circuits maintained the setting of each network.

equalizers), located typically at 50-mile intervals, contained networks whose shapes could be adjusted remotely to compensate for this type of transmission deviation to a first order. The repeater provided all the functions of a basic and regulating repeater as well. Six networks, each affecting a selected portion of the message band, could be independently adjusted from a main station without interrupting service. Two of the networks were placed between amplifiers, and four of the adjustable networks were in amplifier feedback circuits [Fig. 10-14]. The network adjustments were accomplished by activating six test oscillators within a selected equalizing repeater. These sent test tones in the gaps between the signal bands to a receiver at the main station. The test signals could be inserted at either the input or output of the equalizer networks, thus allowing for the measurement of the local networks or of the following 50 miles of line. After the measurements were made, commands sent from the controlling station could adjust the setting of each equalizer to make the necessary corrections in gain or loss. Settings were then maintained by memory circuits in the repeater until new adjustments were required. Each network could be adjusted over a ±4 dB range.

2.2.4 Main-Station Repeaters

The final element in the L4 hierarchy of equalization was the main-station repeater. Unlike basic and regulating repeaters and repeaters for the first level of equalization that were located in apparatus cases in the underground enclosures, main-station repeaters were rack mounted in large underground hardened buildings. The main-station repeater included all the functions of the manhole repeaters and equalizers, plus ten additional bump shapes in a B equalizer for finer-grained equalization. The controls and operation were similar to those in the six-shape equalizing repeater. Main-station equalization was also divided into transmitting and receiving sections to minimize the misalignment penalty.

At the start of the L4 development, hopes were high that transmission would be stable enough so that only the predictable change in line loss with temperature would require continuous, pilot-controlled regulation. A pilot at 11.6 MHz was transmitted for that purpose. Unfortunately, it was discovered quite early that the temperature sensitivity of ferrite inductors and transformers in the repeaters also caused rather large changes in the low-frequency transmission. A tone at 512 kHz, derived from the Bell System frequency standard and transmitted as a multiplex-terminal synchronization signal, was used at main stations as a regulation pilot for the continuous correction of low-end transmission. Later, it was also found desirable to correct the high end continuously at main stations under control of a 20.448-MHz pilot.

Several additional functions were provided at the main stations. It was here that the high-voltage DC was applied to the line. Temperature and band-edge pilots were generated and inserted in the transmitting circuits. Automatic line-switching circuits, to transfer the message signal to a spare coaxial cable in the event of a transmission failure on the working line, were also located in the main station.

The control consoles for equalization and activation of a fault-location system were located at major main stations, such as those in the vicinity of large metropolitan areas and at the junction of coaxial routes. From a console, it was possible to address all equalizing repeaters within its area of jurisdiction and to turn on and measure the transmission test tones. Following measurement, the adjustable equalizers of equalizing repeaters could be reset from the console, if required. The fault-location scheme consisted of a series of 16 oscillators with frequencies at 4-kHz intervals above the band in the range from 18.50 to 18.56 MHz. Each frequency was unique to a repeater location. The control console provided means to turn these on in any section, address them to a specific line, and display the results as a quick check on the health of each repeater.

2.3 Terminals

Existing types of L3 multiplex equipment (channel, group, and supergroup banks) were used to build up a standard 600-channel mastergroup in the range from about 0.5 to 3.0 MHz. A new multimastergroup multiplex (MMX-2) stacked six of these in the range from 0.5 to 17.5 MHz.[11] Sufficiently wide guard bands were provided between mastergroups so that individual mastergroups could be added or extracted by using filters without the need to demultiplex and without disturbing adjacent mastergroups [Fig. 10-15].[12] The solid-state components allowed a sufficient reduction in size so that the complete mastergroup modulation equipment for three coaxial cables and a full set of spares was contained in an 11-1/2 foot double bay about 4-1/2 feet wide. This bay provided test and interconnection access to all signals entering and leaving the multiplex bay [Fig. 10-16].

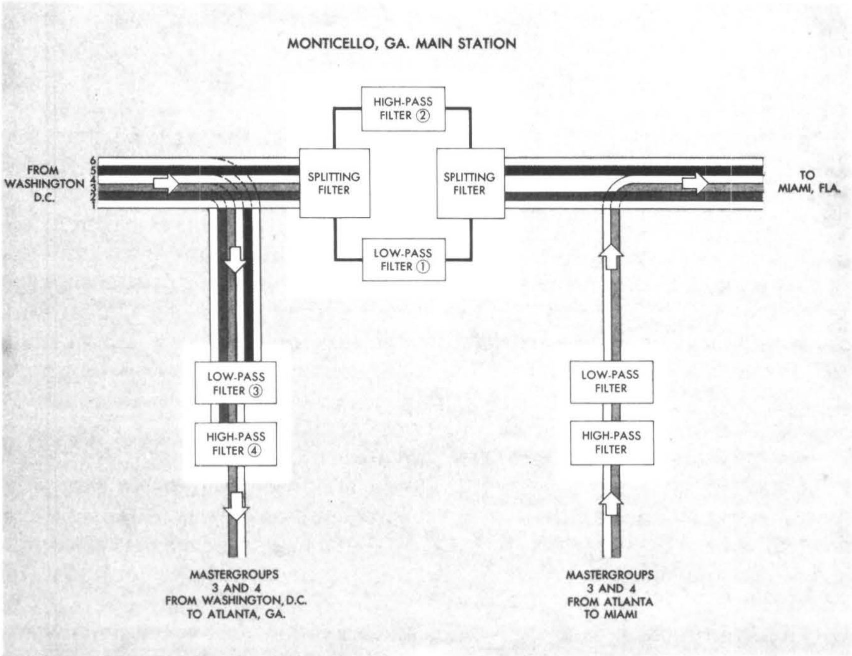


Fig. 10-15. Mastergroup dropping and adding in the L4 system.

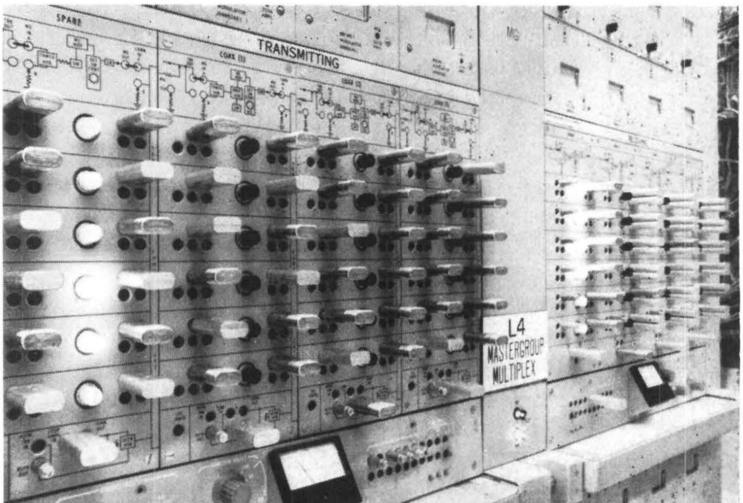


Fig. 10-16. L4 terminal multimastergroup (MMX-2). This panel provided test and interconnection access to all signals entering and leaving the multiplex bay.

2.4 Summary

Line repeaters in the L4 system had twice the bandwidth, were smaller in size, were more reliable, and consumed less power than their predecessors. They made possible a new high-capacity, long-haul transmission system, primarily as a result of the introduction of the transistor. For optimization of signal-to-thermal and intermodulation noise, pre- and postequalization were introduced at every opportunity in the system design. Although the gain element was new, the fundamental and mature concepts of negative feedback and variable equalizers were used throughout. New areas that were entered involved high-voltage DC line power, remote equalizer control, and memory elements to maintain equalizer settings.

In addition to providing 3600 channels on each coaxial tube, the L4 system was configured with the general availability of 20-tube coaxials for new cable installations in mind [Fig. 10-17]. With nine pairs for working systems and the tenth pair as a protection spare, a fully equipped L4 route had a capacity of 32,400 telephone channels.

Following a trial in 1965 and 1966 on 114 miles of line between Dayton and Rudolf, Ohio, the first commercial system was installed between Miami and Washington and was opened for service in October 1967. L4 was widely installed on hardened routes with the cable buried four feet deep and with all repeaters and main stations underground. By 1975, there were over 86 million

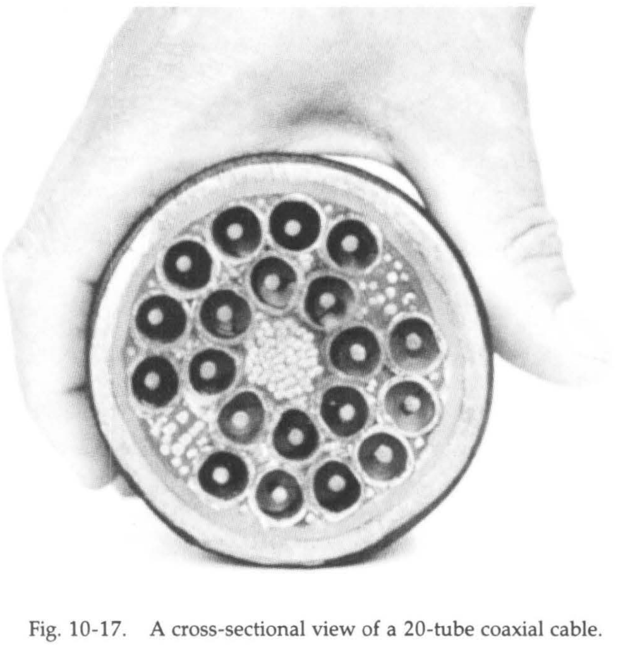


Fig. 10-17. A cross-sectional view of a 20-tube coaxial cable.

voice-circuit miles installed, about twice as much as the maximum combined L1 and L3 capacities.*

III. THE L5 SYSTEM

3.1 The System Plan

In the years after World War II, the development cycle for major systems from first studies and laboratory experiments to commercial service was often five years or longer. But the underlying technology was moving rapidly, making larger and better systems possible at more frequent intervals. As a consequence, a pattern of overlapping developments became the rule rather than the exception in every area of transmission work. In almost every case, work on the next generation was well under way by the time the current design was put in service. Even by the norms of this pattern, however, the transition from the L4 to the L5 coaxial system was unusual. The L4 development was relatively short, requiring only four years from start to service, but exploratory work on L5 was already under way at the Merrimack Valley transmission laboratory in 1965, only two years after the start of L4.

This was possible because of the extraordinary rate of advance in transistor technology. Where L4 settled for a band of 17 to 20 MHz, it was possible by 1965 to contemplate a 60-MHz system with repeaters at one-mile spacings. In fact, the device capabilities were such that, in the course of L5 development, even broader bands and closer spacings were considered, either as options for L5 or as the basis for a follow-up development. However, after some study and initial experiments, it was decided that a 60-MHz system with 2-1/2 to 3 times the capacity of L4 would be a sufficiently challenging step, with enough new problems to keep everyone busy. A design for that bandwidth and one-mile spacings became established as a reasonable target.[13,14] Others, too, were working on solid-state broadband coaxial systems by the mid-1960s. Both L. M. Ericsson in Sweden and the Siemens Company in Germany were developing, and eventually produced, 60-MHz systems at about the same time as L5.[15,16]

Allowing for the broader band and closer spacing, L5 was very much a

---

* From time to time, it has seemed appropriate to mention some of the unplanned occurrences in development. Lest it be thought that coaxial developments were without untoward incidents, a condition on the first L4 route is worth noting. If the main (11.6-MHz) pilot was inadvertently removed, the regulating repeaters rapidly compensated by going to maximum gain, saturating the line with noise and initiating a switch to the spare line. Initially, however, the switch was not quite fast enough, and the next section was also saturated, initiating a switch, and the next and the next in a chain reaction. At the line terminal (Miami or Washington), enough noise was reflected at the circuit terminations to start the process in the reverse direction until some 2500 miles of lines had switched over. To lose the spare was bad enough, but the maximum gain also corresponded to hot thermistors, and the cooling cycle was much slower than the heating. Finally, recovery of any regulator could not start until the preceding regulators had cooled to within normal range. The spare was lost, before anyone could react, and could take hours to recover. It took some circuit changes, a modification of the switch, and the passage of a few years before anyone found the calamitous sequence amusing.

second-generation system and bore many close family resemblances to L4. The total band transmitted was made substantially greater than the signal band to simplify the equalization problem. The top frequency of the nominal 60-MHz system was thus close to 70 MHz. The basic repeater was made as simple as possible, fixed in gain, and designed to offset one mile of cable loss at the mid-temperature range as precisely as possible. Periodic fixed deviation-equalizers were used to correct for the known average difference between repeater gain and cable loss, and, at wider spacings, infrequently changed adjustable equalizers were inserted to offset the unpredictable differences. Dynamic pilot-controlled regulators and equalizers were used only where changes were expected to vary with time, such as the inevitable temperature effects on cable loss and repeater gain. (Temperature effects, in addition to the cable-loss change, were more expected and more cheerfully accepted in L5 than in L4.) Post- and pre-regulation and equalization were used to reduce misalignment, keeping the signal levels as close as possible to the optimum between excess intermodulation noise and excess thermal noise. DC power was fed from main stations over the coaxial center conductors. Protection switching was provided every second power-feed span.

As initially envisioned, the L5 system would have had a channel capacity of 9000 channels in fifteen 600-channel mastergroups, plus 1800 channels in three additional mastergroups, operating with less than standard signal-to-noise performance. The latter three mastergroups were to be used in emergency situations only, for service restoration when other systems failed. However, early in the development process, a noise-reduction technique was devised whereby all 18 mastergroups were brought up to the quality required for standard service. The system then provided 10,800 two-way toll-grade message channels per coaxial pair, a capacity three times that of its L4 predecessor.

Many factors, besides the obvious bandwidth requirement, influenced the specification of the frequency allocation of the L5 system [Fig. 10-18]. Experience with earlier systems indicated that the bandwidth should span something less than five octaves in frequency to avoid troublesome parasitic effects in circuit elements. It was extremely difficult, for example, to design transformers for good very-low-frequency characteristics that would perform satisfactorily at the top frequencies. Thus, the first mastergroup began at about 3.0 MHz in L5, as compared to about 0.5 MHz in L4, even though the shift required more repeater gain. This meant a ratio of 20 to 1 from top to bottom frequencies in L5, compared to 35 to 1 in L4 and 27 to 1 in L3. It is interesting to note that, by this decision, the lowest L5 signal frequency was higher than the highest frequency of the first coaxial system achieved with so much effort 25 years earlier. The frequency allocation was also influenced by the need to transmit and recover certain nonpayload signals. Among these were four pilot signals used to control the automatically adjusted regulators and equalizers, fault-location signals used to control a line-protection switching system, and a reference frequency signal used to synchronize terminal equipment to a frequency standard.

Compatibility with the existing frequency-division-multiplex hierarchy was achieved by adopting the six-mastergroup arrangement of the L4 system, now called a jumbogroup, as the basic payload package for the L5 system. Basic

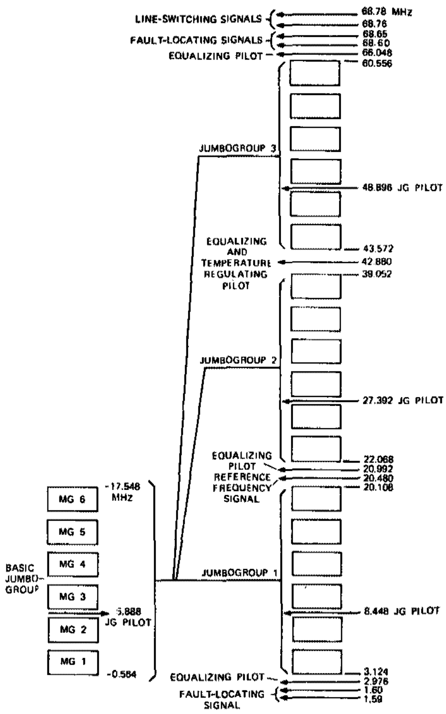


Fig. 10-18. Line-frequency spectrum of 18-mastergroup, 10,800-channel L5 system.

jumbogroups were formed by multiplex equipment external to the L5 system. The L5 system provided all the arrangements, including jumbogroup multiplexing, for the frequency translations and transmission of three jumbogroups. The line frequencies allocated to the jumbogroups were selected in such a way that any of the three jumbogroups, or any combination of two, could be manipulated at line frequencies for branching and dropping operations without the use of multiplexing equipment. Both operations required the use of blocking filters that, in turn, required the provision of guard bands between jumbogroups.

The transmission objectives adopted were essentially the same as those of its L4 predecessor. In addition to the usual telephone-channel signal-to-noise objective of 40 dBrnC0 for 4000 miles, the gain of the broadband line was not to deviate by more than 4 dB from a defined nominal loss at any frequency in the telephone signal band. (Individual channels could be adjusted to much closer limits in the multiplex.) There were other objectives, including some with respect to system reliability, signal administration, maintenance, multiplex compatibility, and the ever-pursued objective of the most for the least cost.

Nominal signal levels were designed, both to maximize signal-to-noise ratios and produce relatively uniform signal-to-noise ratios in each message channel. For L5, as for most coaxial systems, this objective required different signal levels as a function of line frequency. Ideally, it would have required a continuous change in level with frequency, but this would have unduly complicated the terminals. The ideal curve was approximated by a small adjustment in the level of each jumbogroup at the jumbogroup-multiplex output. Signal misalignments from the optimum were expected from a variety of identifiable causes, and elements within the equalization hierarchy were designed to cope with each cause. For reasonable magnitudes of misalignment (less than about ±5 dB), equal pre- and postequalization produced the least signal-to-noise penalty, and, to the extent practical, this strategy was adopted for L5 as it had been in L4.

Schematically, the L5 system was very similar in layout to L4 [Fig. 10-19]. Since the magnitude of any particular cause-associated misalignment tended to be proportional to distance, the various levels of the repeater and equalization hierarchy were designed for deployment on the basis of system length. Simple, fixed-gain, precisely manufactured, basic repeaters were installed at one-mile intervals. After a maximum of seven basic repeaters, a pilot-controlled regulating repeater was inserted to offset loss changes due to temperature changes in the cable. This pattern was iterated up to a maximum of 37.5 miles, where an equalizing repeater was used, incorporating the functions of the basic and regulating repeater and, in addition, the first stage in a sequence of manual and automatic variable equalizers to offset the cumulative gain deviations across the band. At the 75-mile-spaced power-feed main stations, more complex repeaters furnished all the previously mentioned capabilities, plus a finer-grained, manually adjustable equalizer with more gain-frequency shapes (bumps) and

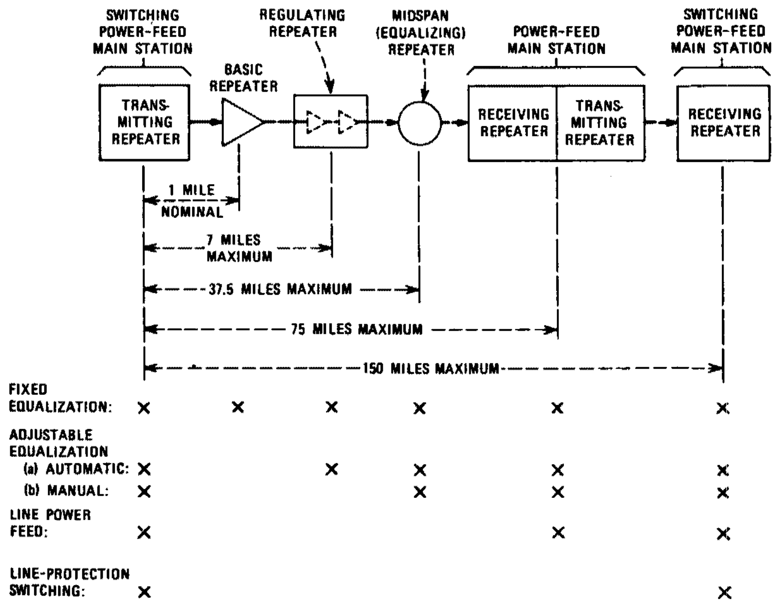


Fig. 10-19. Main features of an L5 switching section. The crosses indicate the location at which the functions listed were found.

a pilot-controlled automatic equalizer. The latter was to offset the time-varying deviations, mostly due to temperature effects on the repeater, and to correct for the failure of the main temperature regulators to offset cable temperature-loss changes perfectly. The basic, regulating, and equalizing repeaters were designed to be housed in manholes. The main station repeaters were designed for an attended building environment, since these locations were used to supply power to the line.

All repeaters were powered by DC along the coaxial center conductor. The 75-mile maximum spacing between powering stations was selected to be one-half of that used for the L4 system. This selection provided a compatible spacing for L4 to L5 conversions, and, at the same time, did not require excessive powering voltage. Experience with the DC powering arrangement used with the L4 system had indicated that voltages greater than 1800 would be likely to generate excessive impulse noise through corona-like discharges in cable, connectors, and repeater equipment. The 1150-v maximum adopted for L5 provided a comfortable margin in this respect.

A new protection switching system was designed with the capacity to serve as many as ten working coaxial lines, in addition to a spare line, to accommodate the largest of the coaxial cables (22 tubes). The design provided for effective operation with only a few coaxials and the addition of working lines as needed, with a minimum of cost and effort. It also provided a robust signaling arrangement to avoid improper switching in response to noise and short interruptions in transmission. (It was also designed with the early L4 experience fresh in mind.) As a result of reliability studies encompassing both cable and repeaters, line-protection switching stations were placed at intervals not exceeding 150 miles. With this spacing, service outage time due to single line failures was projected to be significantly less than that expected as a result of major failures due to such things as cable breakage during road construction.

3.2 Repeatered Line

The basic building module of the L5 repeatered line was the switching span. Such sections were connected in tandem to span the distance between terminal stations. Multiplex, branching, and dropping operations were restricted to protection switching stations. To improve overall transmission and noise performances for very long systems, the mastergroup frequency-frogging concept, used in earlier coaxial systems, was also adopted for L5. In a mastergroup frogging operation, the line signal was demultiplexed to the mastergroup level and then remultiplexed in a pattern that placed each mastergroup in a different position in the line spectrum. The net effect was to reduce the accumulation of certain types of intermodulation noise and produce a more nearly uniform transmission quality for each of the message channels. Where long routes were required, frogging stations were to be implemented at about 800-mile intervals.

The wider band and corresponding larger number of channels, along with the doubling of the number of repeaters, placed an extremely difficult requirement on repeater linearity. The linearity (reduction in intermodulation) needed was 20 dB better than L4, where the problem had been far from easy. The system improvement was found partly in a new power transistor, much improved in linearity. This by itself, however, would not have been adequate for the required performance. A major improvement was made by introducing phase-shifting networks at each regulating repeater to break up the systematic in-phase addition of third-order modulation products that had plagued carrier systems for decades. The modest phase distortion introduced by the networks provided a reduction of about 15 dB in third-order intermodulation noise for a 4000-mile system, and resulted in a system limited essentially by second-order intermodulation and thermal noise.

It might well be asked at this point why, if the solution was so simple, was it not used sooner? The answer is that in earlier systems other means were adequate to control the distortion, but also, apparently, because no one thought about it hard enough. It must be remembered that phase or delay distortion was considered bad for video and data transmission. In L5, however, analog video was not a factor. Some work was done on multilevel data transmission in L mastergroups, but there was little application.[17] The distortion introduced over the bands likely to be used for data was small and could be equalized at the terminals, if necessary. It took just the right amount of poison to cure the patient without killing him.

3.2.1 *The Equalization Plan*

The high frequencies of L5 revealed some equalization problems that had been of no consequence in the earlier systems. The different materials in cables of different vintages cause deviations from the nominal loss that became significant at 60 MHz. An early version of the polyethylene support disks, for example, caused a dissipation loss (the shunt loss) about twice that of the later product. Fortunately, the cable-loss differences as a function of frequency were quite broad and were easily equalized. Within the L5 band, the cable had no sharp or narrow-band transmission aberrations. The current cable-splicing methods, as well as those used for most previous coaxial systems, produced only small deviations that could be neglected. However, splicing techniques used in very early cables introduced relatively severe impedance discontinuities. The resulting transmission deviations varied sharply with frequency and were not closely reproduced from sample to sample. In equalization parlance, they were ill-behaved, and it was not feasible to provide for their equalization. Conversion of these cables for L5 required resplicing using later techniques.

The basic strategy of L5 transmission equalization was essentially unchanged from L4 [Fig. 10-20]. (The depiction, if not the execution, of equalization plans had become simpler since the days of the L3 coaxial system.) Briefly stated, this was to offset predictable deviations that would not change over time with fixed equalizers and provide adjustable equalizers only for deviations whose characteristics could not be predicted or that were expected to vary with time. It was further expected that the adjustable equalizers would fall into one group that would not be changed at all, after initial adjustment, or only be changed at infrequent intervals (the static group) and a second group that would need essentially constant adjustment. Only the latter were to be automatic, i.e., pilot-controlled dynamic equalizers. The plan was carried out by a hierarchy of repeater-equalizers to carry out the successive stages of equalization, spaced

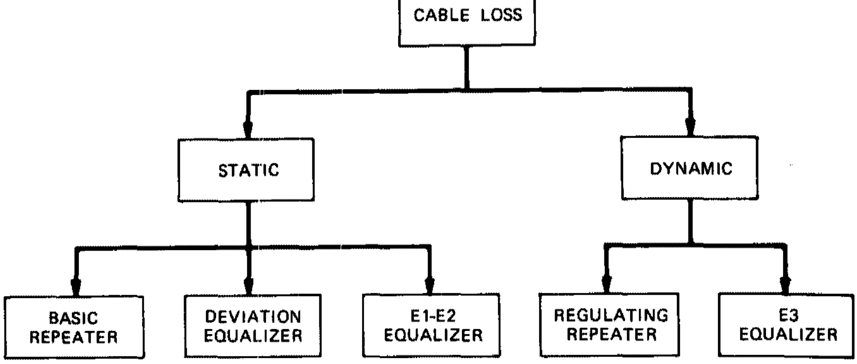


Fig. 10-20. L5 system equalization strategy.

as far apart as the signal-to-noise penalty of signal misalignment would permit. To minimize misalignment, pre- and postequalization were used with only one exception. Generally, each member of the hierarchy provided all the equalization functions included in lower-order members along with additional functions. A feature of the plan was to exercise extremely tight control over transmission in design and manufacture, leaving as little additional adjustment as possible to mop-up stages of equalization.

The basic repeater, the lowest-order member of the hierarchy, was designed to equalize the nominal cable loss at mid-temperature range. It was a fixed-gain design, with six different codes to provide the gain necessary for cable lengths ranging from one-half to one mile, in one-tenth of a mile increments. The plug-in line build-out arrangement as used in L4 could not be built to the precision required in the L5 basic repeater.

The regulating repeater, inserted after seven basic-repeater spans, compensated for variation in cable loss with temperature. Schematically, this repeater was essentially identical to the L4 regulating repeater, with a pilot-controlled (42.880 MHz) postregulator and thermistor-controlled preregulator. While the closed-loop arrangement provided a very precise adjustment, any deviation in the pilot level resulted in an equalizer adjustment. If the pilot changed from its reference level for any reason other than temperature change, the transmission was changed across the entire band as if a line-temperature change had actually occurred. This focused special attention on transmission deviations at the temperature-pilot frequency. The basic repeater was designed not only to have a very precise gain characteristic across the band but also to have a near-zero gain-temperature coefficient at the pilot frequency as well. All repeater spacing was engineered on cable loss at this frequency.

3.2.2 *The Equalizing Repeater*

At about 37-mile intervals (five regulating spans), an equalizing repeater added two additional forms of equalization. In common with past practice, a fixed-deviation equalizer was needed to compensate for the difference between the average loss of the cable and the average gain of the repeaters. The precise design of the basic repeaters permitted this equalizer to be spaced as far apart as the 37-mile intervals without excessive accumulation of signal-level misalignment. A single design offset the average deviation leaving a residue that was relatively easy to correct in the higher-order adjustable equalizers [Fig. 10-21]. The second equalizer in the equalizing repeater was an adjustable, multishape network, referred to as the E1 equalizer, for finer-grained correction of the design error. This was accomplished by a series of ten "bump" shapes, each affecting a different portion of the transmission band. Corrections up to ±4.5 dB could be made in each bump. The E1 equalizer also included a narrow shape at the regulating pilot to permit precise level adjustment after all other adjustments had been made. As in the L4 equalizing repeater, the equalizing networks were between amplifiers or in feedback networks [Fig. 10-22].

The E1 equalizer was usually located in a manhole in a sealed and pressurized apparatus case similar to those used for the basic and regulating repeaters. The next stage of equalization was in a main-station repeater, located in the office environment of a power-feed main station. Both the transmitting and receiving sections of a main-station repeater included an E1 equalizer. In addition, both sections were equipped with a similar, but more elaborate, E2 equalizer. The E1 and E2 equalizers were designed in concert, the E2 providing 18 additional bump shapes intended to complement and offset the transmission-level deviations left by the E1 equalizer. The initial selection of the equalizer shapes was

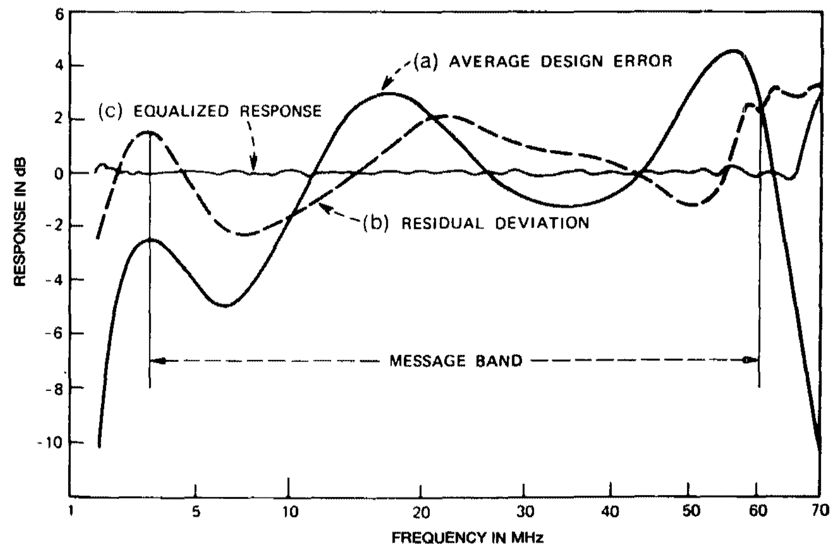


Fig. 10-21. Design-error correction in the L5 system. The average transmission error was reduced at 37-mile intervals by a fixed design-deviation equalizer leaving an easily equalized residue.

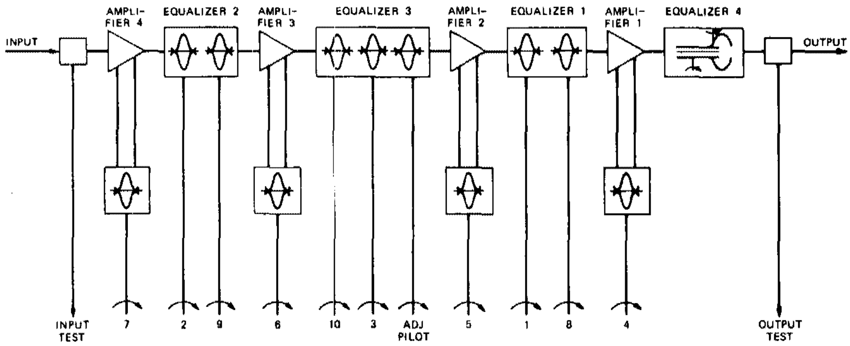


Fig. 10-22. The L5 system E1 equalizer transmission circuitry. Ten adjustable "bump" shapes and an adjustment at the 42.88-MHz pilot supplemented a fixed design-error equalizer at 37-mile intervals.

based on computer simulations of the expected L5 transmission channel. A field trial of the system provided additional essential information that permitted improvements in the initial designs.

Bump shapes were chosen for L5 over other options because the computer simulations showed that they equalized the line better and were easier to adjust. Experience with the L4 system had shown that many of the deviations for which the adjustable equalizers were intended were, indeed, not variable with time. As a result, the L5 static equalizers were adjusted manually at the site and on an out-of-service basis. This procedure was adopted because it was expected that, once set, the equalizers would need to be readjusted only rarely, if at all.

3.2.3 *Time-Varying Deviations*

The largest of the time-varying transmission deviations, the variation of cable loss with temperature, was offset to a very large extent by the regulating repeater. This compensation, however, was not perfect, and a small, but significant, error accumulated over a 75-mile power-feed span. Over the same line span, a somewhat larger time-varying transmission deviation was expected as a result of temperature changes of the line repeaters. While some smaller deviations were expected from other sources, these two represented the principal burden on the final stage of L5 equalization, which was implemented in a network referred to as the E3 equalizer.

The time-varying transmission deviations offset by the E3 equalizer tended to be rather well-behaved broadband deviations and therefore amenable to correction by a few carefully chosen adjustable shapes. On the other hand, they were certain to change continuously with seasonal variations in temperature; equalizer adjustment had to be automatic to avoid frequent manual operations. These considerations led to the adoption of a pilot-driven equalizer with relatively simple transmission networks. Four pilots, including the temperature pilot, placed in frequency above and below the three jumbogroups provided a suitable measure of the time-varying deviations. The relatively small magnitude of these deviations permitted the E3 equalizers to be used in the receiving sections of main-station repeaters for post-equalization only.

3.3 Centralized and Automated Transmission Surveillance

As the long-haul coaxial network evolved to increasingly complex and higher-capacity systems, with more and more electronic equipment at unattended locations, the need for improved maintenance and monitoring capability became acute. At the start of development of the L5 coaxial system, inexpensive minicomputers and remotely controlled transmission-measuring equipment were becoming available. A centralized and automated transmission-surveillance system was included in development plans from the start as an integral part of the L5 network.

The system consisted of a transmission-surveillance center (TSC) located at a principal main station and transmission-surveillance auxiliaries (TSAs) located at associated main stations. The TSC originated all control operations and processed all measured data. The TSAs performed measurement functions as directed and returned the raw data to the controlling center. At each TSA, a digitally controlled oscillator could be connected to a selected line through a switch matrix and requested to apply a sine-wave signal of specified level and frequency within or outside the band [Fig. 10-23]. Similarly, tunable selective

Fig. 10-23. Remotely controlled programmable Type 90 oscillator and detector for the L5 transmission-surveillance auxiliary.

detectors could be connected and controlled to measure received signals. An analog-to-digital converter in the receiver control unit converted the measurements to a digital form suitable for transmission over a data link to the surveillance center. The data link was a four-wire circuit, usually in the same sheath with the coaxials and time-shared with an alarm-reporting system.

The centralized approach provided an area transmission performance overview not possible with local link-by-link maintenance. Since surveillance was essentially continuous, it also provided early warning of slowly developing troubles. Programming and running the tests under computer control saved effort and provided accurate and flexible records. Trouble localization in L5 used test oscillators at each repeater in a manner similar to that used in L4. These oscillators were also under programmed control from the test center, making rapid trouble localization possible.

A TSC was generally located at an attended terminal main station assigned the central role in the overall maintenance and operation of an L5 area. It functioned as a nerve center for originating and processing automatic measurements on line and multiplex equipment. Software routines were written and stored on tape cassettes to minimize operator interaction and to process and present results in an easy-to-use format. The system quickly proved its usefulness, and, in 1976, the system was upgraded to a disk-based system with much larger computer memory and a more sophisticated software system. Application programs were written to permit quick and easy implementation of frequently needed maintenance activities. Routines included measurements of pilots, line transmission, jumbogroup multiplex outputs, and activation of the fault-location sequence.

Data bases were generated, covering the arrangement and extent of equipment to be controlled. This information included manhole designations, route numbers, lines equipped, the allowable range of measurements, and test points available. A separate set of programs was dedicated to checking the surveillance system itself for proper operation. The diagnostic program provided an indication of the health of the entire surveillance system by exercising all the access and measuring equipment at the TSC and TSAs. The transmission-surveillance system provided the means for maintaining the L5 system on a network basis, rather than as a series of independent stations.

3.4 Trial and First Service

L5 equipment was installed in late 1970 for a trial on a 100-mile section of coaxial cable between Netcong and Cedarbrook, New Jersey (the metropolitan junction stations for New York and Philadelphia on the main north-south East Coast coaxial route). The trial was successful in every respect, and construction was started on a 620-mile route for commercial service between Lillyville, Pennsylvania and Hillsboro, Missouri, the junction stations for Pittsburgh and St. Louis [Fig. 10-24]. (Planning work had already been under way for some time.) With its side legs, this installation consisted of 815 sheath miles of cable, 14 main stations, 850 manholes, 3400 manhole repeaters, and over 250 bays of transmission equipment. Xenia, Ohio, a major junction station, was established as the control center housing the transmission-surveillance equipment. Installation and testing went forward without extraordinary incident. (A tornado

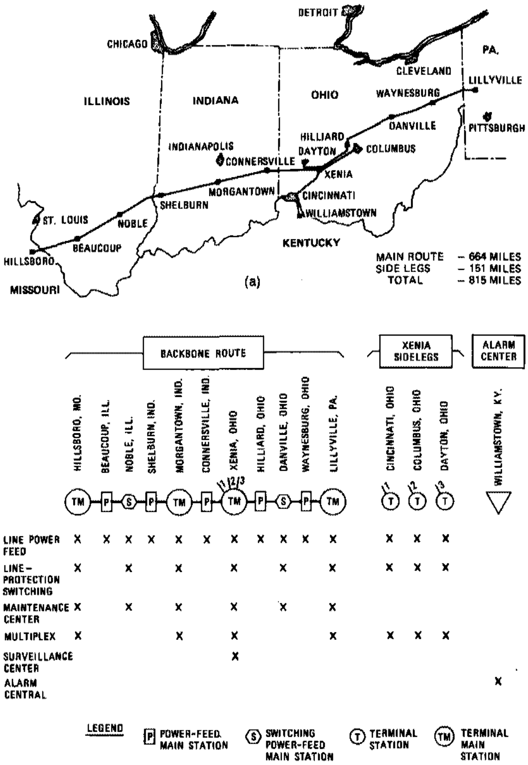


Fig. 10-24. The first commercial L5 route. (a) The route from Lillyville, Pennsylvania to Hillsboro, Missouri. (b) Administrative layout of Lillyville–Hillsboro route. The crosses indicate functions provided at each location.

demolished a substantial part of Xenia, but the underground L5 buildings were unscathed.) The system went into service on January 3, 1974, fulfilling a schedule developed six years earlier, only a year after first service on L4.

3.5 Physical-Design Evolution

In the story of the development of electronic communications, an aspect often neglected is the problems faced and overcome by equipment physical designers. A significant part of the effort required went into this specialized development field, and no small part of the credit for the final achievement belongs to this group of skillful specialists. They were often the sobering influence that kept euphoric circuit designers in touch with reality. They made vital contributions in all systems. The evolution of coaxial-system design will serve to illustrate their role.

All electronic apparatus must be assembled, tested, and maintained. For a laboratory model, baling wire and any kind of skyhook will do; but, to manufacture in quantity, install, and operate in the field, a great deal of discipline and a careful choice of materials and design are required. In the preelectronic era, these necessities led to the practice of assembling equipment on flat plates suitable for mounting on relay racks (the term is indicative of their origin in DC electromechanical switching). Interconnection was by multiwire, formed cables laid out along orderly horizontal and vertical paths. The layouts of the first experimental coaxial repeaters in the 1930s showed the respect paid to this tradition. This approach, however, had little regard to the length of feedback paths and other electrical factors that became critical at the higher frequencies of the new equipment.

3.5.1 L3 Physical Design

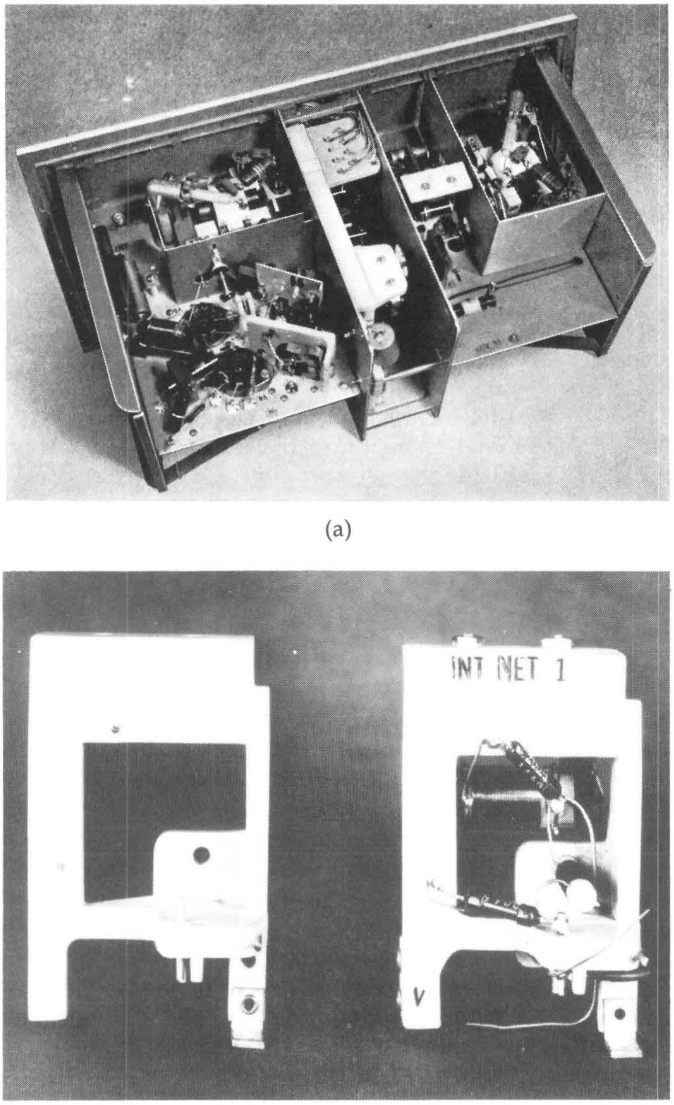
Developers quickly learned the lessons, however, and succeeding amplifier models were made increasingly compact and with less concern for two-dimensional plate layouts. In the L3 repeater, component placement, parasitic couplings, and the inductance of even short leads became important. Circuit elements were placed in the best possible position for optimum electrical performance, and supporting structures were then designed to maintain the desired spatial relationships. These supporting structures were separate units mounted on an amplifier chassis so that the networks could be individually tested before final assembly and removed for repair or replacement if necessary [Fig. 10-25]. This method of design became feasible because of the new availability of cold casting resins, with parts produced in a cheap phenolic mold. The process was practically equivalent to the sand casting of metals, and complex parts were economically manufactured, even in the relatively small quantities required for L3 production.

The L3 amplifiers were housed in a sealed die-cast container to keep out moisture, with a desiccant enclosed in each housing. In this sealed environment, careful attention was paid to conducting the considerable heat generated to the outside world without an excessive temperature rise of the vacuum tubes. Even so, high tube temperature was identified as one of the causes of relatively short L3 tube life. (By ordinary standards, the tubes were still of very long life, but transmission-system designers labored under the "tyranny of numbers." There were over 4,000 vacuum tubes in a transcontinental L3 line. If the average tube life was 100,000 hours—over 10 years—there would be one failure per day on the average.)

In the first experimental line between New York and Philadelphia, there had been some experimenting with repeaters in manholes, but it proved more practical in commercial L1 and L3 to provide aboveground huts for housings. Only some of the later L3 repeaters were placed underground in specially constructed facilities on hardened routes.

3.5.2 L4 Physical Design

The wider band and higher precision needed in L4 required a new approach to physical design. The small size of transistors and other components allowed designers to abandon the three-dimensional L3 approach. A solution was found in the planar world of printed-circuit construction. Copper paths were printed



(a)

(b)

Fig. 10-25. L3 line-amplifier assembly. (a) The wiring side of a complete amplifier. (b) An input interstage, illustrating use of cast resin frameworks.

or etched on the surface of epoxy-glass boards. Printed wiring had the desirable high-frequency properties and provided tight control of component placement. The board assembly was fastened inside a heavy die-cast aluminum box.

In the L4 repeater, the designers had a new problem to solve because of the series end-to-end DC power feed. Only a small series voltage drop was needed for the operation of each repeater, but the entire repeater circuit was at the voltage of the coaxial center conductor. For repeaters close to the power feeds, there could be 1000 v or more between the circuit and the grounded mounting frame. It was therefore necessary to insulate the entire amplifier electrically from the grounded outer case, but still to provide a low-resistance thermal path to dissipate the heat generated. The solution was to bond the off-ground circuit chassis, over a large area, to an epoxy coating on the grounded frame. Such construction ensured that most of the heat was conducted to the repeater frame and eventually to the outside of the apparatus case. In the amplifier sections, thermally critical transistors were installed within heavy-aluminum extrusions mounted on the printed-wiring board and bolted to the aluminum covers. Materials, cross sections, thermal impedances, and the conditions of the environment were analyzed and measured to assure that the transistors were within the temperature range required for long life [Fig. 10-26].[18]

Since L4 was to be installed on hardened routes, the repeater housings were designed for a manhole environment from the outset. A newly developed and less expensive prefabricated manhole was used at the smaller basic and regulating-repeater sites [Fig. 10-27]. Within the manholes, several unsealed amplifier units were mounted in gas-tight cases maintained at cable gas pressure

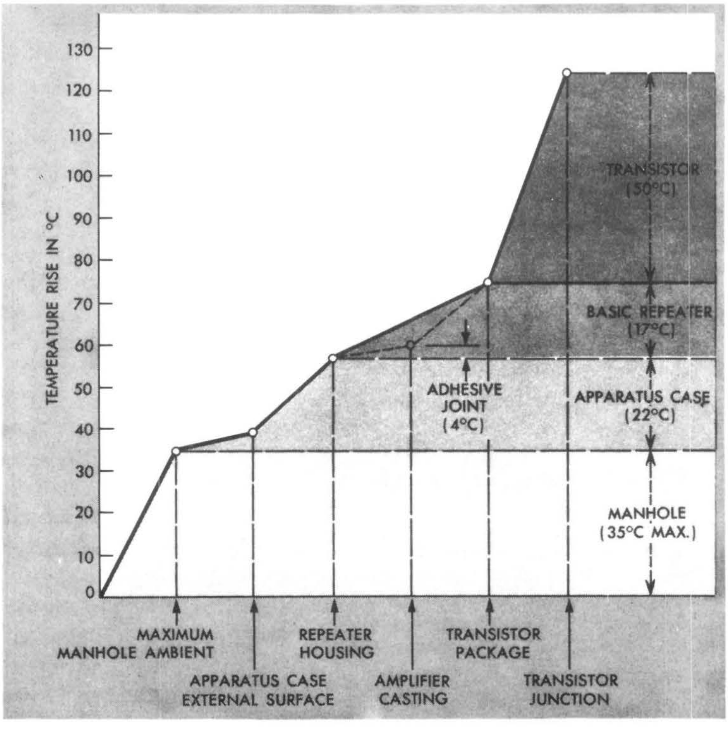


Fig. 10-26. L4 repeater temperature gradients.

Fig. 10-27. A precast-concrete manhole of the type used to house L4 repeaters.

[Fig. 10-28]. The gas pressure prevented moisture from entering through any small defect that might develop in the seal, and it warned of major troubles through low-pressure alarms. The watertight housings further emphasized the need for small size and compact packaging. All of the L4 line equipment not in main stations was designed for mounting in the cylindrical gas-tight cases. Repeaters plugged directly into jacks mounted in the frame of the case. The manholes were often wet and dirty. They presented a poor environment for any kind of work on delicate electronic equipment. These considerations led to a design requiring that all components needing maintenance had to be plug-in elements, so that maintenance work could be carried out under more favorable conditions. Extreme precautions were taken to protect maintenance personnel from inadvertent contact with any high-voltage points.

3.5.3 L5 Physical Design

As physical designers considered the problems imposed by the L5 coaxial system, several different approaches were considered. The high cost of outside plant and the close repeater spacing led to the consideration of such options as small underground lockers, still smaller "handholes," or even the direct burial of repeaters. (The cost of digging and the construction of a concrete

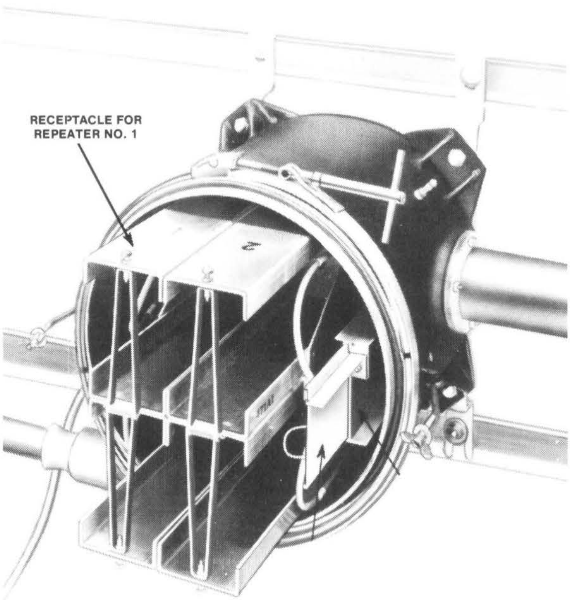


Fig. 10-28. L4 basic repeater pressurized apparatus case with cover removed, mounted on a manhole wall.

manhole considerably exceeded the cost of a full complement of electronic equipment housed in it.) However, the need for access, so that the coaxial units could be equipped only as needed, and the benefits of an easy conversion from L4 to L5, finally determined a plant approach similar to L4. Designers further benefited from not having to characterize a wholly new environment.[19]

Ideally, a conversion from L4 to L5 would have involved only the exchange of plug-in units. However, the need for a more efficient thermal dissipation path (L5 repeaters dissipated one-third more heat then L4) and more surveillance and fault-location leads required a more extensive change. The overall layout and pressurized cylindrical apparatus cases were retained, but the internal framework for mounting the repeaters was replaced. With the new frame, the apparatus case housed four basic repeaters and the associated fault-location circuits [Fig. 10-29]. The repeater was designed so that, when bolted in place, its leading surface was in intimate contact with the apparatus-case frame. Coaxial patch cords were used for insertion of fault-location oscillator tones at both the input and output of the repeaters via twin coaxial jacks located on the apparatus-case chassis. The twin jacks served also as repeater bypass points in the high-voltage, series-powered line. They permitted a power patch around a repeater, without turning power down on a line, in the event repeater replacement became necessary. The apparatus cases were filled with dry nitrogen to 9 psi. Six apparatus cases were required in a basic-repeater manhole to equip

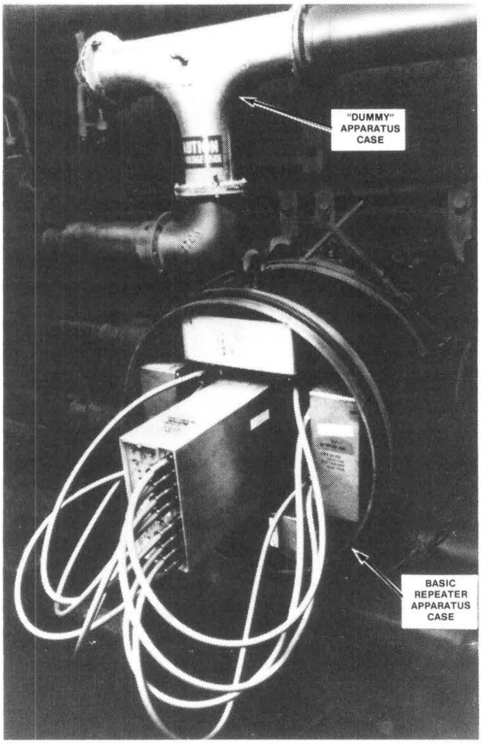


Fig. 10-29. L5 basic repeater apparatus case. The repeaters were located around the periphery of the case. The center box housed fault-location and test access circuits.

a 22-tube coaxial cable. Adapters or dummy apparatus cases were installed to permit an entire manhole to be racked and spliced initially, and apparatus cases were installed only to cover the number of coaxial tubes actually equipped in order to defer expenditures until additional capacity was required.

The L5 amplifier consisted of a preamplifier, a power amplifier, a power-separation network, and a line build-out network, all perhaps 1000 v or more off ground. These were assembled on a chassis located within, but electrically isolated from, a rugged die-cast-aluminum outer housing [Fig. 10-30]. The electrical insulation was provided by lining the outer case with an epoxy coating. As in L4, thermal conductivity was provided by a large-area bond between the off-ground and the grounded chassis. The repeater covers were secured using one-way screws as a safety feature. There were no adjustments to be made and hence no reason to open a repeater in the field. They were effectively sealed in manufacture to avoid the exposure to high voltage that would occur if a coverless repeater were accidentally inserted in a powered-apparatus case.

It was established very early in the L5 development that conventional com-

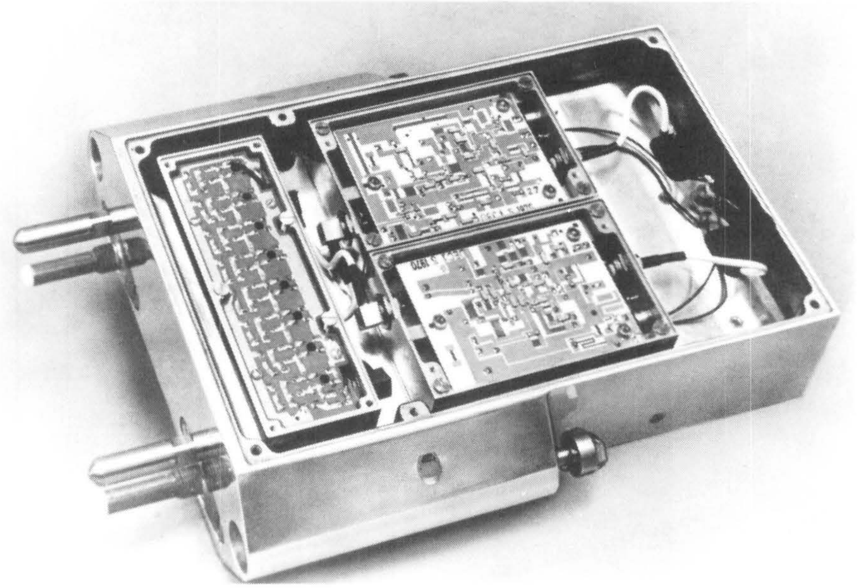


Fig. 10-30. L5 basic repeater. The inner circuits could be at a high DC voltage and were insulated from the grounded outer frame.

ponents mounted on printed-wiring boards could not be used. A recently developed thin-film technology, in which tantalum films on ceramic plates were used for conducting paths, precision resistors, and even low-value capacitors, was coming into general use. (See another volume in this series, *Electronics Technology (1925-1975)*, Chapter 9.) This promised to provide the precision required for L5. But even conventional thin-film techniques were not adequate. L5 required exceptionally large ceramic substrates and a considerable number of discrete components, such as transformers and larger-value capacitors, to mount on the plates.

Heat dissipation, as always, was a problem. The layouts were designed so that the power-dissipating resistors were on the periphery of the ceramic substrate. The transistors, however, had to be centrally located because of circuit requirements for short conductor lengths. Special heat sinks were designed to conduct the transistor heat to the aluminum covers and away from precision resistors and other temperature-sensitive components. The requirements to provide a metallic heat conduction path of large cross section but, at the same time, have enough flexibility to avoid stressing the delicate ceramic plate were at first glance inconsistent. The structural support developed was to mount the substrate directly on the transistor heat sink and merely stabilize its outer periphery on three bosses of the aluminum housing. The L5 heat sink was a precision assembly, with a 32-microinch finish on its mating surfaces, and a high-conduction indium washer inserted at the interface between the transistor and the heat sink surface. Yet it was scarcely visible and, once installed, could

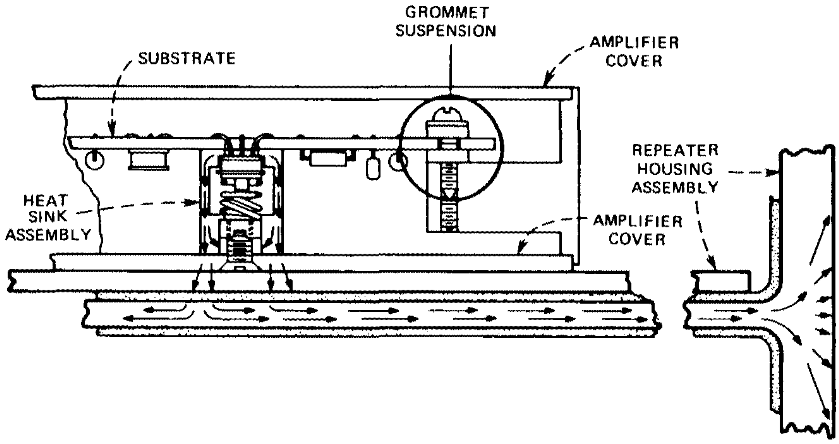


Fig. 10-31. Thermal path for L5 amplifier transistors mounted on a ceramic substrate.

be forgotten, contributing only to the long trouble-free life of the L5 repeaters [Fig. 10-31].

3.6 Increased Capacity—L5E

L5 was an immediate success both technically and economically. Following the initial installation, essentially all significant growth in coaxial routes was by L5 on new cables or conversions to L5 of earlier coaxial systems. In the later 1970s, after several years of relatively slow growth in the long-haul plant, circuit growth again accelerated rapidly and means were sought to increase the capacity of L5. While the networks to break up third-order intermodulation-product addition had proven highly successful, the system was still limited by second-order distortion and thermal noise. Any increase in circuit loading required an improvement in both types of noise.

Fortunately, in contrast to past experience when the best efforts had proven barely adequate, in one respect the L5 designers had built a system better than they needed. Extended studies of the L5 system after commercial service was started indicated that its equalization capability was more than adequate. The extensive circuits of equalizing networks and amplifiers generated excesses of both thermal and modulation noise, and the removal of unnecessary equalization became one of the two major noise-reduction techniques. All manhole-mounted E1 equalizers and all power-feed-station E2 equalizers were removed. In addition, the E3 equalizer was replaced by a pair of simpler equalizers. The other major noise-reduction technique used phase-reversing networks to reduce second-order modulation noise. Such networks were placed along the line at locations formerly occupied by the removed equalizers. Together, the two techniques reduced total noise by about 2.5 dB and permitted the expanded system to meet standard performance objectives in all channels for 4000 miles.

The noise-reduction measures permitted increasing the system capacity by four mastergroups to a total of 22, an increase of 22 percent over the initial service system. Accommodating the new mastergroups required new line-frequency allocations and providing them required new multiplexing equipment. It was essential for coordination with existing systems that many features of the original design be retained, such as the frequency of line pilots and the surveillance and administrative signals. The increased capacity in L5E (L5 expanded) was achieved by increasing the top signal frequency from 60.5 to 64.8 MHz and packing the mastergroups more tightly. Three multimastergroups, two with seven mastergroups and one with eight, replaced the three original six-mastergroup jumbogroups. Two new multiplexing arrangements, both of which provided less guard-band space in the frequency spectrum, were developed to support the new frequency allocation. (See Chapter 15, Section 2.5.) There was no provision for blocking, branching, or through connecting in multimastergroup blocks.

With these changes, a coaxial pair had a capacity of 13,200 high-quality telephone channels. During this period, standard sheaths with 22 coaxials had also become available. In such a sheath, with ten working pairs and one protection pair, an L5 coaxial route could carry up to 132,000 circuits. By 1980, L5 systems were providing almost 66 million voice-circuit miles of capacity.

REFERENCES

1. See special issue on The L3 Coaxial System, *Bell Syst. Tech. J.* **32** (July 1953), pp. 779-1005.
2. R. H. Klie, "The L3 Coaxial Cable System," *Bell Lab. Rec.* **32** (January 1954), pp. 1-4.
3. E. G. Morton, "Equalization in the L3 System," *Bell Lab. Rec.* **33** (October 1955), pp. 390-393.
4. C. G. Arnold, "Repeaters in the L3 Coaxial System," *Bell Lab. Rec.* **33** (May 1955), pp. 182-187.
5. R. S. Graham and J. P. Kinzer, "Adjustable Equalizers for the L3 Coaxial System," *Bell Lab. Rec.* **33** (December 1955), pp. 468-471.
6. D. B. Penick, "Carrier Terminals for the L3 System," *Bell Lab. Rec.* **33** (February 1955), pp. 72-77.
7. See special issue on The L-4 Coaxial System, *Bell Syst. Tech. J.* **48** (April 1969), pp. 819-1093.
8. See special issue on The L-4 Coaxial Cable System, *Bell Lab. Rec.* **45** (July/August 1967), pp. 210-236.
9. P. F. Molander, R. K. Bates, and C. C. Rack, "Amplification and Regulation—The L-4 Line Repeaters," *Bell Lab. Rec.* **45** (July/August 1967), pp. 225-229.
10. J. W. Mulcahy, W. P. Frawley, R. E. Maurer, and F. C. Kelcourse, "The L-4 Equalizing And Fault Locating System," *Bell Lab. Rec.* **45** (July/August 1967), pp. 230-236.
11. W. G. Albert, T. J. Haley, and T. B. Merrick, "L-4 Terminal Equipment—The MMX-2," *Bell Lab. Rec.* **45** (July/August 1967), pp. 218-223.
12. B. C. Meyer, "Manipulating Mastergroups On an L-4 Route," *Bell Lab. Rec.* **47** (October 1969), pp. 297-300.
13. See special issue on The L5 Coaxial-Carrier Transmission System, *Bell Syst. Tech. J.* **53** (December 1974), pp. 1897-2267.
14. See special issue on The L5 Coaxial System, *Bell Lab. Rec.* **51** (November 1973), pp. 290-331.
15. E. J. Eriksen, "The Development of Future Wide Band Carrier Systems," *Ericsson Rev.* **39** (1962), pp. 90-101.

16. P. Echarti, "Supermastergroup Translating Equipment for 60 mHz Carrier Telephone Systems," *Ericsson Rev.* **48** (1971), pp. 89-103.
17. J. F. Gunn, "Digital Transmission On the L-4 Coaxial System," *Bell Lab. Rec.* **48** (February 1970), pp. 51-55.
18. R. J. Wirtz, "Systems Approach to Physical Design," *Bell Lab. Rec.* **47** (December 1969), pp. 354-358.
19. R. J. Wirtz and W. G. Albert, "L5 System: Physical Design," *Bell Syst. Tech. J.* **53** (December 1974), pp. 2147-2194.

Chapter 11

Microwave Radio

I. THE TD2 SYSTEM

1.1 Initial Planning

In May 1946, just as the laboratory models of the first experimental microwave-radio system (TDX) were being tested, a conference was held to plan the next step. It was decided that a system linking New York to Chicago would serve as the basis for thoroughly exploring the technical feasibility of a nationwide relay chain. In October of that year, a Bell Laboratories committee proposed two alternative designs for the first commercial system, one to be completed in June 1949 and the other, one year later, incorporating more extensive improvements. The report was not overly optimistic about the performance of either system; nevertheless it was indicative of the spirit of the times that it was accepted as the basis for preliminary steps at Bell Laboratories to get the program going in advance of formal top management approval. Arrangements were made to shift the TDX group to the new project as soon as possible. Together with development groups in other areas, they started more detailed planning of the new system, to be called TD2.

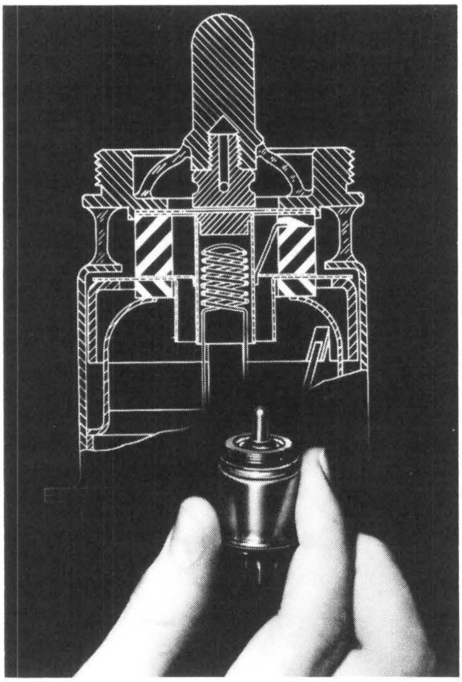
In January 1947, AT&T filed an application with the Federal Communications Commission (FCC) asking authorization for the construction of a microwave relay chain between New York and Chicago. Authorization was quickly granted, and development of TD2 was under way. AT&T selected the second, more ambitious of the Bell Laboratories proposals, but then urged that it be completed by the earlier date proposed for the simpler system! AT&T made this request because of mounting pressure from television broadcasters for interconnecting circuits. At the same time, however, the realization was growing that the Bell System was heading for a nationwide installation of radio relay on a crash basis. The system to be developed had to give at least acceptable transmission performance over transcontinental distances. Its reliability had to be comparable to that of existing, highly reliable cable systems. For maintenance costs to be kept within bounds, repeater stations had to be essentially unattended, with only infrequent routine maintenance visits.

Electron tubes were the biggest problem. The klystrons used in TDX limited the available radio-frequency (RF) bandwidths to about 10 MHz. For the base band of 4.5 MHz needed for television, even with narrow-swing FM, the sidebands were attenuated by the narrow band. There was no real prospect that such a narrow band could be amplitude and delay equalized adequately over

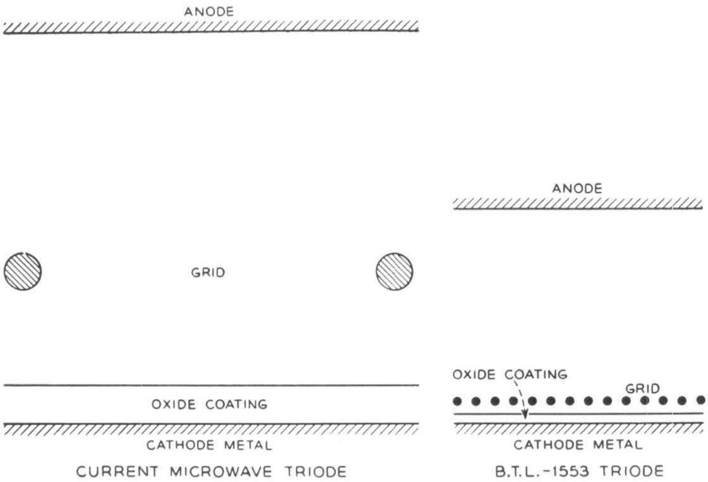
4000-mile distances or be kept accurately tuned to the right frequencies by manual means. In addition, it was unlikely that the life of the klystron could be extended to several thousand hours, even though tube replacement costs indicated that about 10,000 hours was needed. A possible solution to this dilemma was beginning to appear in the research department, where gain-bandwidth studies of triode amplifiers by J. A. Morton indicated the potential for handling wider bandwidths than was possible with klystrons. Morton had been working on new concepts of triode design during World War II. For a time, the work was directed toward a tube that might be used in the input of a 10-cm (3000-MHz) radar receiver. As the design began to take form and show promise and troubles with klystrons mounted, Morton judged that a better place for such a microwave triode would be in a radio-relay system. Some exploratory work on the tube, first called the No. 1553 and later the No. 416A triode, looked very promising [Fig. 11-1].[1,2] There was another competitor by this time in the traveling-wave tube (TWT), but it was generally agreed that a commercial TWT was still a few years away. In use, the microwave triode was incorporated directly into a waveguide cavity and made part of a three-stage amplifier [Fig. 11-2].

With the microwave vacuum-tube decision settled, the TD2 plan began to take shape along the lines proposed by Bell Laboratories in 1947, with one additional exception. The plan had considered the use of a horn-reflector antenna. The decision, however, was to continue to use a horn-lens antenna similar to the TDX type but somewhat redesigned. Since the horn-lens antenna handled only one plane of the polarization and could operate only over the 4-GHz band, this turned out to be a much less-than-optimum choice. As the planning of the project moved further along, it became increasingly clear that TD2 would not really be what had been talked about in the 1947 proposal. It was turning out that almost nothing would be carried over from the TDX, except the experience of having designed a microwave relay system. In fact, because of schedule overlaps, even experience with operation of the TDX system would come after the TD2 development was well started. A late 1949 service-date objective desired by AT&T was retained, however, with the hope that with a best effort and some good luck, it could be met.

The lack of hard numerical data on the propagation and fading characteristics of repeater hops was a major problem. Microwave propagation through the lower atmosphere is affected by rainfall, and refraction depends on temperature and humidity gradients. While it was possible to predict with considerable accuracy the loss along a radio path in conditions of standard atmosphere, predicting long-term fading statistics on a path meant predicting the weather with a detail and precision that would frighten the most optimistic weather forecaster. Instead, a massive effort was mounted in the Systems Engineering Department to acquire actual propagation experience in several widely spaced and carefully chosen locations. By the end of 1949, Bell Laboratories had a much better knowledge of the situation. But the design of TD2 started in 1947, so the designers had to proceed with objectives based on judgment applied to the sparse data then on hand. Fortunately the objectives chosen stood up fairly well; nothing big enough to justify design changes was found.

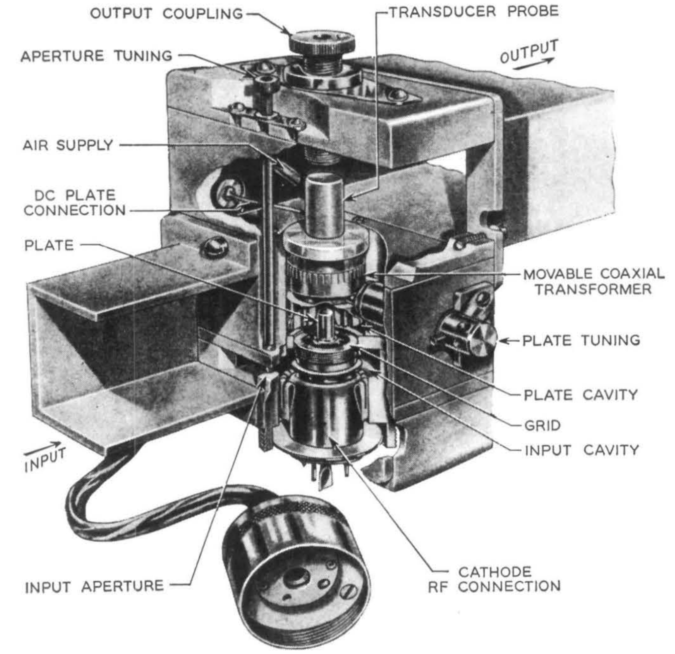


(a)

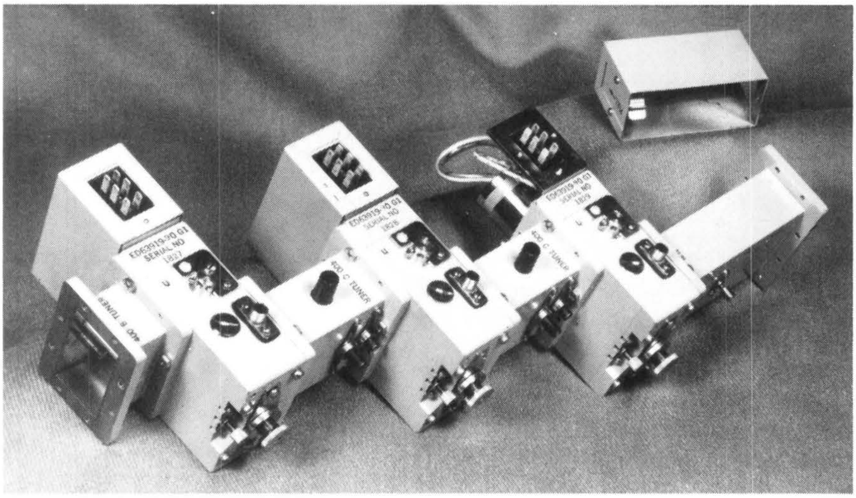


(b)

Fig. 11-1. The Bell Laboratories No. 1553 microwave triode. (a) The No. 1553 triode with a cross-sectional drawing of it in the background. (b) Comparison at the same scale of, *right*, the electrode element spacing of the No. 1553 triode with, *left*, a contemporary microwave triode.



(a)



(b)

Fig. 11-2. The TD2 microwave amplifier. (a) The No. 416A triode mounted in a waveguide cavity. (b) The three-stage microwave transmitter amplifier.

1.2 Scheduling the First Systems

When the TD2 project started in the first months of 1947, a tentative schedule required that Bell Laboratories complete development of the repeater, antenna, power plant, and all other items by the first quarter of 1948. Western Electric preparation for manufacture was to overlap the development, so equipment delivery could start in the last quarter of 1948. The Long Lines Department was to choose and acquire sites, design and construct buildings and towers for the route between New York and Chicago, and have them ready when the first equipment started to flow out of the Western Electric factories. Installation of equipment was to start early in 1949, working from east to west, so that system testing of the eastern end could go on while the western end was still being installed. The service date was to be the fall of 1949.

Even while this schedule was being established, the concept of the TD2 project was growing into one of a nationwide system with long-term commitments. A short experimental system could be built of devices and components in an early stage of development. If failures occurred or maintenance costs were high, no great harm was done. In fact, one of the main motivations for an experimental system was to smoke out the unexpected problems that had not shown up in the laboratory. But for a nationwide system, it was essential that the design concept be sound, and the life and reliability of the devices be assured. The TD2 project began to place increasing demands on activities in all parts of Bell Laboratories, such as systems engineering, development of devices and power equipment, and research in chemistry. It also began to be accepted as an objective that TD2 should be able to carry about 480 telephone channels when not used for television. Telephone needs, which had been set aside by the early concern over television, began to assume their previous importance. The impact of the enlarged effort required and the higher-capacity telephone objective was recognized in 1948, and the projected New York–Chicago service date was moved to the fall of 1950.

At the same time, Bell System commitments to the success of the TD2 project continued to grow. To meet critical service needs between Los Angeles and San Francisco, a chain of stations was projected for essentially the same time as the New York–Chicago system. (In fact, it opened for service only 14 days later.) Also, the long link between Chicago and San Francisco was being engineered and the buildings constructed. The date for the initial service was pushed back a year, but the dates for what had been follow-up systems remained unchanged. There was going to be a tremendous commitment in metal and concrete around the country before any TD2 equipment actually passed its first service tests.

1.3 First Service and Early System Improvements

By the spring of 1950, the first repeater bays were assembled and installed, and testing could be started over loops around the first few repeater stations out of New York [Fig. 11-3]. There were numerous problems at first, such as a poor signal-to-noise ratio, spurious tones, high intermodulation noise on telephone loads, and variations in video response. Early failures of the No. 416A triodes were also experienced. One by one, problems were cleared and

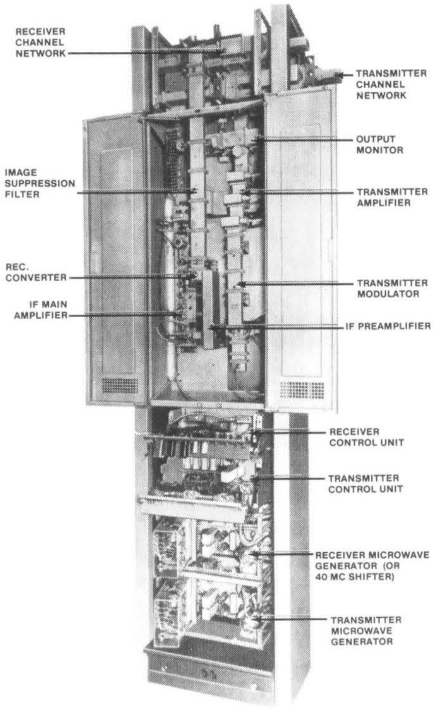


Fig. 11-3. The TD2 radio repeater bay.

performance improved to at least a tolerable level of quality and continuity. By August 1950, good television transmission had been achieved. The New York–Chicago route was opened to service on September 1, 1950 and the Los Angeles–San Francisco link by mid-September. The entire coast-to-coast route was complete in time to carry President Truman's opening address before the Japanese Peace Treaty Conference from San Francisco to New York on September 4, 1951. The design of TD2 had become a reality.[3,4]

System outages were still painfully frequent but began to decrease rapidly in 1951. Some of the worst spurious tones were cleared by adding filters, tightening waveguide joints, and similar measures. An improvement program for TD2 was begun even before the service date on the New York–Chicago route. The most pressing problem was improving the life of the No. 416A tube. This problem was receiving massive effort from the Bell Laboratories and Western Electric engineers at Allentown, Pennsylvania. (In the late 1940s and through the 1950s, branches of Bell Laboratories were established at several locations; those for electronic devices in Allentown and Reading, Pennsylvania, cable and wire products in Atlanta, and transmission equipment at the Merrimack Valley, Massachussetts location were especially important to transmission technology.) At the Allentown laboratory, the defects of bad tubes returned from the field were analyzed, relationships of manufacturing techniques and processes to tube life were studied, and the entire triode was redesigned. In consequence, the life was brought up from 100 hours or so in the earliest models to 4000 hours in 1951. By 1952, it was in the 6000- to 8000-hour range and reached about 20,000 hours in the mid-1960s.

There was, additionally, a whole series of normal follow-up engineering improvements, such as increasing the life of other tubes, strengthening the antennas to withstand higher wind loads, removing maintenance hazards in high-voltage circuits, and designing more portable test equipment. A major improvement was the design of an automatic-protection switching system.[5] This system permitted one RF channel to act as protection for the other five, substituting itself automatically and rapidly for a channel that dropped out, either because of equipment failure or deep fading. This reduced the system outage by a factor of 50 and was an essential element in making radio circuits as reliable as those on cable.

The improvements called for in delay-distortion equalization and in stabilization of the baseband gain-frequency characteristic were a different matter, however. The designers did not have adequate analytical tools for a numerical understanding of what was happening. They had the classical FM theory that made the equations tractable by dropping troublesome higher-order terms. Such theory was adequate for simple single-hop systems, but it gave no quantitative understanding of practical multilink systems. In such systems, gain and delay distortions were followed by amplitude compression, and this process was iterated in the multiplicity of repeaters. It was not until later that high-performance computers made it possible to simulate this process and calculate the results with sufficient accuracy.

In addition, in the early days of TD2, the test equipment available was not accurate and precise enough to make adequate tests on a single repeater or its elements. This meant that observations were made on a string of repeaters, with all the obscuring effects of compression and unknown maintenance variations. At that time, the problem could not be completely understood in numerical terms. But if the Bell System had decided to wait until FM transmission was thoroughly understood before building TD2, there would have been no system. In fact, it was the existence of this nagging problem on TD2 that sparked the extensive work that later gave Bell Laboratories a quantitive understanding of multihop FM transmission and set the stage for the later enormous improvements in microwave systems.

1.4 Increased Spectrum Use

Soon after TD2 was developed, efforts were begun to increase system capacity. In the original design, the number of radio channels that could be transmitted was limited because only half the allocated frequency band between 3.7 and 4.2 GHz could be used; in the late 1940s, it was not possible to design microwave filters that would separate immediately adjacent channels [Fig. 11-4]. In addition, the number of routes that could be developed in congested areas was limited. If two or more routes converged or crossed at a tower, the angular spacing between them had to be about 60 degrees, since only antenna directivity separated potentially interfering channels. In 1951, a set of improved

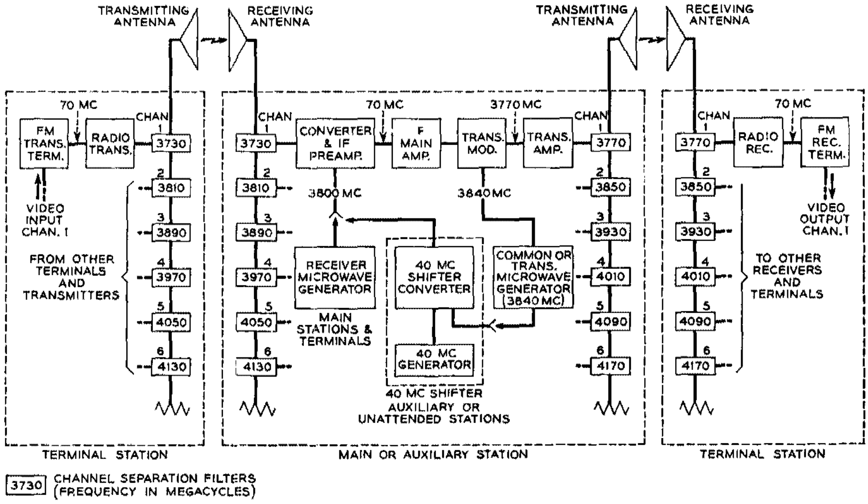
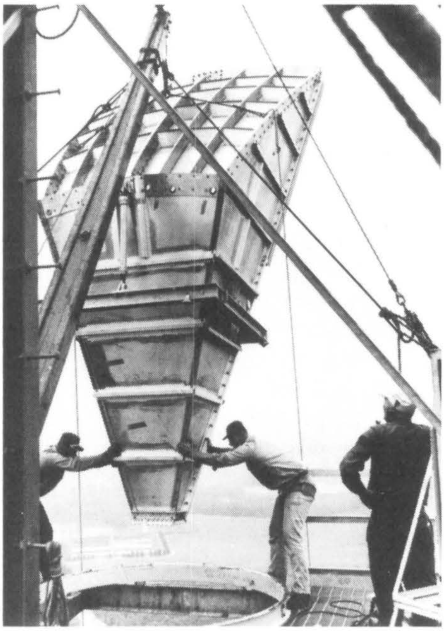


Fig. 11-4. TD2 radio system showing original six-radio-channel frequency allocations.

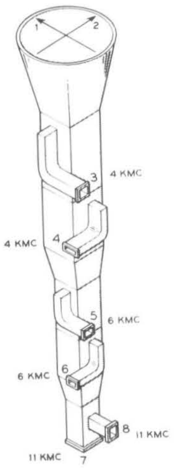
filters was designed so that channels on different routes could occupy the alternate empty slots in the standard TD2 system. About 20 dB of selectivity was provided, which made it possible to reduce the angular separation at junctions to as little as nine degrees. These filters proved very useful in special cases.

In 1955, another factor was added that changed the situation radically. By this time, Bell Laboratories designers had started on the design of the TH microwave system for the 6-GHz common-carrier band. It was expected that the TH system would be widely used as an addition to TD2, utilizing the same sites and towers on routes that already had TD2. However, because the TD2 antenna transmitted only the 4-GHz band, the 6-GHz TH signals could not pass through the same antenna. This raised the problem of adding another set of antennas to existing TD2 installations, obviously a difficult and cumbersome operation.

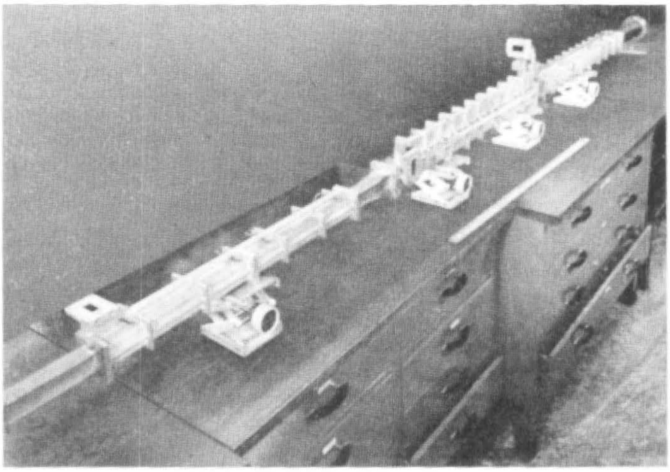
But the horn-reflector antenna was broadband enough to handle 4- and 6-GHz signals simultaneously. (See Chapter 7.) The horn-reflector antenna could also transmit both horizontal and vertical polarizations. There was still a question, however, as to whether the antenna and all the associated common-feed lines could be designed to handle the two frequencies simultaneously. This involved a waveguide that extended down from the antenna to the repeater equipment, large enough to transmit 3.7 GHz, that was round to handle both polarizations, and regular enough not to get into multimoding problems. (At 6.2 GHz, more than one mode was possible in a waveguide large enough for 4 GHz.) At the base, system-dividing networks were needed to separate the signals into the two major system bands. It turned out to be no small job to design such a combining system, especially since the design contemplated the eventual joint use with systems at 11 GHz in addition to those at 4 and 6 GHz. It was only with the pioneering contributions of the Radio Research Department that it was adequately solved [Fig. 11-5].[6] Even after the problem had apparently



(a)



(b)

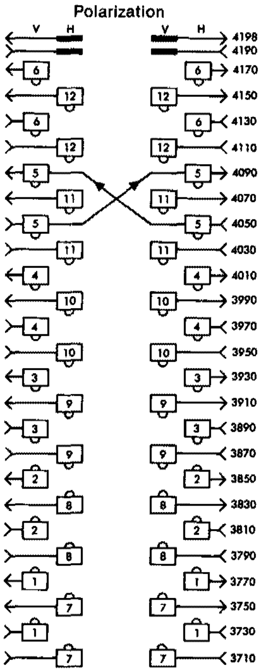


(c)

Fig. 11-5. Multiband microwave system elements. (a) Installing a horn-reflector antenna. The antenna could transmit the three common carrier bands at 4, 6 and 11 GHz and both horizontal and vertical polarizations simultaneously. (b) Dominant-mode terminal pairs for 4-, 6- and 11-GHz systems combining network. (In the 1950s GHz were designated kilomegacycles, KMC.) (c) Laboratory model of systems combining network.

been completely solved and the design was in production, serious multimoding problems remained. These were later solved by redesigning and meticulously aligning the waveguide and antennas.

The first step, therefore, in preparing to use the TH system was to redesign the entire TD2 antenna system and start applying the new design on all new TD2 routes, beginning in 1955. Once TD2 antennas could transmit both polarizations simultaneously on the same route, the questions arose naturally: How much discrimination from filtering and polarization was necessary to use both the regular and the intervening frequency slot channels in the same broad band (i.e., 4 GHz) on the same path? Would the polarization discrimination remain during fading? To answer the latter, tests were conducted between Bell Laboratories locations at Murray Hill and Holmdel, New Jersey, about 23 miles apart. The experiment indicated that polarization discriminations poorer than 20 dB might be expected less than 0.002 percent of the time. This was good enough to contemplate a frequency plan that would use the 4-GHz band very efficiently, doubling the number of channels possible. Transmitters or receivers on the same frequency were on oppositely directed antennas and on opposite polarizations [Fig. 11-6]. Careful analysis of the requirements on separation between transmitters and receivers achieved by the antenna directivities (and



Notes:
1. Indicates local-oscillator 70 MHz below channel frequency.
2. Indicates local-oscillator 70 MHz above channel frequency.
3. Indicates auxiliary channel assignment.

Fig. 11-6. The 4-GHz two-frequency allocation plan used for TD radio systems, showing the realization of 12 two-way radio channels by using cross polarization.

confirmed by field measurement) showed that some added frequency selectivity was required. A filter was added in the intermediate-frequency (IF) amplifier portion of the repeater to sharpen the discrimination against the adjacent channel.[7]

By 1966, over half of the TD2 routes were equipped for interstitial channels, affording a considerable savings by using existing installations more intensively. It is ironic that installing a flexible antenna system on TD2 routes for TH application ultimately delayed the growth of TH. But, of course, the delay came about because it was more economical to add interstitial TD2, another example where the increase in the capacity of an older system raises the competitive hurdle for a new one. The doubling of the channels also had an impact on the protection switching system. The original system used one radio channel to protect the other five working channels. With ten working channels, two such systems would be needed, a cumbersome and costly proposition. Accordingly, a new protection switching system was developed, the 100A, which directly provided two-for-ten protection capability.[8]

1.5 Microwave Antennas and Towers

The horn-reflector antenna was a clearly superior design; it was broadband and transmitted both needed polarizations.[9–13] The reflector, however, called for a shape accurate to within ±1/8 inch over its outside dimensions of 8 by 8 feet. This required factory tooling, which at the time loomed as a serious problem in cost and time. It was clear that it would be easier to build a few delay-lens antennas than a few horn-reflector antennas. The delay-lens antenna was fabricated on the job by hand and with relatively modest dimensional tolerances. Had it been perceived that thousands of antennas were going to be required, the choice of factory tooling over hand labor would have been obvious, though the time factor would still have been serious. In addition, the whole concept of a radio-repeater station had changed over the years. Initially, designers had thought in terms of a small building, housing a couple of transmitters and receivers, with the antennas on the roof. This philosophy was carried out in the design of the TDX system. When necessary, antenna decks were raised a little to get over nearby trees but only by the minimum amount to avoid increasing the length and loss of the connecting waveguide. For this arrangement, the horizontal mounting of the delay-lens antenna was well suited.

This thinking carried over to the first TD2 stations. Structural steel was scarce in the postwar era, and it was thought that there would be economies in a poured-concrete structure built like a silo. Only a few of these were built before it was clear that it was not economical. A building with a square or rectangular cross section is much better suited for mounting equipment in rows of bays. As the system moved into the plains of the Midwest, tower heights were increased to preserve line-of-sight paths over spacings of 25 or 30 miles. In the East, it was usually easy to find a hill to give 100 to 200 feet of elevation, but, in the Midwest, the tower itself had to meet the need. The flat topography presented a problem that at first resulted in a rather uneasy compromise. On the one hand, it was undesirable to mount the radio equipment too far below the antennas because of loss in the connecting waveguide. On the other hand, because of cost and other difficulties connected with installing elevators, it was undesirable to place the radio equipment any further above the ground than necessary [Fig. 11-7]. It was clear, then, that these two functions should be separated. The building would house the radio and power equipment, the tower would support the antennas at the proper height [Fig. 11-8]. Also, as the number of channels in the 4-GHz band was doubled and the plan to add 6-GHz systems on the same antenna took shape, it became obvious that the stations would grow. Expandable buildings were essential; it was much easier to plan for expansion with a building at ground level and a steel tower alongside. The vertical mounting of the horn-reflector antenna became preferable because it allowed the waveguide to start down the tower with a minimum of bends.

The decision to use the delay-lens antenna, then, was made in one set of circumstances, while the application turned out to be in another. With sufficient foresight, or a better crystal ball, the manufacture of several hundred antennas that eventually did not fit the growth pattern might have been avoided. But this was part of the price that went with the very compressed schedule. The same can be said of the approach by successive approximations to the design of buildings and towers. Some steps in design thinking that normally would have existed only on the drawing board were frozen in hardware and concrete.

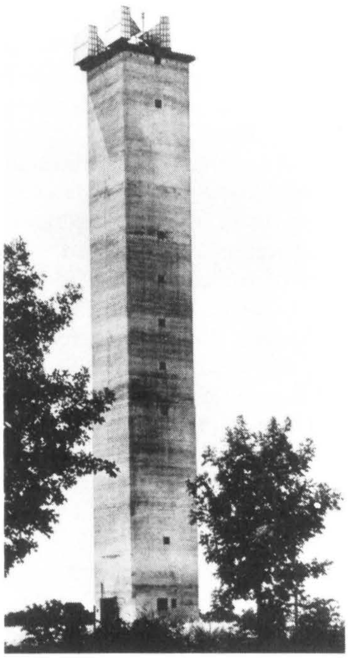


Fig. 11-7. A typical concrete tower used in the first TD2 system between New York and Chicago. The radio repeater equipment was placed part way up the tower to reduce waveguide runs.

Fig. 11-8. A 125-foot steel tower for TD2.

1.6 TD3—The Next Version of TD2

In April 1962, it was decided to redesign TD2 into an entirely new version, called TD3, taking advantage of all the advances in technology available.[14] The TD3 system made use of the band from 3.7 to 4.2 GHz in the same way as TD2, combining regular and interstitial channel versions. It was to be all-solid-state except for a TWT amplifier with a 5-w output. The noise-performance objective was several decibels better than TD2, so that it could carry 1200 telephone channels in a single RF channel (compared with 480 to 600 channels for TD2). This increased capacity required not only the higher transmitter power of the TWT but also a better receiver input-noise figure. The latter was achieved by the use of new low-noise Schottky-barrier diodes in place of the original silicon point-contact rectifiers. The intermodulation requirement, and hence the equalization of delay distortion in TD3, was so severe that it required the development of new analytical techniques and precision laboratory measuring equipment. Work on computer modeling had advanced to the point where modulation performance could be predicted from measured or any given shape of amplitude and delay distortion over the radio band, including the nonlinearities and amplitude-to-phase conversion effects of the TWT.

The first TD3 installation was on a new radio route consisting of a five-hop single switching section about 140 miles long between Alexander, Arkansas and Arkabutla, Mississippi. This section was the first portion of a new radio route planned for service between Alabama and Oklahoma in 1967. Manufacture of TD3 extended from 1966 to 1971, with a total of 2358 transmitter-receiver units produced.

TD3 met all objectives and worked very satisfactorily, but some of the new components were quite expensive. At the same time, as experience was gained in the manufacture of TD2, its performance was improved and costs were reduced. In the late 1960s, as the original TD2 equipment was constantly upgraded, joint studies by Bell Laboratories and Western Electric indicated a number of areas where cost reduction and equipment simplification in TD3 could be obtained. On this basis, a decision was made in 1970 to proceed with an improved version, TD3A. In 1973, the best and least costly components of TD2 and TD3A were further amalgamated into a new system, TD3D, which also incorporated better maintenance and testing capabilities. When the TD3 system was projected, a TWT was absolutely essential to obtain the required transmitter power. Continued improvements in the microwave triode had been made, however, until it yielded power comparable to the TWT and at much lower cost. In TD3D, the TWT was replaced with the triode amplifier, but it was anticipated that a solid-state transmitter would soon be available. This was a striking sequence. The new generation design, with the objective of doubling the capacity of the original TD2, was overtaken by incremental improvements in the old model until a new hybrid proved to be better and less costly than either.

1.7 Capacity Increases of TD2

The increase in the telephone channel capacity of the TD radio systems from the first tentative estimates of a few hundred channels to the final accomplishment is a remarkable story. Unlike single-sideband AM (SSBAM) systems on cable or wire, the FM radio systems were not limited by the 20-MHz bandwidth of the radio channels. If the signal-to-noise ratio could be improved, more telephone circuits and a wider baseband could be transmitted by lowering the frequency deviation. Obviously, if this could be done with relatively minor changes in the installed equipment, the saving over installing additional radio channels or building new routes was substantial. In addition, this increased route capacity without new frequency assignments in the always-crowded radio spectrum, a very good benefit indeed.

The first step was to raise the power output of the microwave-triode amplifier from 0.5 to 1.0 w. With the adjustments available, it was easy to raise the amplifier gain and drive the output stage to produce 1 w. However, when this was done, the performance with the wider baseband was poor. The cause was carefully analyzed and detected to be amplitude-to-phase conversion in the microwave amplifier. The surprisingly simple cure was to move the connection of a resistance from one side of a capacitor to another. The change could have been made at a much earlier date improving overall baseband transmission performance, but the design defect was not revealed until the power was increased.

The noise in a microwave system carrying hundreds of telephone conversations came from two sources. The first was thermal noise, whose effects could be reduced by increased transmitter power or lower receiver noise figures. The second source was due to the intermodulation of the telephone signals. In FM systems, a major cause of intermodulation was delay distortion in the repeater circuits. This source had been worked on throughout the history of TD2. Intermodulation noise could also arise from multimode propagation effects in the waveguide systems connecting the radio repeaters to the antennas. Discontinuities and poor impedance matches caused conversions from desired to undesired modes or echoing and reechoing of the desired modes at the discontinuities. These effects were recognized in early TD2 planning and were important factors in the desire to keep the antennas as low as possible on the roof of the equipment building.

As network growth into the Great Plains resulted in the use of taller towers that were separated from the equipment building, longer and more complicated waveguide runs were needed. In addition, the shift to an antenna system that operated over several frequency bands introduced the complicated system-dividing networks. So a source of intermodulation noise, minor in the early days, became major. The next step in the improvement program was to clean up the entire antenna system. This involved, among other things, better aiming of the antennas when it was realized that radiated waves slightly off the axis of the antenna also excited unwanted modes. New test equipment, some new parts with better impedance match, and better alignment of the waveguide flanges were also important [Fig. 11-9]. Tests in 1966 on the improved version, designated TD2A, established that the system could carry up to 900 circuits per channel and still meet the long-haul objectives.[15]

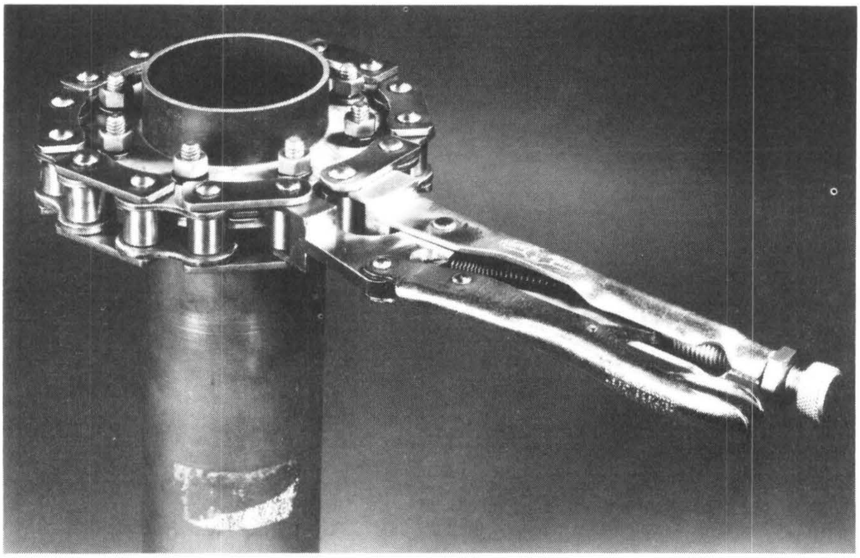


Fig. 11-9. A special wrench designed to align round waveguide runs on microwave radio-relay towers. It consisted of a gripping wrench with a length of stainless-steel roller bearing chain, which held the two mating flanges precisely in position while the bolts were being tightened.

A year later, further improvements were tested on a route between Dallas and Oklahoma City. Western Electric engineers had been working on the 1-w triode microwave tube to introduce ceramic insulation in place of the glass that had been used previously. The improved No. 416C tube could dissipate more power because of the high thermal conductivity of beryllium oxide and thus permitted raising the microwave power to 2 w. With improved components from the TD3 system, such as the Schottky-barrier low-noise down-converter, and improved FM terminals, it proved possible to carry 1200 circuits and still meet long-haul objectives.[16] Encouraged by this success, designers started to work on further increases in capacity. Additional tests on the No. 416C triode showed that a transmitted power of 5 w was possible with selected tubes, although the life expectancy was reduced. The increase in capacity to 1500 circuits, possible at the higher power, was judged to be worth the additional maintenance effort required. By 1973, this system, called TD2C, had been installed on a route linking Kansas City and St. Louis and on several others.[17] This was the last gasp for the microwave triodes. Even at the higher power, its life expectancy was about a year of continuous operation. It had come a long way from the days of 0.5 w and 100 hours.

About this time, the FCC introduced a regulation limiting the number of protection channels permitted in the common-carrier bands. Whereas TD2 had been operated for a number of years with two protection channels for ten working channels, the new regulations allowed only one protection channel. This provided a strong motivation to convert to more reliable solid-state components. The IF amplifier and microwave generator were among the units converted. Other improvements about this time included the introduction of maintenance-only-when-needed procedures at TD stations, easier methods for tuning the RF amplifiers, and the introduction of a carrier resupply system at each repeater to prevent downstream repeaters from going to full gain in the event of signal loss on a hop.

1.8 The Final Changes

The trends established earlier towards lower cost, better performance, higher capacity, and conversion to solid state continued after 1973. During the next six years, all electron tubes were replaced, including the 4-GHz/RF power-amplifier. Two versions of the solid-state microwave amplifier were produced at 2 and 5 w, using the newly-developed gallium arsenide field-effect transistor (GaAsFET). In addition, a low-noise GaAsFET RF receiving-preamplifier was introduced for use on long hops with relatively poor signal-to-noise ratios.

As transmission improved throughout the plant, talkers spoke at a lower volume. In the mid-1970s, it was ascertained that the average talker volume on the long-distance network was about 4 dB less than the value used in establishing the earlier radio system design-parameters.[18] This meant that the signal from the sum of the talkers was not driving the FM radio system to full deviation as planned, and hence the RF spectrum of even 1500 such talkers was narrower than allowed by FCC rules. In March 1980, the system load was increased to 1800 circuits per radio channel, which could be done and still meet noise objectives if the transmitters were operated at 5 w and the receivers were equipped with the new low-noise RF preamplifier.

TABLE 11-1: Increase in TD System Capacity

Year	Transmitter Power-Watts	Telephone Circuits Per Radio Channel	Radio Channels Per Route	Telephone Circuits Per Route
1950	0.5	480	5	2,400
1953	0.5	600	5	3,000
1960	0.5	600	10	6,000
1967	1.0	900	10	9,000
1968	2.0	1,200	10	12,000
1973	5.0	1,500	11	16,500
1980	5.0	1,800	11	19,800

Other improvements continued to occur. System modeling with more powerful computers led to a better understanding of system delay-distortion and resulted in significant improvements in overall baseband response. New microwave technology using barium titanate made possible the introduction of microwave filters of smaller size and better temperature stability. Finally, a standard method was developed to combine signals from two vertically separated antennas on a tower (space diversity) in a smoothly continuous way to provide a steadier signal during periods of multipath fading. This method was to be essential in conditioning the TD systems to carry high-speed data in later years.

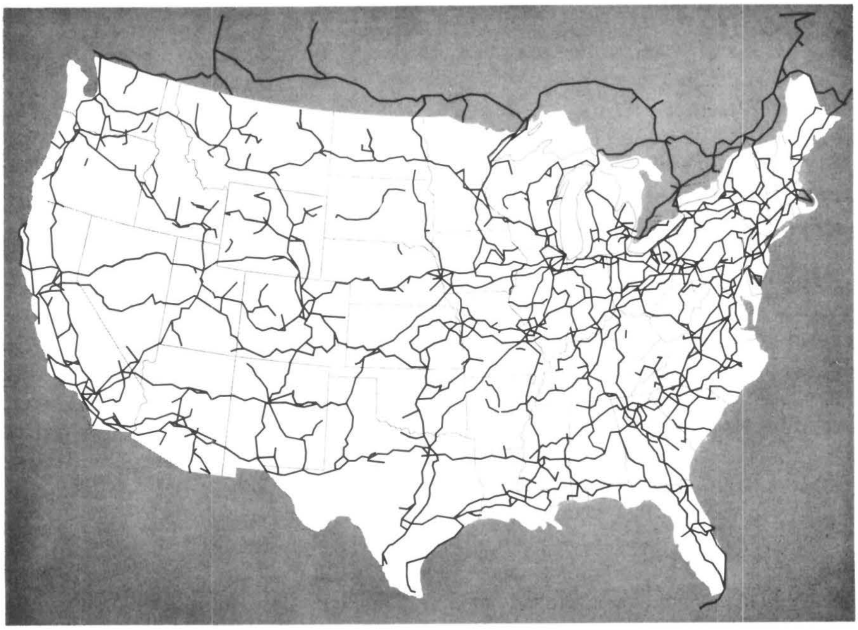


Fig. 11-10. TD2 system broadband routes, 1980.

In the 25 years following initial production in 1950, some 35,000 TD2 repeaters and 5000 TD3 repeaters were produced and installed around the country. Dramatic increases in radio-channel and route capacity were achieved [Table 11-1]. By 1980, microwave radio provided over 600 million voice-circuit miles in the Bell System plant; almost three-quarters of the total was on TD2 and TD3 systems. TD systems touched virtually every city of any size in the United States [Fig. 11-10]. TD was a tremendously successful development. It is an outstanding example of the constant application of new technology to achieve a steady increase in capacity and performance over the years, with a corresponding decrease in the cost of long-distance service.

II. TH SYSTEMS

2.1 TH1—The First 6-GHz System

Bell Laboratories engineers began to plan for a 6-GHz system in 1955, anticipating the need for additional circuit capacity when the rapidly expanding TD radio network would be completely filled, at least in its busiest sections. The design objective was to provide the same quality long-haul service as TD2, using as much of the TD2 plant as possible. The FCC had established the 500-MHz band from 5.925 to 6.425 GHz for common-carrier use. In the TH plan, this was divided into sixteen 30-MHz slots paired to provide eight 2-way radio channels, plus some narrower bands for auxiliary communications. Each broadband channel was to be capable of transmitting a baseband signal of 1860 frequency-multiplexed telephone channels (the same as L3 coaxial) or, alternately, a full-bandwidth color-television signal.

Compared to very-high-frequency FM radio, the microwave systems were low-index FM systems. The frequency deviation required for an adequate signal-to-noise ratio in a given telephone channel depended on the transmitter power, the receiver noise figure, the repeater spacing, and the antenna gain, all of which could be traded to a considerable extent. In the first microwave systems, the frequency swing was initially about ±4 MHz, comparable to or even less than the baseband signal band. By Carson's rule (a formula for the frequency band occupied by the principal components of an FM signal, derived in the 1920s by J. R. Carson of AT&T), the band required with FM is twice the band of the modulating signal (the baseband) plus twice the frequency deviation. With a ±4-MHz swing, the 8 MHz of the multiplexed telephone baseband required a radio band of about 25 MHz. Another factor in the choice of 30 MHz (29.65 MHz, to be precise) was the expectation that high-definition theater television with a band as wide as 10 MHz would also be carried. The TH frequency plan differed from the alternating transmitter-receiver assignment of TD2 in another major respect. In the TH system, one-half of the band was used as a block for transmitting and the other half for receiving, alternating hop-by-hop [Fig. 11-11].

The principal challenges in the TH system design included achieving a sufficiently high signal-to-noise ratio to handle additional circuits, providing new RF apparatus, such as channel-dropping networks and amplifiers, providing better delay-distortion equalization over the broader channel bandwidth, and integrating the new system into the existing TD2 network. For the higher power

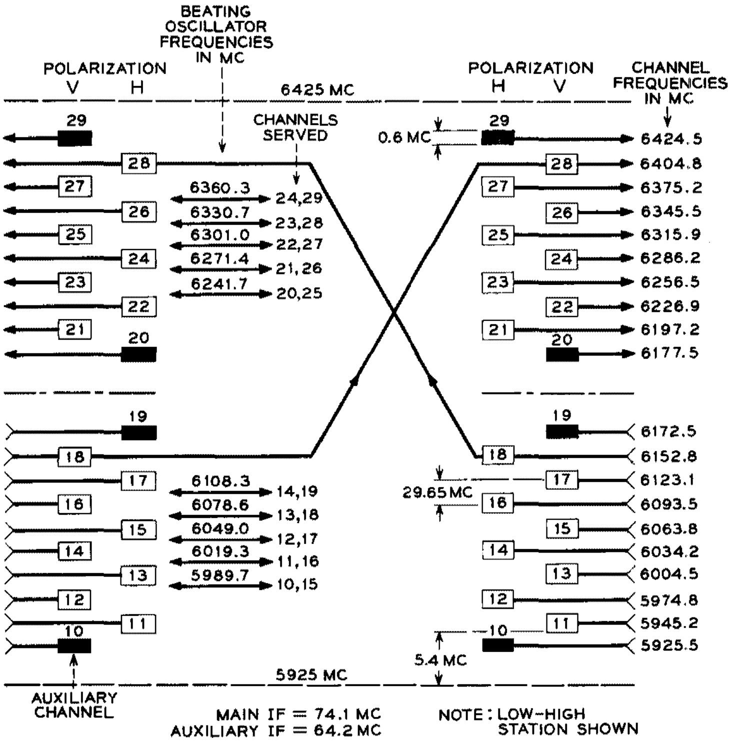


Fig. 11-11. The two-frequency plan—frequency allocation and polarizations for the TH system channels.

and bandwidth required, the TWT had no serious competitor. This device had been invented in England during World War II by R. W. Kompfner and much improved in the Bell Laboratories research area by J. R. Pierce and his associates [Fig. 11-12]. The TWT overcame the principal limitations that had bedeviled all the earlier microwave amplifiers by providing high power and gain over an extremely broad band and, in addition, had the promise of very long life. The version developed for TH radio was focused with permanent magnets and cooled with forced air. To eliminate interaction with adjacent amplifiers and protect personnel from the high voltages required, a massive magnetic shield enclosed the entire structure [Fig. 11-13]. The TWT was capable of an output power of 5 w over the entire 500-MHz band at 6 GHz. In conjunction with a new balanced down-converter and a low-noise IF preamplifier at the receiver, it provided an adequate signal-to-noise ratio over the hop.

A new approach was taken in the design of the microwave generator used to supply the various 6-GHz carriers for frequency conversions. Instead of each microwave repeater having its own generator, a common microwave carrier supply furnished all the carrier frequencies required at a radio station. With a common supply, it was expected that the cost per repeater at a fully loaded

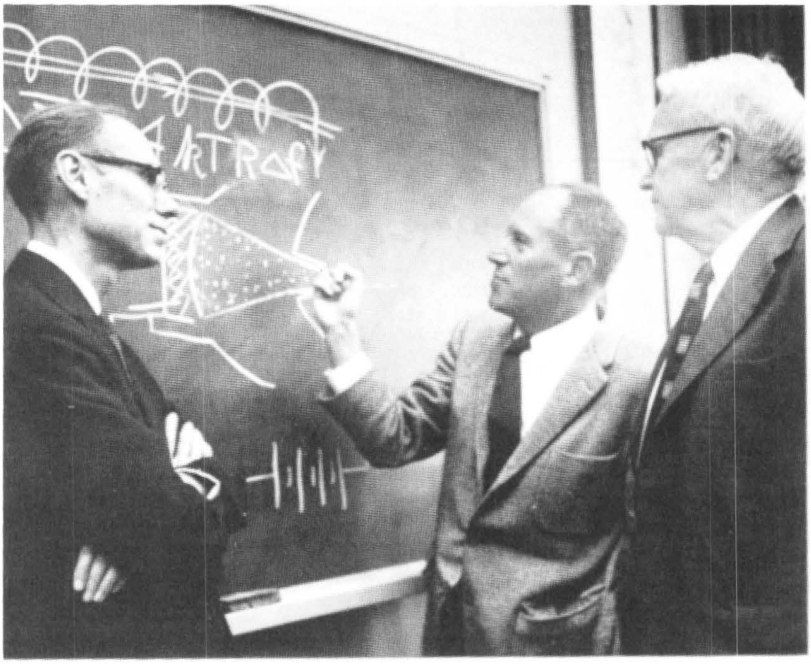


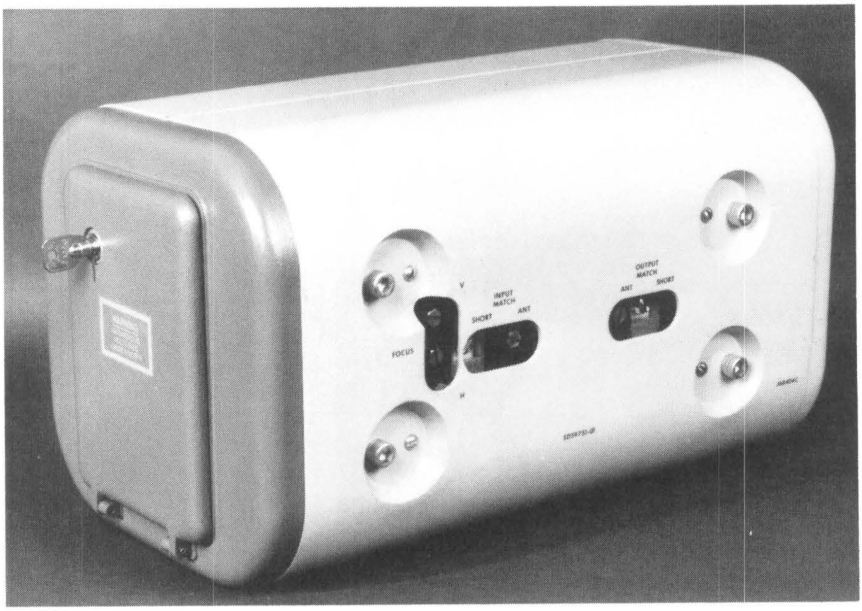
Fig. 11-12. *Center,* R. W. Kompfner, the inventor of the traveling-wave tube, discussing its properties with, *left,* J. R. Pierce, who played a key role in perfecting it, and, *right,* H. Nyquist, who made important contributions to its theory of operation.

station would be greatly reduced. Maintenance access was provided by mounting TH equipment in sliding racks, which required a flexible waveguide connection to the channel-separation filters and antenna waveguide [Fig. 11-14].

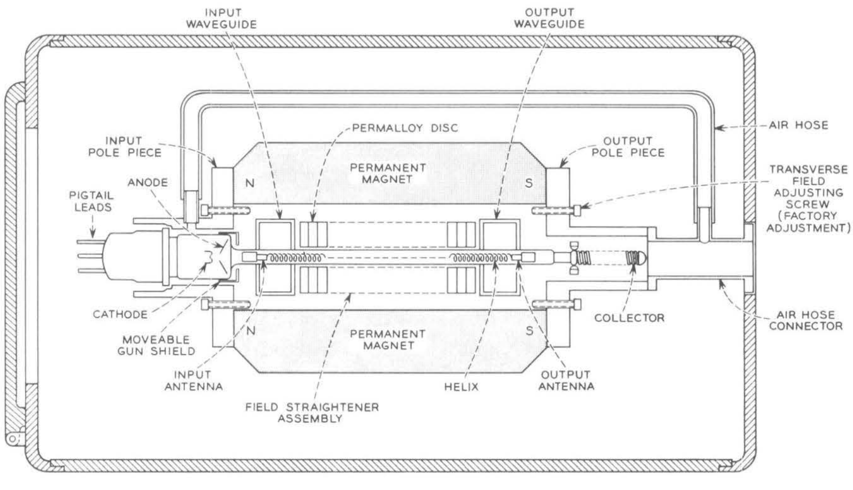
A field trial of TH on a route between Salt Lake City and Denver in 1960 revealed a few unanticipated sources of interchannel interference, but otherwise the trial went well. The interference was eliminated with relatively minor design changes. Carrying over 1800 telephone circuits per radio channel, the TH1 system provided a significant addition to the capacity of the microwave network.[19] About 4000 repeaters were installed by 1970, when manufacture was discontinued.

2.2 Improvements

TH1 radio met its objectives and filled a need, but it had some shortcomings. A great deal was being learned about the practical needs of the field with the widespread installation of the several generations of TD radio. The principal problem with the TH system was that it had been designed with the expectation of a rapid growth in signal load to the full capacity of the system. Under such conditions, the cost per channel would have been about as low as expected. But even in a rapidly expanding system, there were many situations where



(a)



(b)

Fig. 11-13. The TH radio system traveling-wave tube. (a) The traveling-wave tube enclosed in its large magnetic shield. (b) Cross section of traveling-wave-tube amplifier.

growth was not fast. In these cases, the start-up cost of the microwave carrier supply designed for the full complement of channels, along with other equipment furnished in common, put a heavy cost burden on the first 6-GHz channel

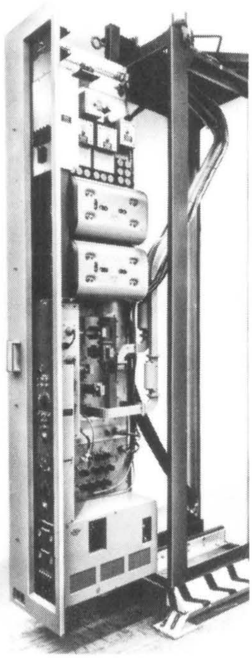


Fig. 11-14. TH1 radio repeater transmitter bay showing traveling-wave amplifiers in sliding rack mounting.

to be equipped. The experience with TD was demonstrating that a self-contained transmit-receive bay was most desirable.

In 1966, work started on an improved TH system, called TH3. TH3 was designed as a new long-haul facility to provide modern solid-state equipment with improved performance for use in the 6-GHz band.[20] Many of the new system components were the same as those used in TD3. The radio bay required new circuitry because of the higher RF frequency and wider IF bandwidth, although design techniques were similar to those used in TD3. The IF frequency was also the same as TD3, i.e., 70 MHz instead of the 74 MHz used in TH1. This allowed TH3 to use the FM terminals (upgraded to 1860 circuits), a protection switching system, and some of the IF circuits developed for TD3. A new TWT was designed for the RF power amplifier that dispensed with the massive shield and was capable of operating at 10 w instead of 5 w, as in TH1 [Fig. 11-15].

Laboratory testing with prototype and preproduction bays was carried on in 1967 and early 1968. TH3 was first installed on a nine-hop route between Vega, Texas and Dodge City, Kansas in 1969 and 1970. Only minor problems were experienced in the first system, and these were quickly corrected. TH3 then met its long-haul objectives for 1800-circuit loading. About 2000 repeaters of TH3 were eventually manufactured and installed. As in earlier systems,

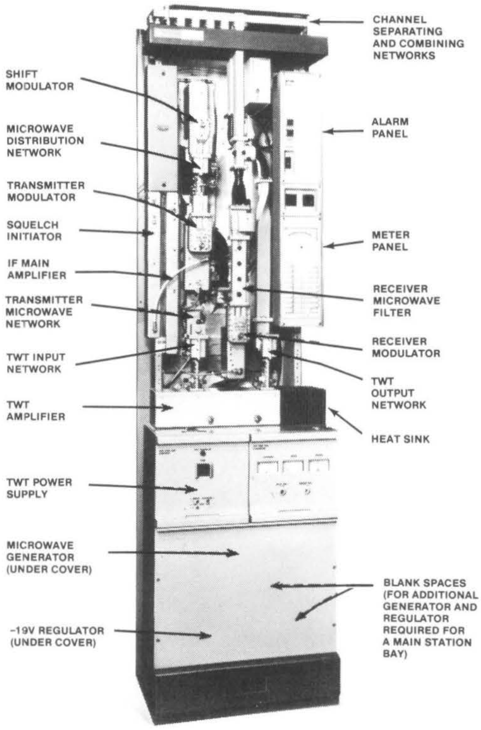


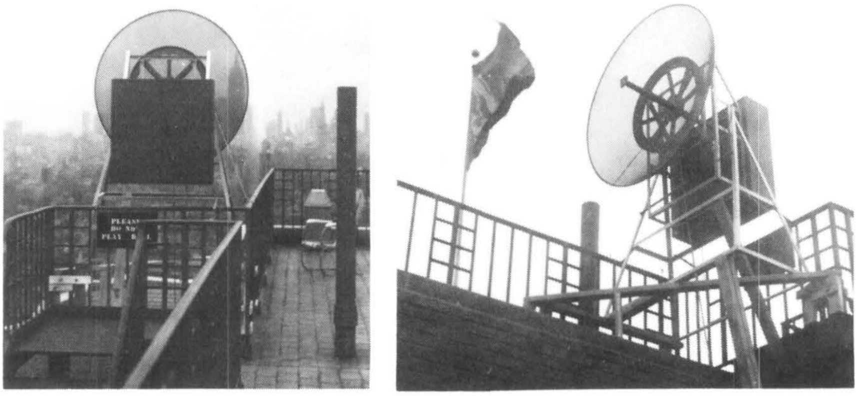
Fig. 11-15. TH3 transmitter-receiver bay (repeater bay), 1970.

improvements were introduced in TH3 as time passed. By adding an RF filter in the receiver and increasing the baseband drive in the transmitter to accommodate the observed lower talker volume on the network, it was determined by 1978 that TH3 could carry 2400 circuits. It was not until early 1983, however, that the FCC officially allowed the increased drive level.

III. SHORT-HAUL SYSTEMS

3.1 The First Short Haul System—TE1

The initial stimulus for the development of short-haul microwave radio-relay systems was television. Television frequently needed a temporary transmission link between the site of a newsworthy event and the local broadcast transmitter or an entry point to the larger network. Using technology from microwave radar and a modified military microwave system (AN/TRC-6) developed in World War II, designers created the 4-GHz TE1 system in 1946. (See another volume in this series, *National Service in War and Peace (1925–1975)*, Chapter 5, Section 6.)



(a) (b)

Fig. 11-16. The TE1 short-haul microwave link for television, 1946. (a) Receiving antenna on the roof of the AT&T Long Lines building at 32 Avenue of the Americas in New York City. From the transmitter on the roof of Yankee Stadium the beam passed down Fifth Avenue. (b) Front view of Long Lines antenna and receiver.

The TE1 transmitter used a low-power reflex-klystron frequency modulated by the baseband video signal. The transmitters and receivers were packaged in box-like containers, small enough to be called portable, and built to withstand adverse weather conditions [Fig. 11-16]. Later, an improved version of the system, TE2, was developed and first placed in commercial service by The Bell Telephone Company of Pennsylvania in June 1949, in a television studio-to-transmitter link.

3.2 The First 11-GHz System

Following the success and rapid growth of the first long-haul radio system, it appeared that a microwave system would also be attractive for short-haul applications, if it could be made low in cost. The envisioned applications were for side legs to main routes in areas of lower population density and television transmission off the major network. The expansion of network television and long-haul message transmission on radio was concentrated in the 4-GHz common-carrier band with some use of the 6-GHz band, starting in the late 1950s. A third common-carrier band existed at 10.7 to 11.7 GHz, which was unused in the early 1950s. These factors, plus pioneering research work in the field of components and system concepts for low-cost microwave systems, led to the development of the TJ system.[21]

RF transmission at 11 GHz posed a number of new problems. Not only was multipath fading more severe, but heavy rainfall in the path caused severe signal attenuation. Attenuation at 11 GHz was also a very significant factor if waveguide runs to antennas were used on high towers. Repeaters that brought the signal to baseband were desirable for short-haul systems to give access to the signal for frequent adding and dropping of circuits and for order-wire and alarm functions at repeater points. Reflex klystrons fitted the transmitter needs well, since they could easily be frequency modulated by the baseband signal. On the other hand, baseband repeaters required remodulation of the signal at every repeater point. Any nonlinearity in the relationship of baseband signal amplitude to RF frequency deviation generated intermodulation noise in the recovered signal, and the achievement of the required linearity was difficult with klystrons. One of the problems was that the RF frequency also depended on the impedance presented to the klystron, and this impedance could change rapidly with frequency, especially if there was a long waveguide run to a poorly matched antenna feed.

Meeting these problems had major effects on the design of the TJ system. The signal loss between the klystron and the antenna was minimized by using periscope-antenna systems with a minimum of waveguide. In a periscope (sometimes called a flyswatter), a parabolic dish close to the transmitter was directed vertically. The beam was then reflected horizontally toward the distant station by a 45-degree plane reflector at the top of the tower [Fig. 11-17].[22] The periscopes provided relatively poor back-to-back isolation between opposite directions of transmission, however; and it was necessary to use shifted frequencies (the so-called four-frequency plan) and crossed polarization for trans-

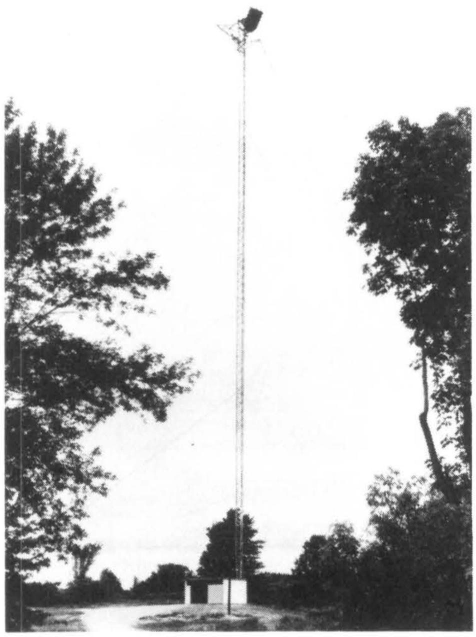


Fig. 11-17. TJ radio-relay repeater site at Oxford, Michigan, 1956. Vertically-directed parabolic dishes are on top of the repeater hut in a typical periscope arrangement.

mitters in the opposite directions. With 0.5 w of transmitter power, proper microwave-receiver design, moderate antenna sizes, and conservative hop lengths, large fading margins were achievable to accommodate attenuation caused by rain. A new device, the ferrite isolator, had the characteristic of low attenuation in one direction of transmission and high attenuation in the reverse direction. An isolator in the klystron output passed the outgoing signal with low loss but prevented wave reflections from the antenna or transmission line from affecting the impedance terminating the klystron. This made highly linear operation possible.

A plan for the use of the 11-GHz band was derived to use the spectrum as efficiently as possible while accommodating the special needs and characteristics of the short system.[23] The plan provided three 2-way channels of 40-MHz bandwidth, each protected for multipath fading or equipment failure by a protection channel on a one-for-one basis [Fig. 11-18]. The plan also took into account overreach interference between hops on a route, separation of working and protection channels to minimize simultaneous fading, cross polarization of the channels to ease filter requirements, and appropriate separation of transmitting and receiving bands to reduce interference and cost. The system originally was designed to carry up to 240 voice circuits or a color television signal for up to ten hops.[24]

The first application of TJ was between Phoenix and Flagstaff, Arizona. It

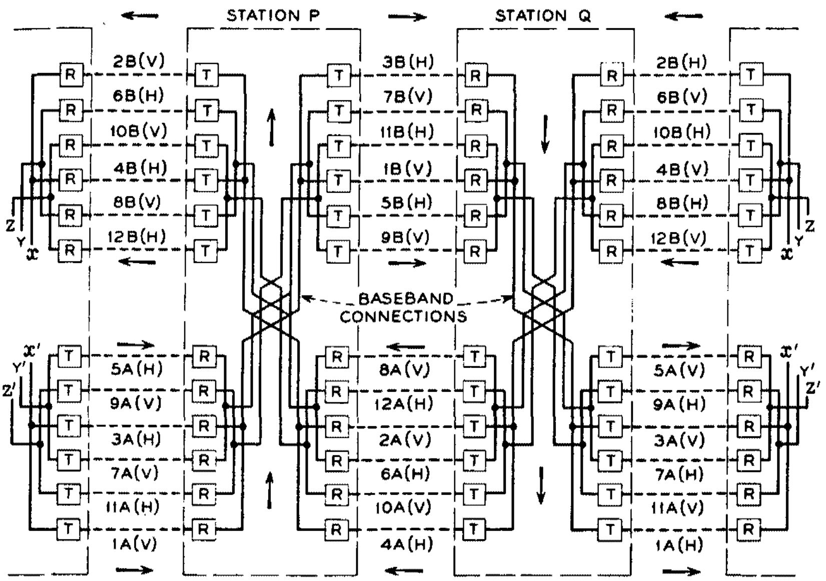


Fig. 11-18. The TJ frequency allocation plan. Three 40-MHz-wide two-way radio channels were derived in the 10.7 to 11.7 GHz common carrier band. An RF frequency shift with polarization reversed for each channel was necessary at each repeater. A protection channel was provided for each working channel on a one-for-one basis.

was a five-hop system, initially carrying 48 circuits and expected at that time to grow to over 300 circuits for the route. (Improvements in TJ made the larger load possible.) Almost 1500 TJ transmitter-receiver bays were produced by the mid-1960s and were configured in routes totaling over 5600 route miles.

3.3 The 11-GHz Pole-Line System

As the TJ system was being designed, the Radio Research Department continued to develop concepts for still more economical short-haul microwave systems.[25] They proposed a system that would use very short hops (five to ten miles), permitting antennas to be relatively low and mounted on poles or short towers (hence the designation pole-line radio). Such antenna structures would require little land and could be located along existing roads, which would provide easy access and readily available commercial power. Simplified low-power electronics for the transmitter and receiver would permit mounting a repeater in a cabinet rather than requiring a building and could use low-cost storage batteries for standby power.

Based on many of these concepts, TL radio development (later called TL1 when the succeeding TL2 system came about) was started in late 1959.[26,27] Using the 11-GHz frequency band, TL was designed to use many of the concepts of TJ, including its frequency plan, isolators, and short waveguide runs to obtain transmitter-klystron linearity. There were new problems, such as achieving adequate frequency stability with the wide temperature variations in unheated cabinets and keeping the total power required low enough so that a few storage batteries could provide adequate standby power.

By this date, diffused-base transistors of sufficient frequency range and life were available to permit building baseband and 70-MHz IF amplifiers of greatly reduced power and size.[28] With the exception of the klystrons for the transmitter and beating oscillator, this was the first Bell System solid-state microwave transmitter-receiver. A new klystron was designed for long life at 100-mw power output. The transmitter frequency was made stable over a wide temperature range by keeping the klystrons at the nearly constant temperature of a boiling liquid (a vapor-phase cooling system). Thus, two of the primary system objectives were achieved: highly reliable low-maintenance operation and low-cost standby power.

The reduction of engineering and installation costs was also an important objective of the development. For the outdoor repeater locations, all the equipment (except for the lead-acid storage batteries) was installed and tested in a cabinet at the factory. It was shipped assembled to the site, where it could be mounted and made ready for service with minimum installation effort [Fig. 11-19]. For larger routes or for other situations where a structure was necessary, the equipment was factory packaged and tested in various sizes of trailer-type prefabricated shelters and shipped to sites to be placed on prepared footings. A simplified and low-cost order-wire, alarm, and trouble-location system was designed to operate over the system itself, using signal frequencies below those used for message-circuit transmission. A simple one-by-one protection-switching system to guard against multipath fading and equipment failures was also designed, although it was expected that the short hops and the high reliability

Fig. 11-19. The antennas and radio cabinets of a TL repeater on Chink's Peak, Idaho, one of the first installations of TL radio.

of the equipment would make diversity protection unnecessary, and many unprotected systems were used.

The first TL installation was operative in September 1961 between Billings and Hardin, Montana. It was a four-hop nondiversity system carrying 32 circuits multiplexed by Type ON carrier terminals. Almost 700 transmitter-receiver units were built in the next two years to form about 2600 route miles.

3.4 The TM1 and TL2 Systems

The TL1 system was targeted for light telephone routes with cross sections under 100 circuits. Even during its initial use, however, the rapid increase in demand for television and message circuits made it apparent that the low-cost equipment would be of greater value if it could be redesigned to carry television and a greater number of message circuits. In addition, as the number of circuits was increased to several hundred, the reliability of the system, judged adequate for small cross sections, needed to be greater. While the 11-GHz band could have been used quite successfully in most areas, in some parts of the country, heavy rainfall would have attenuated the entire band below usable levels. This made it difficult to achieve the reliability desired for the heavier message cross sections and still have hop lengths long enough for the systems to be economical.

On the other hand, the 6-GHz band would be much less affected by rain attenuation. For the same outage time, a crossband diversity pair at 6 and 11 GHz could have longer path lengths than a 11/11-GHz pair in areas of heavy rainfall. The susceptibilities in the two bands were complementary. The 6-GHz system would not suffer from multipath fading under the conditions where the 11-GHz system was wiped out by rain, and the 11-GHz system would not be subject to rain attenuation when the 6-GHz system was encountering multipath fading, since the two weather-related phenomena do not occur at the same time. Therefore it was decided to undertake the development of two new short-haul systems, TL2 for the 11-GHz band and TM1 for the 6-GHz band, designed to be complementary in application and using as much common equipment as possible.[29]

The objectives for the TM1/TL2 radio pair were more ambitious than for TL1. They were intended to yield higher reliability, carry up to 600 circuits (from L multiplex), give better noise performance, and furnish television capability in the short-haul field. They were to provide at low cost what a short time before would have been considered a heavy-route radio facility.

The frequency plan for TL2 was the same as that for TL1 and TJ. There were three plans used for TM1: the channel plan was the same as the long-haul TH plan where TM1 and TH would be used on the same route, another plan split each TH channel into two channels, and the third plan was an arrangement with fewer channels staggered with respect to TH to minimize interference. Different antenna arrangements were also possible, yielding different numbers of radio channels. The maximum number was six 2-way crossband-diversity channels with TL2 when both horizontal and vertical RF polarizations could be used in both bands (as was possible with the horn-reflector antenna).

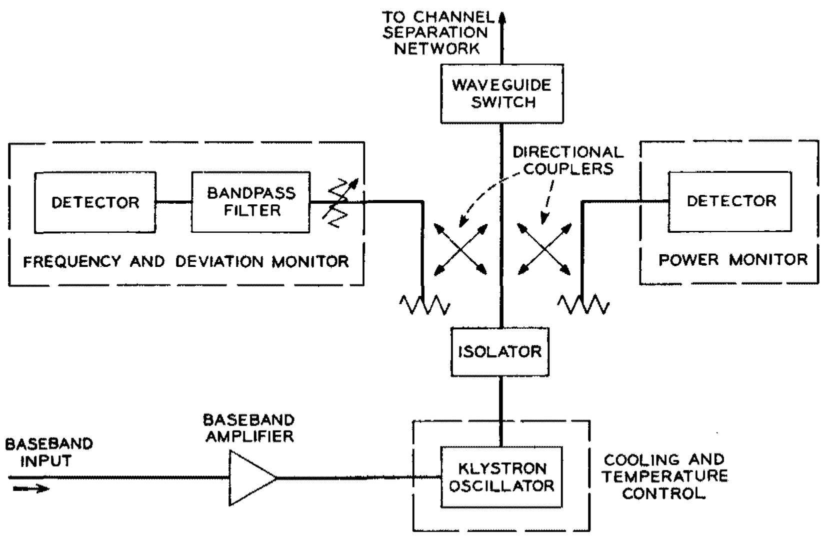


Fig. 11-20. The TL2/TM1 radio-transmitter block schematic.

The transmitter and receiver block diagrams in both bands were the same [Fig. 11-20]. The IF and receiving baseband circuitry were identical for the two systems. The RF waveguide, networks, and klystrons could not, of course, be common to both bands. One of the most difficult TM1 design problems was to obtain the 0.02-percent frequency stability required in the 6-GHz band, in contrast to the 0.05 percent that was adequate at 11 GHz. A new 6-GHz klystron was designed to be consistent with the TL1/TL2 klystron characteristics and was made suitable for operation with a vapor-phase cooling system. The new klystron had a specially designed temperature compensator, and the vapor-phase cooling system used for TL1 was redesigned for the closer temperature control needed. The klystron power supply was also precisely controlled and temperature compensated to meet the requirements.

The TL2 and TM1 systems were very similar in appearance with transmitters or receivers mounted into the same equivalent spaces. Six transmitters, six receivers, or three transmitter-receiver pairs could be mounted in one seven-foot bay [Fig. 11-21]. Both had transmitter power outputs of 100 mw. To permit "hop-stretching," a TWT amplifier was later designed to work with TM1 and boost the power to 2 w. When this amplifier was used, a one-by-one crossband-

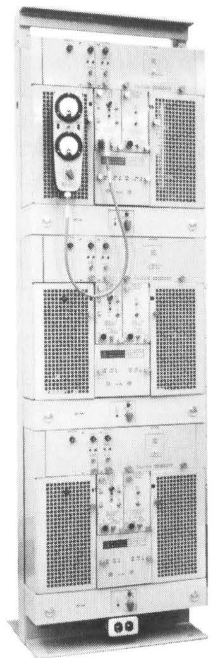


Fig. 11-21. The TL2/TM1 11-GHz/6-GHz crossband diversity system. A 7-foot bay could be equipped with three transmitter-receiver units and diversity switch panels.

diversity switch was designed so that the signal would revert to and stay on the higher-power system, except when it was being subjected to heavy fading or had an equipment failure. Other features—such as the order-wire and alarm system, equipment shelters, and maintenance philosophy—were much like those of TL1.

The first TM1/TL2 crossband-diversity system was installed between Charlottesville and Richmond, Virginia in 1964. During the next eight years, over 5500 TL2 and TM1 radio repeater units were manufactured, providing about 45,000 one-way radio channel miles for message and 2,000 one-way radio channel miles for television.

3.5 An Improved System for 6 GHz—TM2

Continuing demands for expansion of the message and television networks and advances in semiconductor technology led, in the early 1970s, to the development of a second generation of the TM system. A new semiconductor microwave device, the impact avalanche transit-time (IMPATT) diode, made it feasible to design a much lower-cost, all-solid-state radio transmitter requiring less primary power. Operating within the same frequency plan as TM1, the TM2 system was designed to carry 1200 message circuits with lower noise and better frequency stability. The improved frequency stability and antenna systems with better back-to-back isolation permitted operation with a two-frequency plan, that is, with transmitters on the same frequency in opposite directions, permitting more channels per route. The transmitter in one direction was vertically polarized, and the other was horizontally polarized.

As the routes became larger in cross section (greater number of radio channels and more circuits per channel), the spectrum became more and more crowded, and the FCC placed tighter restrictions on the number of radio-protection channels that could be used on a route. The simple one-by-one protection switching system used in earlier short-haul systems was no longer adequate, or permitted, in many applications. A new, more sophisticated, protection switching system (the 400B) was designed to operate with these new short-haul heavy cross-section systems. It permitted one protection channel to serve as many as seven working channels. In addition, TM2 provided options for hot standby equipment to guard against equipment failures and separated antennas (space diversity) to protect against multipath fading, where these arrangements were preferred to frequency-diversity protection.

The components of TM2 were almost all new designs. The utilization of the IMPATT diode was a significant technical challenge, since the IMPATT structures were basically oscillators. The transmitter amplifier was actually an oscillator, with a power output of 1 w, in which the frequency was locked to a lower-power frequency-modulated signal. The low-power FM microwave locking signal was obtained, first by varying the frequency of a 70-MHz oscillator in proportion to the voltage of the baseband signal and then converting up to the desired RF frequency. This required an RF oscillator (another IMPATT device) to drive the up-converter. The receiver also used an IMPATT oscillator to drive the down-converter and obtain the IF signal. The up-converter and down-converter were variations of the high-performance designs from the TH3 system. New wider-band transmitter and receiver baseband amplifiers were required to accommodate the larger number of channels. The design permitted four transmitter-receiver pairs to be mounted in a seven-foot bay.

The first TM2 system went into service between Toledo and Castalia, Ohio in 1973. In 1976, a further improved version of TM2 was introduced, called TM2A. With a new IMPATT power amplifier, capable of 1.6-w output, and an improved microwave generator, the capacity was increased to 1800 voice circuits. Thus, the metamorphosis of the short-haul system was nearly complete. From the concept of an unprotected single radio-channel pole line carrying fewer than 100 circuits, system designs evolved to routes of 8400 circuits with one-by-seven frequency-diversity switching. By 1975, short-haul radio was providing more than 20 million voice-circuit miles in the Bell System.

Although this story of short-haul systems has concentrated on radio systems designed by the staff of Bell Laboratories and manufactured by Western Electric, the field was always hotly competitive. Many systems made by other manufacturers were also being used by the Bell System companies for short-haul message and television purposes.

IV. SINGLE-SIDEBAND SYSTEMS

4.1 Early History

In a SSBAM system, the signal occupancy of the transmission channel is the same as the bandwidth of the baseband signal. For radio applications, this offers the opportunity for better utilization of the limited radio spectrum as well as attractive economics. This opportunity could not be realized for early radio systems, however, because of the stringent requirements on nonlinear distortion for the microwave transmitters and receivers.

The repeater-section path loss in radio systems is fundamentally different as a system parameter, compared to wire-line or coaxial-cable loss. In the guided systems, the loss in decibels is directly proportional to the length. If, at 10 MHz, 10 miles of coaxial cable has a loss of 40 dB, 20 miles will have 80 dB. Repeater spacing is therefore extremely critical. In radio, by contrast, if a 10-mile hop has 40 dB loss, a 20-mile hop will have only 46 dB. Conversely, a 5-mile hop reduces the loss only to 34 dB. Because of the inverse-square spreading of the radio beam, doubling or halving the spacing changes the loss by only 6 dB. As a consequence, the repeaters are usually spaced as far apart as other considerations (i.e., terrain clearance or fading) will permit. The repeater-section loss is high, typically about 60 dB, and the transmitter power output is pushed close to its limit, a condition in which an AM signal would suffer intolerable distortion. An FM signal is relatively insensitive to amplitude distortion, however, and can be transmitted through nonlinear amplifiers with little penalty. Whatever the initial doubts as to the best form of modulation for microwave radio, the advantages of FM became universally accepted and unquestioned as system capacity was raised from a few hundred to thousands of channels.

Fortunately, there were always a few people ready to question received wisdom and reexamine the assumed limitations. In 1966, aware of the anticipated frequency congestion in the decade ahead, A. J. Giger, of the Bell Laboratories transmission development group, analyzed the requirements and proposed that work be started on a 6-GHz SSBAM system that could transmit 1200 voice channels. Giger proposed using a TWT with 35-w maximum (saturated) power and other improved circuits in the radio transmitter and receiver to achieve a total intermodulation noise to meet long-haul system objectives. Adequate linearity in the TWT was to be achieved by operating well below its maximum power capacity. He further recommended that additional development work be done on linearization techniques to improve the microwave transmitters by at least 20 dB. This would permit expanding the initial system capacity to 3600 voice circuits.

The linearization technique that appeared to offer the most promise was the feedforward distortion-cancellation scheme, based on the 1925 invention by H. S. Black that preceded his discovery of negative feedback.[30] (See Chapter 4, Section I.) Feedforward was an especially interesting prospect for linearizing a microwave transmitter because a TWT is, by design, a slow-wave device. It has a large electrical delay and phase shift. With feedforward, the linearity might be improved without the serious stability problems associated with closed-loop negative feedback. Moreover, the circuit would not suffer the reduction in gain inherent in a feedback amplifier.

In 1966, H. Seidel, of the device development group in Bell Laboratories, had proposed a feedforward circuit with features that appeared attractive for applications in the microwave range. Seidel's first investigations were with lower-frequency amplifiers intended for use in coaxial-cable systems.[31] There had been a continuing interest in the use of feedforward for cable systems from the 1950s onward, with particular emphasis on submarine cable systems, because of a redundancy feature provided by the use of two amplifiers in the circuit.[32,33] In 1969, new studies indicated that at least a 40-dB improvement in linearity would be required for a TWT in an AM system, compared to the TWT in the TD3 radio transmitter. This prompted further work with feedforward; and, in 1970, Seidel demonstrated in the laboratory a reduction of greater than 40 dB in nonlinear-distortion noise of a TD3 No. 461A-TWT amplifier.[34] This demonstration provided assurance that a transmitter meeting the linearity requirements was at least feasible. AT&T then initiated a focused, experimental study program to address all of the issues involved in the use of single sideband in the 4-GHz long-haul radio network.

4.2 Single-Sideband Amplitude Modulation—The Promise and the Challenge

The advantage in spectrum efficiency of SSBAM over FM was substantial. As noted earlier, the minimum bandwidth required for FM transmission is twice the sum of the top baseband signal frequency and the peak-frequency deviation, that is: the bandwidth required = $2(f_t + \Delta F)$ where f_t = top baseband signal frequency and ΔF = peak frequency-deviation. For the same transmitter power, and with the appropriate preemphasis of the baseband signal applied to the FM modulator, the signal-to-noise ratio of the demodulated FM signal can be made equal to that of a SSBAM signal when $\Delta F = f_t$. Thus, ideally, the FM radio-channel bandwidth must be four times the baseband signal bandwidth, whereas, with SSBAM, the two are equal. Wherever maximum exploitation of available bandwidth and power is desired, SSBAM is the method of choice, provided it is technically feasible.

In practice, as the channel load and baseband width had been increased, FM radio systems in the Bell System network had peak frequency-deviations considerably less than the top signal frequency, so that the use of SSBAM was expected to provide about a three-to-one advantage in spectrum occupancy. This implied that SSBAM also could have a signal-to-thermal-noise advantage, if the peak transmitted power indeed could be made equal to that of an FM system. However, as noted, at the required power level the TWT amplifiers used in FM systems in the late 1960s fell far short of the linearity required to achieve the three-to-one circuit advantage. The nonlinear distortion could be improved by dropping the signal levels; but, at the normal operating level intermodulation and thermal noise were about equal; that is, the systems were in balance with respect to the two noise sources. Any reduction in signal level would quickly produce excessive thermal noise. With SSBAM, the required power for adequately low thermal noise was about the same as that for the FM transmitters. And, at that power level, intermodulation products in a long system would be 35 to 40 dB above thermal noise and equally far above requirements.

There were, in addition, several other nontrivial problems. At the outset of the experimental-study program, it was recognized that a propagation loss that varied across the radio band because of multipath fading would have a much greater effect on SSBAM than on FM transmission. In an FM system, changes in signal level from fading change the signal-to-noise ratio but not the telephone channel loss. Moderate signal-to-noise changes, especially if they are of only short duration, are quite tolerable. By contrast, in an AM system, any uneven change in level from fading will result in a corresponding change in end-to-end channel loss unless it is offset. In the direct-dialing network, this loss must be held within very close limits. Consequently, dynamic amplitude equalization of the radio channel was necessary to achieve satisfactory transmission.

It also was recognized that special frequency-correcting features would be necessary because of the frequency errors introduced by tandem radio repeaters. In an IF-type radio repeater, the repeated frequency translations from RF to IF and back would inevitably result in a shift of many hertz in the recovered baseband AM signal. This shift had to be reduced to within 1 Hz of the original signal frequency to avoid significant degradation of received speech signals. It was planned to accomplish this by transmitting a tone of accurately-known frequency along with the multiplexed speech channels. Restoration of this tone to its correct frequency at the receiving terminal by shifting the entire band would correct all channels. Finally, it was essential that the SSBAM system be capable of introduction into the existing FM network, which meant living with the interference caused by FM transmitters on the same route and at route-junction stations.

The choice of 4 GHz rather than 6 GHz for the early SSBAM work was made for a number of reasons. Because of the longer wavelength, dimensional tolerances would be less critical for the new microwave components. Multipath fading occurrences would be less, and there would be a narrower-bandwidth radio channel to equalize. Perhaps most importantly, the 4-GHz band was used exclusively for common-carrier systems. It was well developed, and its use was highly coordinated by relatively few users. The interface environment could be characterized with confidence. In contrast, the 6-GHz band was used by a variety of entities, in a much less tightly coordinated way. As it turned out, this decision was less than optimal, and some time and effort were lost when the project was later shifted to 6 GHz. The system feasibility was demonstrated, however, and essentially all the critical problems were solved first in the 4-GHz band.

4.3 Early System Studies

There was a compelling economic advantage in designing a SSBAM system that could be placed in the deployed FM network, because existing facilities—such as antennas, towers, waveguide feeders, and buildings—could be used. There was no question of building an AM system from scratch. This meant radio path lengths and antenna gains would be essentially the same as for the FM systems. Since the FM sideband power was known, the antenna discrimination characteristics could be used to estimate interference noise levels for use in a noise-allocation budget that included thermal, intermodulation, and interference noise contributions. A linearity requirement could be determined for the radio repeaters that assured meeting the long-haul noise objective of 40 dBrnC0 for a 4000-mile connection.

A noise-allocation tree was devised during the early exploratory work [Fig. 11-22]. Because of the systematic in-phase addition of third-order intermodulation products, the intermodulation noise objective for a single repeater had to be 16.8 dB below the thermal noise contribution. This result was based on

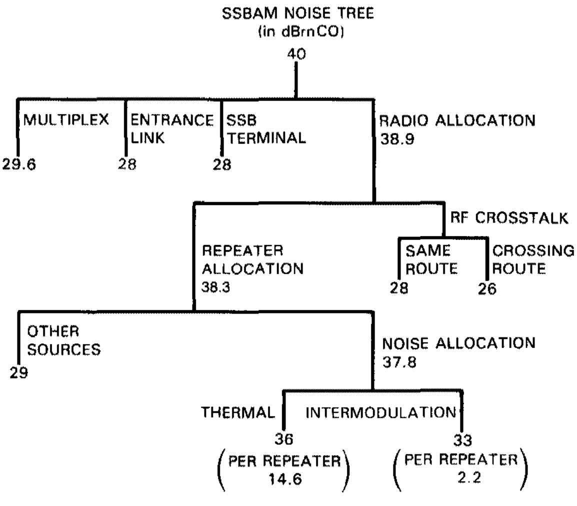


Fig. 11-22. Single-sideband AM radio noise-allocation tree. Noise objectives for each contribution are in dBrnC0 for a 40 dBrnC0 total.

the assumption that the intermodulation noise added coherently from repeater to repeater within a terminal section and that, at terminals, the message spectrum was rearranged (frequency frogged) to decorrelate the noise in successive sections. Considerable attention was given to understanding the mechanisms controlling the accumulation of intermodulation noise and investigating methods for reducing the correlation. One of the problems was to measure intermodulation noise in a single repeater where it was actually below the thermal-noise level. A special pseudorandom noise-loading test set was designed that provided accurate measurements of the intermodulation noise at normal operating conditions in the presence of the higher thermal noise.[35]

A detailed study of the TD3A FM repeater revealed that not only the TWT but almost every other component as well would require substantial improvement in linearity. The FM repeater had not been designed with low amplitude distortion as a requirement. But, since the requirements for AM were so much more stringent than the actual performance, it was evident that considerable development work would be necessary to improve the repeater and the transmitter would be a particularly difficult problem. The demonstration in 1970 of 40-dB improvement in TWT linearity using feedforward indicated that the required amplifier performance could be achieved in the laboratory, but there were important problems remaining concerning circuit complexity, stability, cost, and practicality for field use. The two signal loops in the feedforward circuit would have to maintain their alignment to within 0.6 dB in amplitude and 0.4 degree in phase over the system operating-temperature range and for extended periods of time.

If the improved linearity could be achieved, the achievable voice-circuit capacity of the 20-MHz radio channel would then be limited primarily by the expected interference from the TD2 and TD3 FM transmitters in the network.

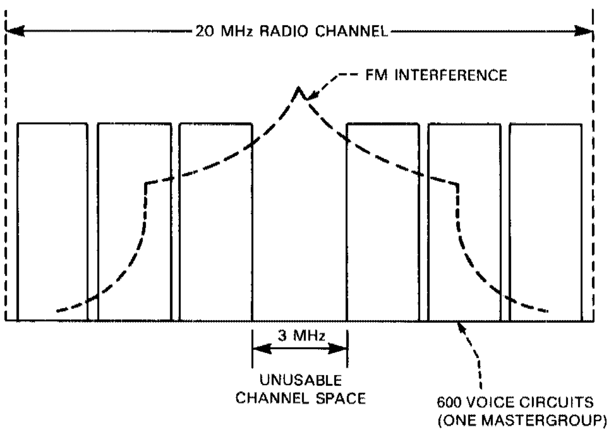


Fig. 11-23. Exploratory single-sideband AM frequency plan. Six mastergroups (3600 channels) could be transmitted in the 20-MHz radio channels assigned in the 4-GHz band. The center frequencies were unusable because of high interference from FM carriers.

This interference at a radio junction would be extremely high at the frequency of the FM carriers in the center of each band, tapering off with frequency at each side of the carrier. About 3 MHz of the channel spectrum was not usable, but 3600 voice circuits could be placed in the remaining parts of the 20-MHz/4-GHz channels using tightly packed mastergroups [Fig. 11-23].

The allowable amplitude and phase fluctuations of a telephone channel under the dynamic conditions of multipath fading were investigated. Subjective testing showed that listeners were not much disturbed by fairly substantial fluctuations, but some data sets under worst-case conditions would produce errors for dynamic variations of about ±3 dB in gain and ±30 degrees in phase. The equalization objective was set at ±2 dB and ±20 degrees for any circuit within the 3600-channel load to provide some margin. Later in the program, it was further established that this performance should be provided at least 99.9 percent of the time.

4.4 Repeater-Linearity Improvement

During the early work on the radio transmitter, the frequency up-converter was included in the feedforward circuit, since a linearity improvement of about 40 dB was required in this unit, as well as in the TWT. This increased the complexity of the feedforward circuit because two additional frequency converters were required [Fig. 11-24]. The linearization desired was from the input to the up-converter at IF to the TWT output at RF. It was necessary to attenuate and convert the output back to IF to derive the error signal (the distortion) by subtraction of the input. It was then necessary to convert the error back to RF and amplify it to combine it with and correct the output signal.

Other factors further added to the complexity. Automatic control circuits were necessary to maintain the two loops in proper amplitude and phase alignment because of expected transmission variations due to changes in temperature. Much of the early program was concentrated on reducing the feedforward improvement required to a point where the automatic control circuits could be

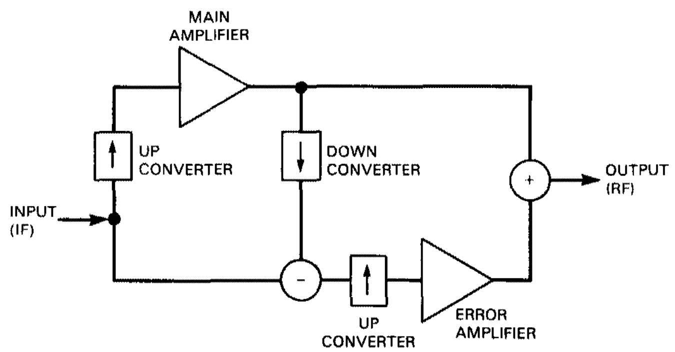


Fig. 11-24. Experimental feedforward circuit operating from IF (70 MHz) to RF (4 GHz). A down-converter was necessary to compare output and input and to derive the error signal at IF, and an up-converter and error amplifier were required for the distortion correction at RF.

discarded. Methods were studied for improving the inherent linearity of the TWT and the frequency converter. It was observed that, while these components fell far short of the needed performance, the nonlinearity was quite stable suggesting that a compensating nonlinearity or predistortion might be introduced to offset their characteristic.

A number of commercially available TWTs were characterized that had higher gain and higher output-power capabilities than the TWTs used in the Western Electric FM transmitters. As expected, these tubes had lower nonlinear distortion at a given power level than the FM system TWTs. Moreover, their higher gain lowered the required level at their input, alleviating the linearity requirement on the transmitter up-converter. High local-oscillator drive power was also used to reduce the converter intermodulation noise. When the high drive power was used within a feedforward loop that included a predistortion circuit at IF, the results were very gratifying. The needed intermodulation improvement was reduced to about 25 dB, and it became possible to eliminate the automatic loop-balance controls. A laboratory model of this transmitter demonstrated very stable operation and met system linearity requirements for a period of several months without adjustments.

At this time, the primary concern was the cost of the transmitter, because it still consisted of two TWT amplifiers, three frequency converters, a predistortion circuit, and delay equalization networks. The application of predistortion was studied with renewed vigor. Because it had been so successful, there was a prospect of using it alone, although both the TWT and frequency converter would have to be further improved by a considerable amount if feedforward was to be eliminated.

The most troublesome intermodulation products were the third-order ones produced by the cubic term in the transmitter's transfer characteristic. The predistortion introduced, therefore, was by means of a solid state "cuber" at IF. The cuber generated third-order distortion controllable in amplitude and phase, and that, in principle, could be adjusted to offset the distortion from the up-converter and TWT [Fig. 11-25].[36] But the early work with predistortion yielded only a moderate improvement in the TWT distortion, and the improvement could be maintained over only a small range of power output. The problem was that the TWT distortion was due to fifth- and higher-order distortion terms as well as to third-order effects. When the development group discussed this problem with C. C. Cutler, one of the original Bell Laboratories TWT designers, he immediately deduced that electrons in the collector region were being scattered in the backward direction, forming a weak reverse beam that was saturating at low power levels. A small external magnet at the collector end of the TWT eliminated this reverse beam, and immediately predistortion improvements of 30 dB or more could be obtained for the transmitter over a large output-power range [Fig. 11-26].

This unexpected result presented a dilemma, because it created pressure to terminate work on feedforward before complete confidence could be established that predistortion alone was viable under field conditions. In November 1972, however, the leaders of the project, W. C. Jakes and R. E. Markle, decided to terminate all work on feedforward and set a goal of designing a predistortion circuit that could provide 35 dB of improvement in the laboratory and assure

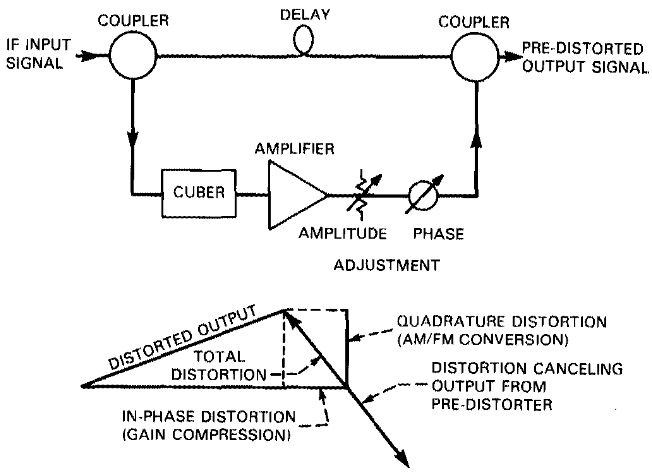


Fig. 11-25. The IF predistorter using a cuber to generate third-order distortion controllable in amplitude and phase.

at least 25 dB for three months without adjustment under field conditions. A laboratory repeater without feedforward was assembled late in 1972 for further study and experimentation [Fig. 11-27].

And thus ended the amazing story of feedforward and microwave AM radio. The feedforward concept was an essential element in the process. Without the prospect of the large improvement in linearity offered by feedforward, the project would not have been undertaken. It led, however, to extremely complex and fussy circuits that were not really practical. Finally, the efforts to reduce the improvement in linearity that was demanded of feedforward were so suc-

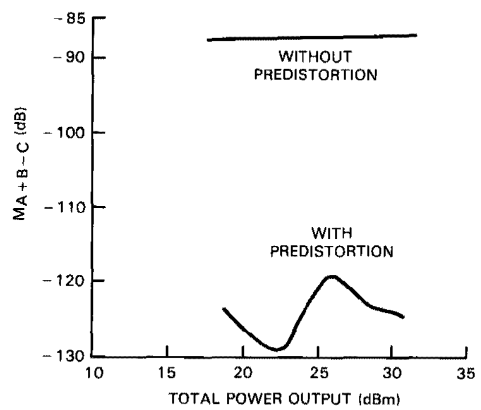


Fig. 11-26. The improvement achieved in critical third-order modulation of an up-converter and traveling-wave tube by predistortion.

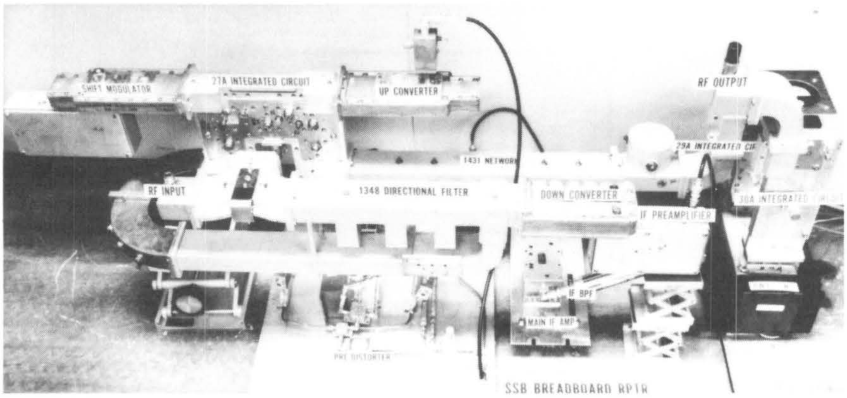


Fig. 11-27. Laboratory breadboard AM microwave repeater using predistortion only, 1972.

cessful that it was dispensed with altogether. It entered the project only as a catalyst but, like other catalysts, it was essential and served an extremely useful purpose. It advanced the realization of SSBAM radio by many years.

4.5 Combatting Radio-Propagation Effects

Multipath fading, rain attenuation, and scintillation produce variations in received signal strength that must be offset by automatic gain control (AGC) to avoid excessive level misalignments in all microwave systems. With SSBAM, the signal power was directly affected by talker activity, so AGC had to be based on monitoring some component of the signal that was always present, such as a deliberately inserted pilot. However, AGC provided only a flat gain adjustment, and dynamic equalization was essential to reduce the amplitude shapes across the channel caused by frequency-selective fading. In 1970, the radio path characterizations used for FM were not adequate for SSBAM, and these had to be extended by an extensive program of experimental work to devise an adequate equalization plan.[37,38] At first, as many as five pilots were considered necessary, but further tests and analyses in 1973 and 1974 showed that only linear and parabolic shapes would be necessary to meet the equalization objective, if space-diversity reception were employed on the radio hops with the worst fading.[39] The three radio pilots necessary to control this equalizer were also used for the AGC to provide better control of power levels. The AGC action was based on maintaining a constant average for the detected pilot powers at the receiver output, while upper and lower pilots in the channel were compared with the center pilot and the differences used to control bump equalizers in each half of the channel [Fig. 11-28].[40] It was later determined that the shape equalizers were needed only in protection-switching stations (i.e., about every four or five radio hops). Thus, by the end of 1975, the system-equalization plan had been considerably simplified, along with a corresponding reduction in cost.

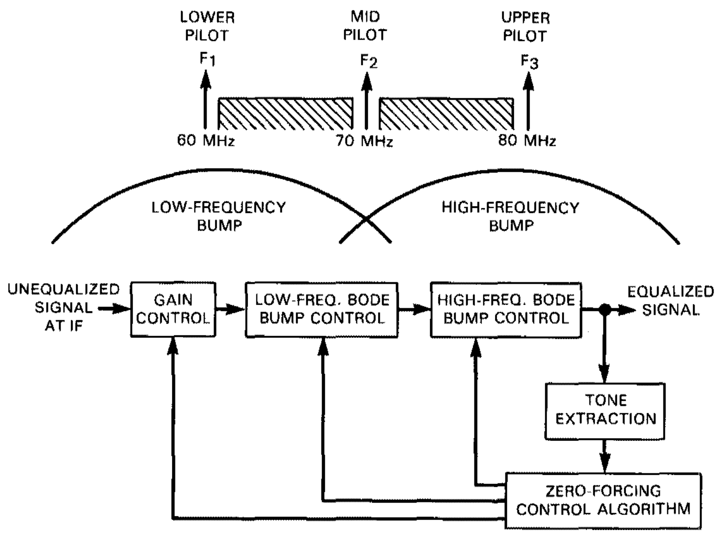


Fig. 11-28. Single-sideband radio channel amplitude equalization. Three pilots and two bump shapes were adequate for overall gain control and amplitude deviations across the band.

4.6 System Trials and 6-GHz Development

By the end of 1973, the work on linearization and equalization had proceeded so well that an extensive field test was planned to demonstrate both feasibility and practicality and to gain some initial experience with SSBAM in the existing network. The work up to this point had all been carried on at 4 GHz. In November 1973, as a result of further study of the existing situation at 6 GHz, a decision was made to design the first commercial system for the higher band. It had been found that many more frequency-coordinated routes were available than had been originally estimated, and the wider band (30 versus 20 MHz) would make more channels available at correspondingly lower costs. Nevertheless, it was decided to carry out the field trial with the 4-GHz repeaters for which complete designs were available as the results obtained were expected to be readily translated to the higher band.

A four-hop field test using brassboard repeaters was carried out in Georgia from mid-1974 through mid-1975 to obtain transmission data in a realistic radio-station environment. Operation and general feasibility were successfully demonstrated and a number of important issues investigated. These included the addition of intermodulation products,[41] the tolerance to transmission misalignment, the dynamics of the equalization control-loop response, and the adequacy of the baseband frequency correction circuit. The information obtained was then used to guide the design of the 6-GHz system, designated AR6A (AM radio, 6 GHz, A version), that was initiated early in 1975.

During this same time period, a potential intelligible-crosstalk problem was

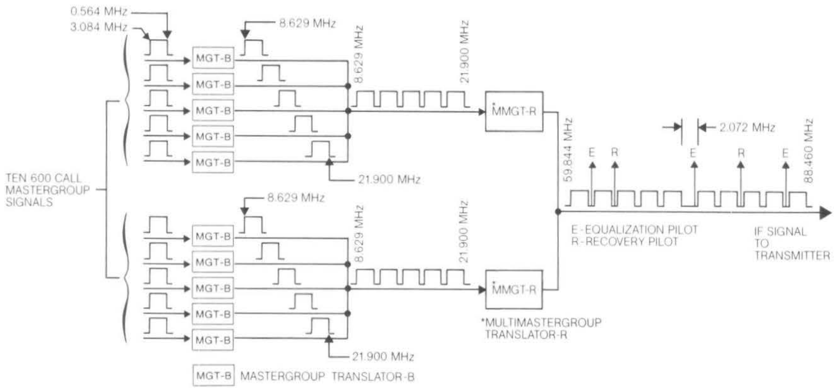


Fig. 11-29. Multiplexing for 6-GHz single-sideband AM radio, AR6A. The mastergroups were first translated to a five-mastergroup block from 8.6 to 21.9 MHz. These blocks were then translated to the upper and lower halves of the 30-MHz IF band with a 2-MHz gap left at the center to avoid interference from FM carriers.

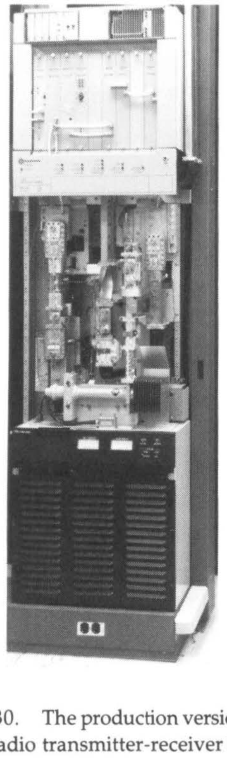


Fig. 11-30. The production version of the AR6A radio transmitter-receiver bay.

solved by phase locking all SSBAM microwave sources in a radio station to a single reference oscillator. The reference oscillator was needed in any case to provide the frequency stability required to pick off pilots in narrow-band filters and to restrict the frequency-tracking range required for the final voice-frequency correction in the terminal.[42,43]

Work on the AR6A proceeded rapidly, even while the trial was in progress. The 6-GHz protection-switching system was modified for SSBAM and new multiplexing equipment was designed to stack the mastergroups into the radio channel. Ten mastergroups were transmitted in the 30-MHz band by closer spacing than originally planned and by leaving only a 2-MHz gap at the center [Fig. 11-29].[44,45,46]

Early in the development program, a basic decision was made not to jeopardize the system success or schedule by introducing any more new technology into the design of the radio units than was required by the new radio band. The microwave circuits were based on those developed for TH3, while the IF circuits were based on both L5 coaxial technology and FM radio-design experience. The resulting physical design of the radio bay was similar in appearance to TH3 [Fig. 11-30].

A new IF protection-switching system, the 500A, was designed for AR6A; it was controlled by microprocessors and used three pilots and noise slots within a channel to determine when switching should be initiated. The new protection system provided access and remote-testing capabilities that were not available in previous radio systems.[47] These remote transmission-performance tests were conducted under the control of a central minicomputer that served a large radio region. A dedicated test set associated with the receiving protection-switch equipment responded to the central computer orders and made the necessary measurements. This system was designated the Transmission Surveillance System—Radio (TSS-R).

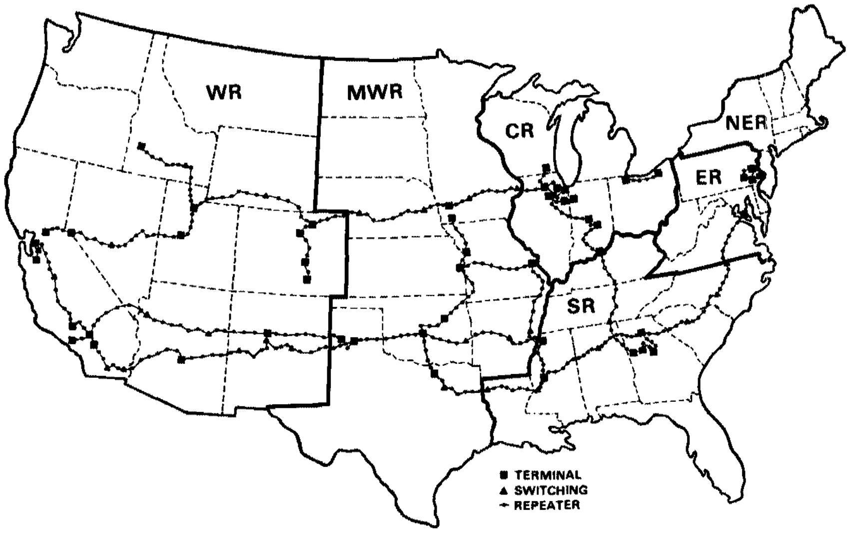


Fig. 11-31. AR6A single-sideband AM microwave radio routes, 1983. The letters indicate AT&T Long Lines Department operating regions.

The development was completed with the design of a portable test set for adjusting the predistorter in the field.[48] This set enabled predistortion improvements of 30 dB to be easily achieved. A complete AR6A system was installed on six radio hops in Missouri and field tested from mid-1979 through mid-1980. This route was extended by three radio hops in late 1980 to connect Hillsboro, Missouri to LaCygne, Kansas, and became the first SSBAM long-haul system to be placed in commercial service on January 12, 1981. By the end of 1983, 3500 radio bays had been installed in the United States to expand the capacity of radio routes coast to coast [Fig. 11-31].[49] The AR6A system became the principal means for expanding radio capacity in the 1980s.

REFERENCES

1. J. A. Morton and R. M. Ryder, "Design Factors of the Bell Telephone Laboratories 1553 Triode," *Bell Syst. Tech. J.* **29** (October 1950), pp. 496-530.
2. A. E. Bowen and W. W. Mumford, "A New Microwave Triode: Its Performance as a Modulator and as an Amplifier," *Bell Syst. Tech. J.* **29** (October 1950), pp. 531-552.
3. A. A. Roetken, K. D. Smith, and R. W. Friis, "The TD-2 Microwave Radio Relay System," *Bell Syst. Tech. J.* **30** (October 1951), pp. 1041-1077.
4. A. C. Dickieson, "The TD-2 Story: From Research to Field Trial," *Bell Lab. Rec.* **45** (October 1967), pp. 282-289;
   A. C. Dickieson, "The TD2 Story: Vacuum Tubes and Systems Engineering," *Bell Lab. Rec.* **45** (November 1967), pp. 324–331; and
   A. C. Dickieson, "The TD2 Story: Changing for the Future," *Bell Lab. Rec.* **45** (December 1967), pp. 356–363.
5. I. Welber, H. W. Evans, and G. A. Pullis, "Protection of Service in the TD-2 Radio Relay System by Automatic Channel Switching," *Bell Syst. Tech. J.* **34** (May 1955), pp. 473-510.
6. E. T. Harkless, "A Network for Combining Radio Systems at 4, 6 and 11 kmc," *Bell Syst. Tech. J.* **38** (September 1959), pp. 1253-1267.
7. H. E. Curtis, T. R. D. Collins, and B. C. Jamison, "Interstitial Channels for Doubling TD-2 Radio System Capacity," *Bell Syst. Tech. J.* **39** (November 1960), pp. 1501-1527.
8. H. D. Griffiths and J. Nedelka, "100A Protection Switching System," *Bell Syst. Tech. J.* **44** (December 1965), pp. 2295-2336.
9. H. T. Friis, "Microwave Repeater Research," *Bell Syst. Tech. J.* **27** (April 1948), pp. 183-246.
10. W. E. Kock, "Metal-Lens Antennas," *Proc. IRE* **34** (November 1946), pp. 828-836.
11. A. H. Lince, "Antennas for the TD-2," *Bell Lab. Rec.* **30** (February 1952), pp. 49-55.
12. R. W. Friis and A. S. May, "A New Broad-Band Microwave Antenna System," *Trans. Amer. Inst. Elec. Eng.—Pt. I. Commun. Electron.* **77** (March 1958), pp. 97-100.
13. A. B. Crawford, D. C. Hogg, and L. E. Hunt, "A Horn-Reflector Antenna for Space Communication," *Bell Syst. Tech. J.* **40** (July 1961), pp. 1095-1116.
14. See special issue on the TD-3 Microwave Radio Relay System, *Bell Syst. Tech. J.* **47** (September 1968), pp. 1143-1530.
15. U. S. Berger, "TD2: Eighteen Years and Still Growing," *Bell Lab. Rec.* **46** (July/August 1968), pp. 210-216.
16. U. S. Berger, "The Old TD2 Becomes the New TD2C," *Bell Lab. Rec.* **51** (October 1973), pp. 278-284.
17. See reference 16.
18. W. C. Ahern, F. P. Duffy, and J. A. Maher, "Speech Signal Power in the Switched Message Network," *Bell Syst. Tech. J.* **57** (September 1978), pp. 2695-2726.
19. See special issue on the TH Microwave Radio Relay System, *Bell Syst. Tech. J.* **40** (November 1961), pp. 1459-1743.
20. See special issue on TH-3 Microwave Radio System, *Bell Syst. Tech. J.* **50** (September 1971), pp. 2085-2353.

21. L. C. Tillotson, "A Short-Haul Microwave Transmitter," *Bell Lab Rec.* **33** (April 1955), pp. 131-134.
22. W. C. Jakes, Jr., "Single-Surface Reflectors Used for Microwave Relay Systems," in *Antenna Engineering Handbook*, ed. H. Jasik (New York: McGraw-Hill, 1961), pp. 13-1–13-8.
23. S. D. Hathaway and H. H. Haas, "The TJ Radio System," *Bell Lab. Rec.* **37** (April 1959), pp. 129-133.
24. J. Gammie and S. D. Hathaway, "The TJ Radio Relay System," *Bell Syst. Tech. J.* **39** (July 1960), pp. 821-877.
25. C. L. Ruthroff and L. C. Tillotson, "An Experimental 'Short-Hop' Microwave System," *Bell Lab. Rec.* **38** (June 1960), pp. 202-206.
26. U. S. Berger, "The TL Microwave System," *Bell Lab. Rec.* **40** (February 1962), pp. 44-50.
27. S. D. Hathaway, D. D. Sagaser, and J. A. Word, "The TL Radio Relay System," *Bell Syst. Tech. J.* **42** (September 1963), pp. 2297-2353.
28. W. E. Ballentine, V. R. Saari, and F. J. Witt, "The Solid-State Receiver in the TL Radio System," *Bell Syst. Tech. J.* **41** (November 1962), pp. 1831-1863.
29. R. W. Friis, J. J. Jansen, R. M. Jensen, and H. T. King, "The TM-1/TL-2 Short Haul Microwave Systems," *Bell Syst. Tech. J.* **45** (January 1966), pp. 1-95.
30. H. S. Black, U.S. Patent No. 1,686,792; filed February 3, 1925; issued October 9, 1928.
31. H. Seidel, H. R. Beurrier, and A. N. Friedman, "Error-Controlled High Power Linear Amplifiers at VHF," *Bell Syst. Tech. J.* **47** (May/June 1968), pp. 651-722.
32. B. McMillan, U.S. Patent No. 2,748,201; filed September 21, 1951; issued May 29, 1956.
33. R. W. Ketchledge, U.S. Patent No. 2,751,442; filed September 26, 1952; issued June 19, 1956.
34. H. Seidel, "A Microwave Feed-Forward Experiment," *Bell Syst. Tech. J.* **50** (November 1971), pp. 2879-2916.
35. R. B. Swerdlow, "Pseudorandom Noise Loading for System Evaluation," Panel Session on Microwave Noise Measurements and System Effects, *1973 IEEE G-MTT International Microwave Symposium, Digest of Technical Papers*, Boulder, CO (June 4-6, 1973), p. 227.
36. R. I. Felsberg and H. Miedema, U.S. Patent No. 3,825,843; filed June 8, 1973; issued July 23, 1974.
37. G. M. Babler, "A Study of Frequency Selective Fading for a Microwave Line-of-Sight Narrowband Radio Channel," *Bell Syst. Tech. J.* **51** (March 1972), pp. 731-757.
38. G. M. Babler, "Selectively Faded Nondiversity and Space Diversity Narrowband Microwave Radio Channels," *Bell Syst. Tech. J.* **52** (February 1973), pp. 239-261.
39. A. Vigants, "Space Diversity Engineering," *Bell Syst. Tech. J.* **54** (January 1975), pp. 103-142.
40. G. J. Mandeville and D. I. McLean, U.S. Patent No. 4,003,006; filed October 6, 1975; issued January 11, 1977.
41. R. H. Moseley, S. B. Pirkau, and R. B. Swerdlow, "Pseudorandom Noise Loading for Noise Power Ratio and Law of Addition," *Bell Syst. Tech. J.* **56** (April 1977), pp. 511-533.
42. R. B. Swerdlow, U.S. Patent No. 4,019,140; filed October 24, 1975; issued April 19, 1977.
43. M. Emshwiller, U.S. Patent No. 4,029,968; filed May 11, 1976; issued August 2, 1977.
44. R. E. Markle, "The AR6A Single Sideband Long-Haul Radio System," *International Conference on Communications* **3** Chicago, IL (June 12-15, 1977), pp. 78-82.
45. R. E. Markle, "Single Sideband Triples Microwave Radio Route Capacity," *Bell Lab. Rec.* **56** (April 1978), pp. 104-110.
46. M. M. Luniewicz, "MGTB—A Closely Spaced 600 Channel Translator," *International Conference on Communications* **2**, Chicago, IL (June 12-15, 1977), pp. 17-21.
47. J. Gammie and J. P. Moffatt, "The AR6A Single-Sideband Radio System," *International Conference on Communications* **1**, Denver, CO (June 14-17, 1981), pp. 3.5.1-3.5.7.
48. R. P. Hecken, U.S. Patent No. 4,274,047; filed December 7, 1979; issued June 16, 1981.
49. See special issue on the AR6A Single-Sideband Microwave Radio System, *Bell Syst. Tech. J.* **62** (December 1983), pp. 3247-3489.

Chapter 12

Submarine Cable Systems

I. THE BEGINNING

In 1858, only 23 years after Samuel Morse built his first telegraph apparatus, the first electrical communication between North America and Great Britain was accomplished by an undersea telegraph cable. Although Alexander Graham Bell had invented the telephone in 1876, it was not until 1956, 80 years later, that voice transmission by cable was achieved between the same points. The reason for this long delay was that a satisfactory voice signal was much more difficult to transmit over the transatlantic distance than a telegraph signal.

The first telegraph cables were painfully slow, capable of transmitting only two or three words per minute, equivalent to a bandwidth of only 1 or 1.5 Hz. Many improvements increased transmission speed over the years, and, in 1924, continuous loading with permalloy increased the transmission rate to about 400 words per minute, four or five times faster than the fastest unloaded cables. This was far faster than the fastest manual operator, and, at this stage, multiple operators prepared messages on tape to feed to mechanical transmitters. Even so, the useful band hardly exceeded 100 Hz. By proper design, continuous loading could also reduce the loss over the voice band. In 1921, three continuously loaded telephone cables were installed between Key West, Florida and Havana, Cuba, a distance of about 100 miles. Each of these was capable of transmitting a single two-way telephone channel, plus three carrier telegraph channels. At that early date, however, a transatlantic cable with sufficiently low loss was beyond the reach of the technology.

The first transatlantic telephone circuits were established by low-frequency radio (about 60 kHz) in 1927 and at the so called high frequencies (6 to 25 MHz) in 1928 and 1929. The radio circuits were successful, but quality was often degraded and the circuits were often interrupted by static and ionospheric disturbances. Despite the shortcomings, overseas traffic grew rapidly following the establishment of the radio circuits, and, in 1928, as cable-loading technology improved, Bell Laboratories began to investigate the possibility of a loaded undersea telephone cable to supplement the radio circuits. In 1930, the Bell System made a proposal to the British Post Office for a continuously loaded cable across the Atlantic Ocean to carry a single voice channel without intermediate amplification.

Viewed as either an extension of the band of the telegraph cables or of the length of the existing Key West–Havana telephone cables, a voice-frequency transatlantic telephone cable was a formidable undertaking.[1] Advances in materials had brought such a link within the range of feasibility but just barely. The design was for a cable with 165-dB loss at the top frequency of a 3-kHz voice band. The permissible receiving level was limited by thermal noise and the maximum transmitted signal to about 50 v by nonlinear distortion in the magnetic loading tapes. Transmission was to be over a single cable in one direction at a time with voice-actuated switching for two-way transmission.

An experimental 20-mile length of loaded cable was constructed and laid, first in the deep water of the Bay of Biscay and later, from the western shore of Ireland out to sea. The deep-water tests showed the cable transmission characteristics to be stable at great depths and unchanged by the stresses of laying and recovery, but they also revealed problems in cable structure and handling that profoundly affected later thinking. The shore-based tests showed that the loading tapes were stress sensitive, producing a kind of microphonic noise due to wave action on the rocky bottom. Both types of problems may have been solved by continuing development, but the project was suspended as the economic conditions worsened and as high-frequency radio was improved and expanded to meet the still growing traffic.

In 1931, AT&T installed an unloaded submarine cable from Key West to Havana.[2] Modified Type C carrier terminals were used to derive three 2-way voice channels in addition to a voice and telegraph circuit. In 1940, by a further modification of the terminals, seven carrier channels were obtained on this cable. There were no intermediate repeaters. In all the telegraph cables and earlier telephone cables to Cuba, steel armor wires and sea water provided the return circuit. In the 1931 cable, a cylindrical copper return circuit was used that was outside the insulation and concentric with the inner conductor. This formed a complete metallic coaxial structure and was the first commercial application of that structure for long-distance telephony. The early telegraph cables were insulated with gutta-percha, a natural tropical gum resembling rubber. The telephone cables were insulated with paragutta, an improved insulator with lower loss, based on the same material but with further processing and the addition of rubber and wax.

II. CABLES WITH SUBMERGED REPEATERS

Transatlantic traffic continued to grow, stressing the capacity of the improved high-frequency radio links and making it apparent that a single-channel system of any type would be of very little value. The only solution appeared to be in a broader-band cable with repeaters on the ocean bottom. With this objective in view, Bell Laboratories launched a long-range program of research and development in electron-tube and component technology to assure an operating life of at least 20 years in the undersea environment. Work was also started on the design of repeaters to transmit multichannel signals on a coaxial cable. The designs used the techniques being applied on land, but there were many new problems to be considered.

By 1942, a plan was in existence for a 12-channel system with repeaters at 50-mile intervals that could cross the Atlantic Ocean. The installation plan for long deep-water systems was always based on continuous laying, including the repeaters. Any procedure that stopped cable payout or that required man-handling of repeaters was to be avoided. These features seemed essential to the integrity of the cable, the safety of personnel and equipment, and the achievement of predictable performance. Therefore, the repeater was designed to be only slightly larger in diameter than the cable and was made of articulated sections so that it could be bent around and traverse the drum of the cable-payout engine, just as the cable did.

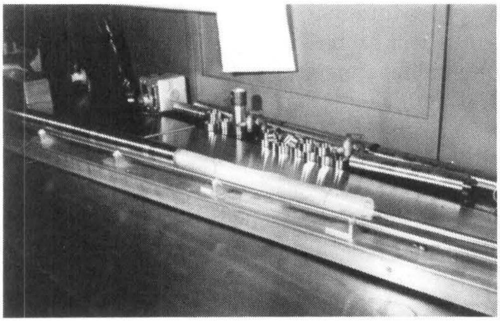
In the first design, the length of the active portion of the repeater was about eight feet and the overall length, including the cable tails, approximately 70 feet. The extremely limited space in the repeater and the long electrical lengths dictated rather narrow bandwidths and eliminated any fancy circuitry such as that required for bidirectional transmission. This further dictated the need for a separate cable for each direction of transmission. The design was thus, like the Type K carrier, a physical four-wire system using the same frequencies in separate cables for the opposite directions of transmission. While this simplified the repeater design, it added the expense of a second cable to achieve even the first two-way circuit. The basic articulated design was retained for all the two-cable systems.

2.1 The First Bell System Repeatered-Cable Project (Florida-Cuba)

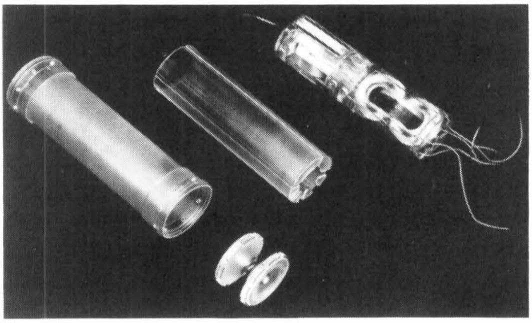
Along with other civilian projects, work on submarine cable was suspended for the duration of World War II. When work was resumed at the end of the war, improvements in both the cable and repeaters made it possible to plan a system to transmit 24 channels over two cables. To test the design and provide relief on an always busy route, designers planned a short system, designated the SA ocean cable, from Key West to Havana.[3]

The cables were very similar to the 1931 telephone cable, but polyethylene insulation was used instead of paragutta. The cable was slightly less than one-half inch in diameter to the inside of the outer conductor. Various weights of steel-wire armoring and layers of pitch-impregnated jute were applied outside the conducting structure, depending on the bottom conditions expected. Repeater components and vacuum tubes were assembled inside unit plastic cylinders, which were then enclosed in an assemblage of abutting steel rings each three-quarters of an inch wide. The entire assembly was enclosed in a long tube of soft copper 1-3/4 inches in diameter. The steel rings provided support against collapse under sea-bottom pressure, while the copper tubing and end seals kept the interior dry and at sea-level pressure [Fig. 12-1]. The assembled articulated repeater consisted of 15 sections and was about 7 feet long. When armored, it constituted a bulge in the cable about three inches in diameter and could be bent around the sheaves and drums of the cable ship to a radius of three feet without damage [Fig. 12-2]. Transitions at each end, tapered from the diameter of the the repeater bulge to the normal cable diameter, increased the total length to about 35 feet [Fig. 12-3].

The transmission band provided 24 one-way circuits in the range from 12 to 120 kHz in each cable. Repeaters were spaced at approximately 36 nautical miles (42 statute miles, 67 km) along the cable, with a gain of 65 dB at the top frequency. Power for the repeaters was transmitted as DC along the center



(a)



(b)

Fig. 12-1. The Key West–Havana SA submarine-cable repeater. (a) Repeater-network assembly unit. (b) View of repeater assembly.

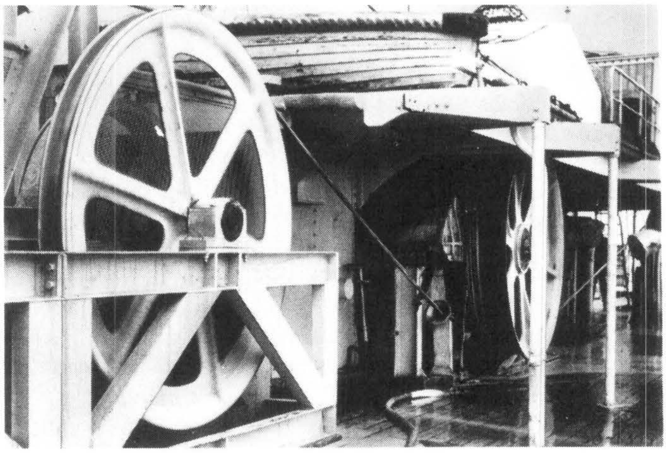


Fig. 12-2. Cable gear on the CS *Monarch*. The articulated repeater was designed to bend around sheaves of this diameter without damage.

Fig. 12-3. The Key West–Havana submarine-cable repeater, 1950. The repeater appeared as a bulge on the cable three inches in diameter. With end tapers, it occupied about 35 feet of cable.

conductor of the cable, very accurately regulated to a constant current. The electron-tube heaters were in series, and the anode potential, about 50 v, was derived from the total voltage drop across the three heaters of a repeater. Preliminary sea trials were conducted in Long Island Sound and, in 1948, on 15 nautical miles of cable off the Bahamas, at depths up to 2 miles. The cable was laid between Key West and Havana, and service over the system was established without incident in May 1950. The Key West–Havana repeaters, operating at depths of over a mile, were the first deep-sea designs capable of service on a transatlantic route.

III. THE FIRST TRANSATLANTIC SYSTEMS—TAT-1

In 1952, negotiations were begun with the British Post Office for a telephone cable between the United States and the United Kingdom, designated TAT-1. The main transatlantic link was to be from Newfoundland to Scotland, a distance of 1950 nautical miles, with depths up to 2300 fathoms. The route via Newfoundland was chosen because it provided the shortest direct deep-sea path from North America to Britain. A second submarine link of 270 nautical miles was to carry the circuits from Newfoundland to the eastern tip of Nova Scotia. Overland links were by a dedicated microwave system from Nova Scotia to New York and by conventional paired and coaxial cables from Scotland to London. Both underwater links were carefully sited to avoid the maze of existing telegraph cables [Fig. 12-4].

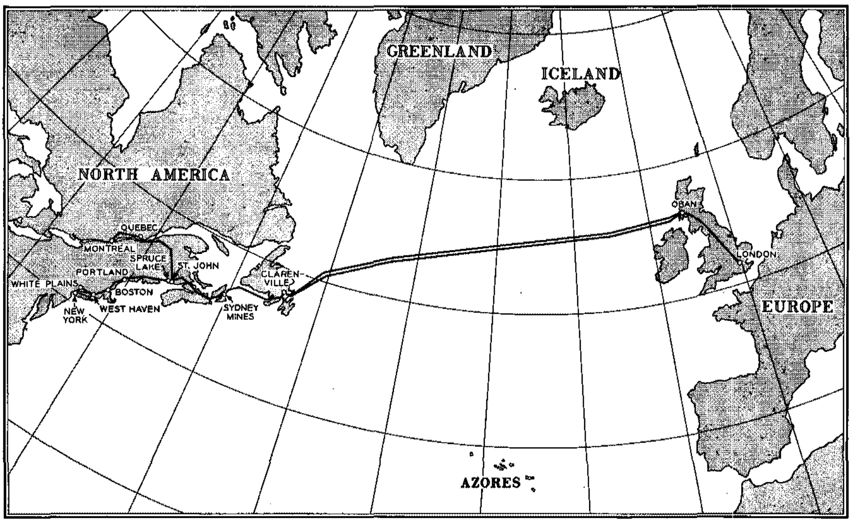


Fig. 12-4. Route of the first transatlantic submarine-telephone-cable system.

The Bell System people brought to the discussion their design of the flexible repeater and their experience with the successful Key West–Havana system. The British had pioneered designs for repeaters submerged in relatively shallow water and had successfully installed them in short systems as early as 1944.[4] Their design provided 60 two-way channels on a single cable in a band up to 552 kHz. They used more modern tubes of much higher gain than in the Bell Laboratories design but of less certain life and reliability. The two-way operation and consequent higher frequency made more compact circuitry essential. To house the repeater, they used a rigid canister that required the cable drum to be bypassed when laying the repeater [Fig. 12-5]. This required stopping the ship and was not considered by the Bell System representatives to be a satisfactory method of operation for long, deep-water links.

Many conferences and much correspondence followed, which included a full share of disagreements along with a high degree of cooperation. It was realized that compromise was essential but innovation and relatively untried methods were too risky. The phrase *proven integrity* was coined early in the project and became the guiding philosophy. An agreement was signed on November 27, 1953, specifying the two-cable method and the Bell System repeater design for the deep-water transatlantic link. The 60-channel British system would be installed between Newfoundland and Nova Scotia where the shallow water presented fewer hazards.

A cable with a diameter of 0.62 inch over polyethylene insulation was used for both systems and also for the overland link across the Newfoundland. The latter 63-mile span was, in effect, an extension of the British single-cable system. All sections were armored and otherwise protected as required by the local conditions [Fig. 12-6].

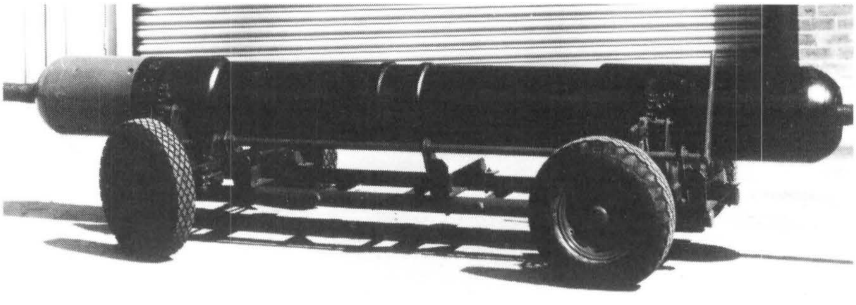


Fig. 12-5. The British rigid-repeater assembly, 1952.

In the Bell System design, designated the SB, the deep-sea repeater was patterned after the Key West–Havana design, but, by taking advantage of the lower loss of the larger-diameter cable, it was capable of transmitting thirty-six 4-kHz channels in the band from 20 to 164 kHz. Repeater spacing was 37.5 nautical miles; 51 repeaters were required for the 1950-nautical-mile link. The

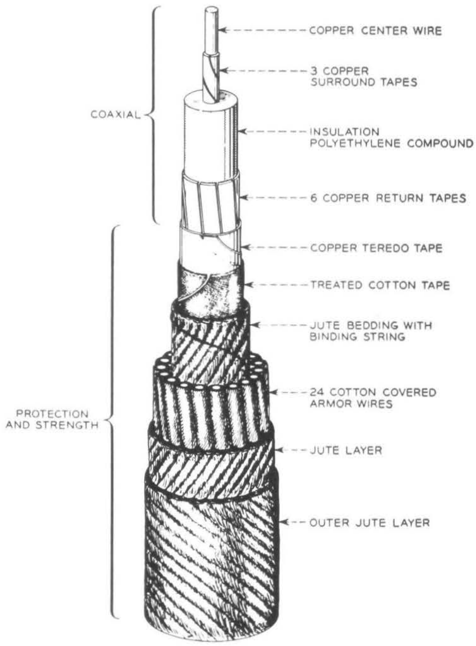


Fig. 12-6. Structural features of the deep-sea-type cable used on the first transatlantic system (TAT-1). Cable on rocky bottom near shore or where trawlers were active was more heavily armored.

required gain was furnished by a three-stage feedback amplifier of conventional design using the low-gain, highly reliable tubes.[5]

Power was transmitted as a direct current of about 0.25 ampere over the center conductor, which required 2000 v, oppositely poled, at the ends of the cable [Fig. 12-7]. The current was maintained constant within narrow limits. For this purpose, the voltage between the ends could be varied as much as 1000 v to offset the effect of earth-potential differences, such as might occur during a solar magnetic storm. The life of capacitors under a stress of 2000 to 2500 v to ground was a major factor in determining the permissible number of repeaters and hence the length of the deep-sea section. As in the SA design, the line current passing through the heaters of the repeater tubes in series furnished about 50 v for the amplifier-plate voltage. If a short circuit developed between the inner conductor and the sea near the ends of the system, the entire 2000 v with the stored charge on the line would be applied across one, or very few, repeaters. For their protection, a gas-tube bypass was included, which, by breaking down, would limit the voltage that could appear across a repeater.

In addition to the message bands, each system provided telegraph and telephone channels outside that band for maintenance and administration [Fig. 12-8]. Test and monitoring signals were also generated outside the band to locate faulty repeaters. In the Bell System repeater, this was accomplished by a very narrow band-pass crystal-filter circuit that bypassed the feedback circuit, providing a peak of noise at a frequency unique to each repeater. In the event of a transmission failure, the last functioning repeater could be identified at

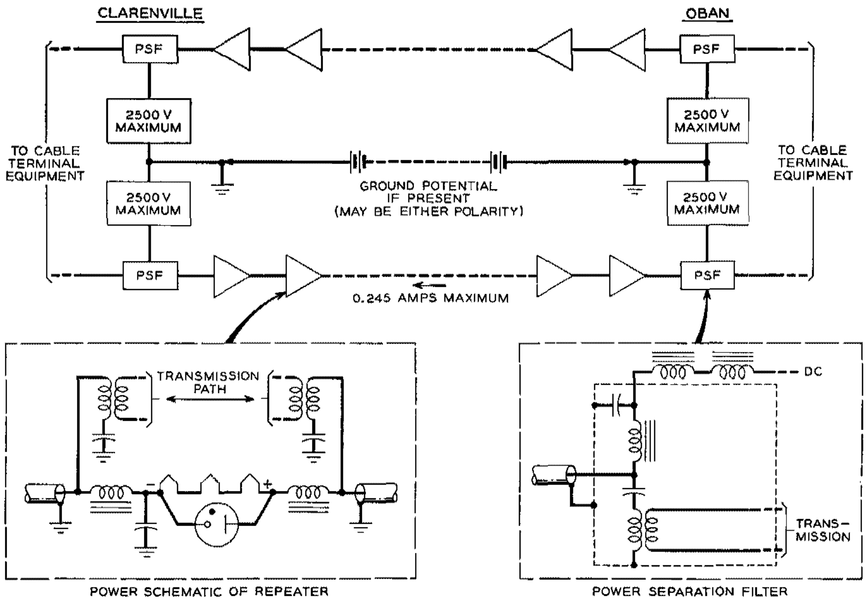


Fig. 12-7. The TAT-1 cable DC-power loop schematic.

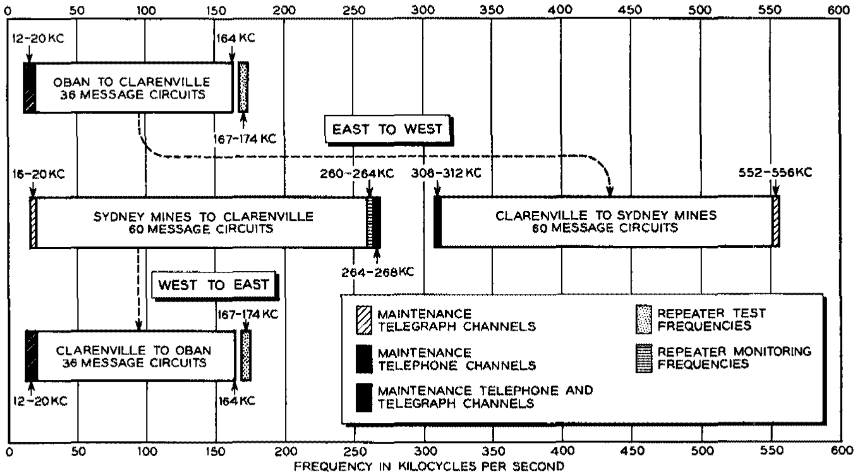


Fig. 12-8. Frequency allocations in TAT-1 submarine-cable links.

the receiving end by the frequency of the first missing frequency in the "picket fence" of noise signals.

Many new techniques were introduced and many hazards dealt with in the manufacture and installation of TAT-1. New manufacturing methods were required, particularly in the area of quality control. The TAT-1 components were assembled under conditions of a surgical clean room, and inspection methods established new standards of thoroughness.

The entire TAT-1 facility was installed in 1955 and 1956 and went into service on September 25, 1956. It was taken out of service in 1979 after exceeding its 20-year design life. There were no system outages from electronic or mechanical failures, despite numerous breaks by fishing trawlers. The circuits realized were of high quality, and traffic was enormously stimulated. The number of transatlantic calls increased by about 20 percent per year, a compound rate of growth that was to persist for at least two decades. Although it was realized that the flexible repeater and the systems built around it would have little prospect for higher capacity, the pressure for improved service to other locations at early dates was very high. Before the flexible-repeater era was over, additional SB systems were installed to France (TAT-2), Alaska, Hawaii, and Puerto Rico, for a total of 8000 nautical miles. As of January 1983, almost 3500 miles were still in service.

Installations were not without incident. The ship encountered a hurricane during the laying of the first transatlantic link. During the installation of the second system, TAT-2, the cable ship *Ocean Layer* caught fire at sea. The ship was destroyed, but everyone escaped and no one was seriously injured [Fig. 12-9]. As they watched the burning ship from the lifeboats, more than one telephone engineer (and seaman) may have wondered, if only for the moment, whether he had indeed chosen the right profession.

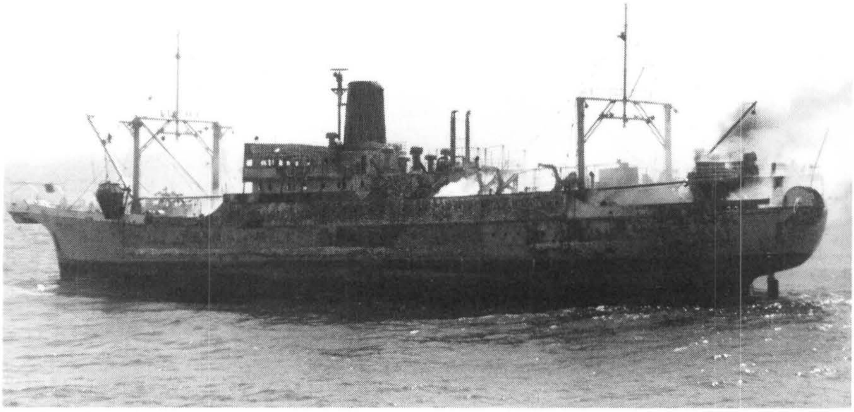


Fig. 12-9. The CS *Ocean Layer* burning at sea.

IV. DEVELOPMENTS IN ENGLAND

Submarine telephone-cable developments were pursued in several other countries in addition to the United States. France and Germany had active programs, but the next major advance came from England. During and immediately following the TAT-1 work, the English had improved their rigid-repeater system to the point where they were ready for a deep-sea installation. In April 1957, the British and Canadian governments agreed to provide a direct link between the countries via a system, designated Cantat-1, based on the British design. The transatlantic link was viewed as the first part of a worldwide British Commonwealth system.[6]

In addition to the rigid two-way repeater, the system featured a new lightweight cable of a radically new design, also developed by the British. Considerable strength is required in a submarine cable because of the length suspended between the ship and the ocean bed and because of the large dynamic stresses attending laying and the even larger stresses if recovery is necessary. The strength and protection of the conventional cable, used since the early telegraph days, derived from a helical layer of steel armor wires outside the coaxial transmission structure. Under laying tension this helix uncoiled like a giant spring, through many turns, between the ship and ocean bottom. On the bottom the tension was relieved and the helix regained its original coil, since slightly more cable was paid out than the linear distance traveled in order to avoid suspensions over irregular terrain. In this situation, if the ship stopped for any reason and the tension changed abruptly, as was likely, for example, from the vertical motion of the ship, the cable would tend to flip into bights. If tension were then reapplied, the bights were likely to pull into kinks that could easily damage the cable. In the British Post Office design, the steel-strength member was placed at the center of the cable and was torsionally balanced. Much less steel was required in the protected location, and the cable was much lighter, further reducing stresses. The inner conductor was formed by cladding the steel strand with copper, and the outer conductor was protected by a tough plastic sheath. Jute and pitch, trademarks of the art since Cyrus Field, appeared to be gone forever. (They reappeared, at least temporarily, in the 1980s after some bad experiences with plastic as protection for some armored sections of cable.)

The repeater amplifier made use of new long-life British "valves" operating at lower voltage than the vacuum tubes used in TAT-1. Even so, with repeaters 20 miles apart, 10,000 volts, 5000 at each end, oppositely poled, was necessary for the power feed. The larger container made physically larger capacitors feasible, so that long life was possible even under the high stress. The cable ship was also modified to provide for a degree of automation in repeater handling and to permit continuous laying of the cable and repeaters.

Cantat-1 was completed in November 1961. It provided 60 two-way 4-kHz circuits (or eighty 3-kHz circuits). The transatlantic link was from Oban, Scotland to Corner Brook, Newfoundland. An even wider-band link, designed along the same general lines, provided 120 two-way circuits in a single cable to the Canadian mainland. Another innovation in Cantat-1 was the use of ocean-bottom equalizers placed in canisters similar to the repeater housing. Their adjustments were made on shipboard, based on measurements as the cable was laid, before sealing the container. The system had a noise objective of 1 pw (picowatt, 10^{-12} watt) per kilometer, which, although not perfectly achieved in all parts of Cantat-1, became an accepted international standard for long submarine cables. The standard of 1 pw per kilometer differed only slightly from the Bell System standard for long circuits. Both were weighted for the difference in subjective effect of noise across the voice band.

V. SAVING BANDWIDTH

5.1 Sixteen-Channel Banks

Compared to the land-based systems, submarine-cable channels were extremely expensive; in 1956, the first system cost roughly one million dollars per two-way channel. One way to reduce costs was to increase the band and obtain more channels with approximately the same material outlay. The other approach was to realize more channels in the same bandwidth.

Spacing at 4-kHz intervals had long been used for carrier systems in the land plant, permitting the use of inexpensive filtering and modulation techniques, which, in the case of short systems, was very important to the total system economy. In the expensive undersea plant, this was not the case. Money could be well spent in the terminal to get more voice channels on a given system. For this reason, a channel bank was designed with more complicated circuitry; it provided 16 channels spaced at 3 kHz in the band formerly occupied by 12 channels at 4-kHz spacing.[7] This equipment was more expensive but very cost-effective.

The actual band transmitted was from 200 to 3050 Hz. The useful 2850 Hz, occupying 95 percent of the intercarrier space, was achieved by extremely sharp crystal filters and a double-modulation scheme that made transmitting speech frequencies to 3050 Hz possible. While the lowest speech frequency was 200 Hz, the carrier was only 75 Hz lower. The double-modulation scheme made it possible, in effect, to slide the transmitted speech band about, making the choice of the specific top and bottom frequencies for the 2850-Hz transmitted band independent of the spacing from the carriers. The narrower band was judged to be equivalent to about 2 dB greater loss, compared to channels derived from the normal 4-kHz spacing, but the sacrifice was deemed to be well worth the one-third gain in channels. The 16-channel equipment was never manufactured in the United States but was available from French, British, and Japanese manufacturers. Starting in 1959, most undersea systems used 3-kHz-spaced channels.

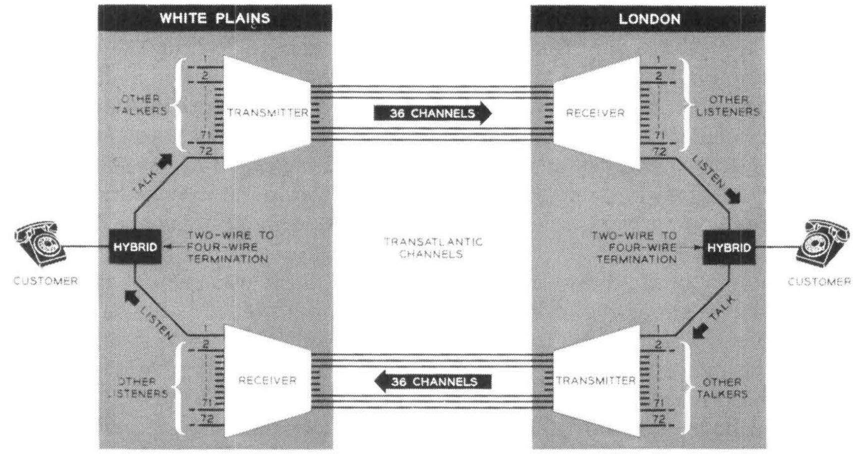
5.2 Time-Assignment Speech Interpolation—TASI

Another approach to multiplying circuits was to improve the time fill on the circuits. Time-assignment speech interpolation (TASI) was an idea that had been around since at least the 1930s. TASI was based on the fact that in four-wire circuits, used in all long-haul systems including the submarine cables, a one-way channel was dedicated to each talker, and, in normal conversation, the average talker actually spoke less than 40 percent of the time. Using sensitive speech detectors, fast electronic switches, and a signaling system and terminal circuitry to keep the ends in synchronism, the system permitted a larger number of voice circuits (trunks) to time-share a smaller number of channels. This was accomplished by interpolating talk spurts.

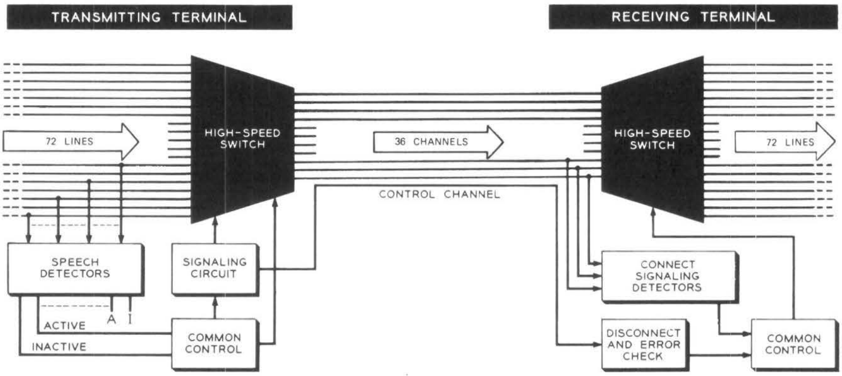
As long as the number of channels exceeded the number of talkers, all talkers were connected full time. When more talkers were connected than the number of full-time channels, a newly active talker was served by disconnecting a momentarily inactive talker and switching the channel to the active talker. The terminals kept track of the talker-to-channel connections, and the end-to-end signaling system ensured that the talkers were connected to the correct listeners at all times [Fig. 12-10].[8]

With TASI, as with any line-concentration scheme, there was always a possibility that more talkers would be seeking access to the channels than there were channels available at that moment. When that happened, the speech of the overflow talkers would be clipped or "frozen out." The statistics with 72 talkers on 36 channels was such that the freeze-out was hardly perceptible at the concentration ratio of two to one. Over the years, many improvements in interpolation terminals were made as circuit techniques and components improved. In even the first TASI, the speech signals were switched in time-sampled form (pulse-amplitude modulation format). It was, indeed, the first commercial time-division switch, although the samples were not coded.

The first TASI system (TASI-A) was put in service in 1959 in New York and London, providing 72 talking circuits over 36 channels in the TAT-1 system. Additional installations followed on submarine cables to France and from the West Coast to Hawaii. Subsequent generations of TASI designs reduced the cost and power consumption, increased the capacity, and improved the performance. In one version, the terminal processing and switching was carried out on coded binary samples via pulse-code modulation. The impairment due to overload was made much less perceptible, and concentration ratios of at least four to one appeared feasible. Speech interpolation terminals have been, and probably will continue to be, used over much of the undersea network.[9]



(a)



(b)

Fig. 12-10. Time-assignment speech-interpolation (TASI) terminals. (a) The basic TASI plan on the first transatlantic circuits, where 72 talking paths were provided on 36 channels. (b) The major parts of a TASI system for one direction of transmission. The opposite direction of transmission required the same circuits, but arranged in reverse order.

One impact of both the 16-channel terminal and TASI was to increase the total broadband signal power. This, in turn, increased the required load-handling capacity of the transmission system.

VI. CANISTERS, CABLE, AND COMPONENTS

During the late 1950s and throughout the 1960s, long-distance traffic was growing at a rapid rate. Overseas calling shared this boom, and the success of the first submarine cables provided an additional stimulus on the transatlantic route. The number of transatlantic calls was growing at a compound annual rate of 20 percent with no sign of slowing down. The limited capacity of the

SB design was inadequate to handle the projected traffic growth, and the flexible-repeater approach was not suited to a much higher capacity. Exploratory work on a system better adapted to the growing need had been started in 1952, four years before the first system was installed and ten years before the new system was completed. While the development cycle was shorter in later generations of design, this was indicative of the long lead time required to establish the necessary integrity and reliability in undersea systems and the need to overlap project developments in order to have higher-capacity systems available in a timely way.

As the new phase of exploratory work began, there was general agreement in the Bell System on the following points: (1) A rigid-repeater housing had to replace the flexible repeater to provide the space and component arrangements that would permit wide bandwidth and bidirectional transmission on a single cable. These features were needed to meet the growing channel demand as economically as possible. (2) A new cable design was required that would put the strength member inside the inner conductor and thus allow a larger coaxial cable in a given outside diameter. A large coaxial cable would help offset the higher losses associated with higher frequencies. (3) The functions of cable and repeater stowage on shipboard, cable-payout control, and overboarding should be reexamined and approached as a major system problem. The first two items, of course, reflected decisions the British had already made in the design of their higher-capacity, short, shallow-water system. The challenge was to adapt these essential advances to the conditions of a long, deep-water link. The third item was the signal for a full-scale attack on the shipboard and installation problems raised by the use of rigid repeaters.

6.1 Cable

The new cable was similar in general design to the type pioneered by the British [Fig. 12-11]. The strength member consisted of a bundle of steel wires of high tensile strength at the center. No external armor was used in the deep-sea portions, but armored versions were provided for rocky locations near shore or where trawler activity was expected. The inner conductor was a continuous tube of copper over the steel strands with a welded seam. The outer conductor had a diameter of one inch over the polyethylene dielectric and a simple overlap seam of about 0.25 inch. The overlap seam accommodated dimensional changes from pressure and temperature, while the outer polyethylene jacket maintained the contact and prevented buckling of the copper tubing during bending and unbending.

6.2 Components

While these basic issues were debated and the options explored, many other questions were raised and decisions made. By the late 1950s, transistors had emerged from the experimental stage and had to be seriously considered for submarine-cable use. They were especially attractive because of the lower power and potentially greater life and reliability they promised. But the time required to establish life and reliability in the new devices, which otherwise were so tempting for the application, was not available. A commitment to proceed

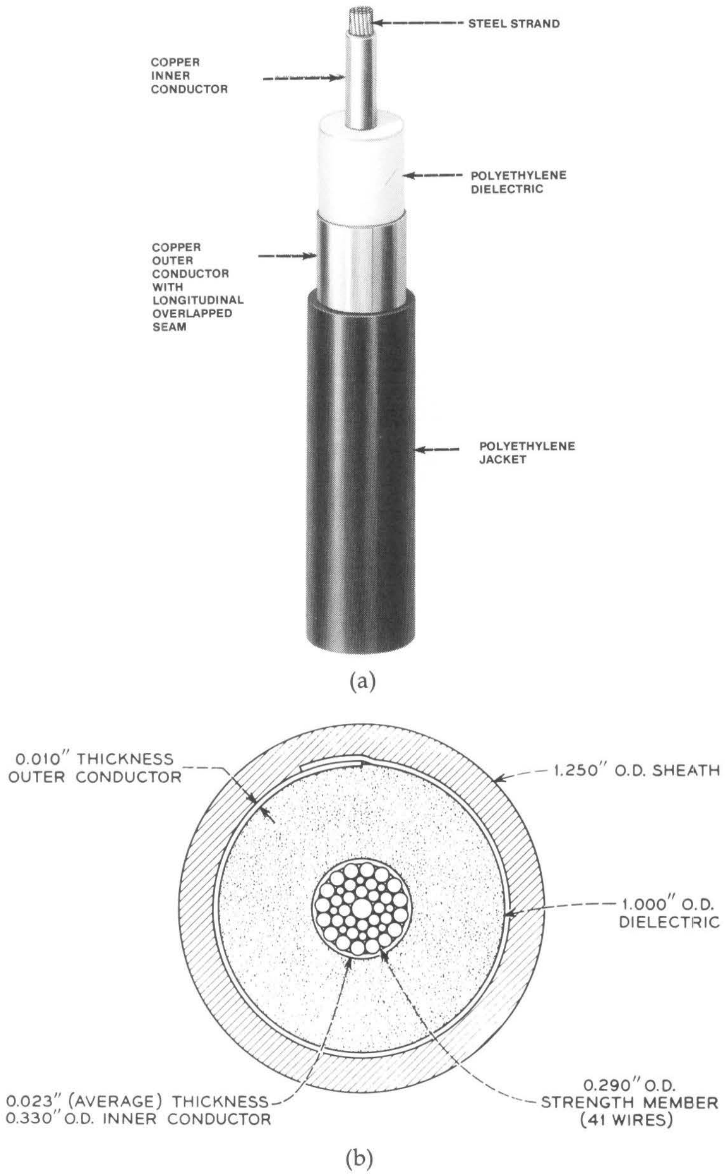


Fig. 12-11. Lightweight armorless ocean cable. (a) Appearance and general construction. (b) Cross section.

along a development path had to be made several years before the required service date; confidence in the components was essential at the time of commitment. In the mid-1950s, transistors with the required performance were just beginning to be available. Several years of testing on large numbers under carefully controlled conditions would be necessary before the required life could be established with any certainty. (See another volume in this series, *Electronics Technology (1925-1975)*, Chapter 1, Section 5.2.) In contrast, a great deal of experience had been accumulated on the reliability and long life of carefully designed and manufactured vacuum tubes. Many of the tubes used in the SB design had been on life test for over 15 years with no failures, and various operating conditions had been tried that lent confidence to a tube design based on that experience. The reluctant decision was that transistors would have to establish a track record before they could be used in submarine cables. In 1955, work was started on a tube with a transconductance of 6000 micromhos. This was a conservatively designed tube by the standards of that time.

VII. THE LINEAR CABLE ENGINE AND A NEW CABLE SHIP

7.1 The Linear Cable Engine, Cable, and Repeater Payout

While the basic decisions on the housing, cable, and components were being made, a great deal of attention was directed to the ship and installation problem. The flexible repeater had been coiled in the cable tanks along with the cable. The rigid repeaters, in contrast, would be stored on the cable working deck and the associated connecting leads carried into the cable tank in an orderly, preplanned fashion. The cable-tank restraining device, called a crinoline, and the working deck around the tanks had to be designed to allow these leads to leave the tank as the cable was payed out, to pick up a repeater wherever one had been spliced into the cable, and to payout the repeater and the cable that followed. This was to be done without assistance, beyond the minimum required to position the repeater for the takeoff, and without stopping the ship.[10]

Submarine cables varied in weight, but all were considerably heavier than water and, unless restrained, would run out of the tanks at a rate that had no relation to the distance traversed by the ship. It was necessary to control the rate of payout so that the cable would lie smoothly on the bottom. A taut wire, payed out in parallel with the cable, gave an accurate measure of the distance traveled over the bottom and a small percentage of slack was allowed in the cable payout to avoid suspensions between high points. For telegraph cables and the SB flexible-repeater system, cable payout was controlled by multiple turns of the cable around a braking drum. But this made for an exceedingly awkward arrangement in launching a rigid repeater. A new cable-payout engine was needed at the stern to handle the cable and rigid repeaters without interruption, with no turns around a drum. Lastly, a smoothly contoured chute was proposed for the stern of the cable ship to avoid a large-diameter sheave at that point.

The study of these problems started with the construction of a one-quarter scale model of a new in-line (linear) cable engine and the basic ship arrangements. The next step was construction of a full-scale mock-up of the cable-storage and repeater-payout arrangements on a hillside in Chester, New Jersey [Fig. 12-12]. This very valuable facility was dubbed the CS *Fantastic*. It taught the designers many lessons and was used to establish the guidelines for the design of the cable-handling features on a new cable-laying ship for the new system.

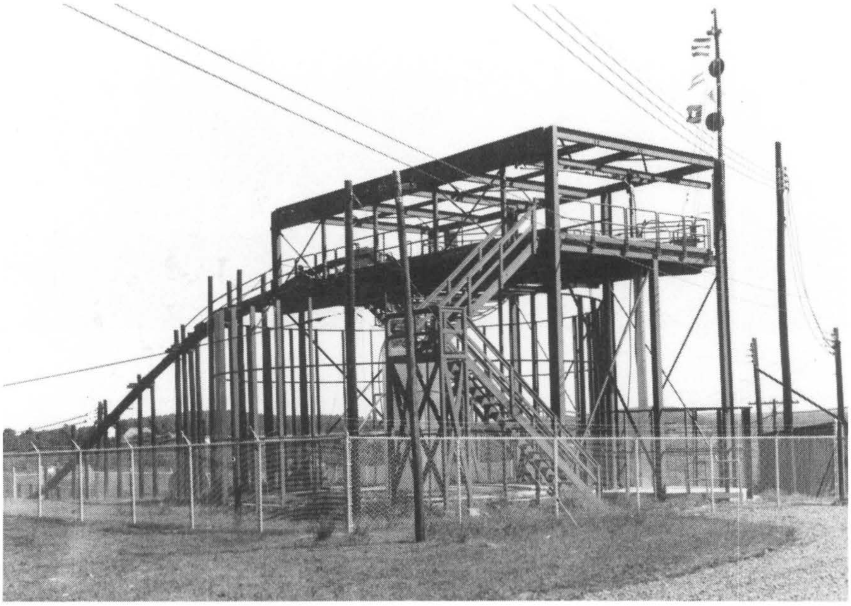
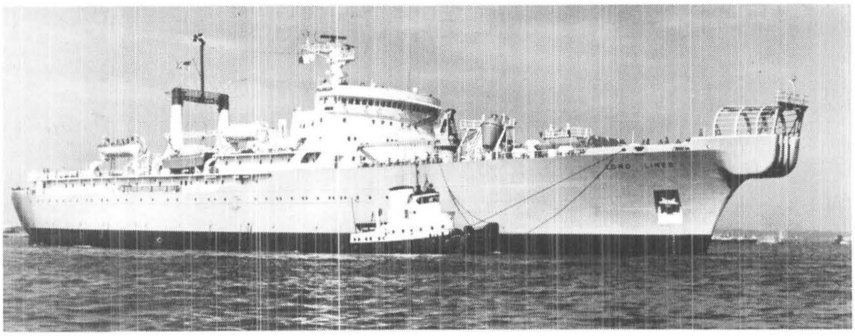


Fig. 12-12. Full-scale skeleton model of the CS *Fantastic*. The signal flags proclaim, "Stand clear, I am laying cable."

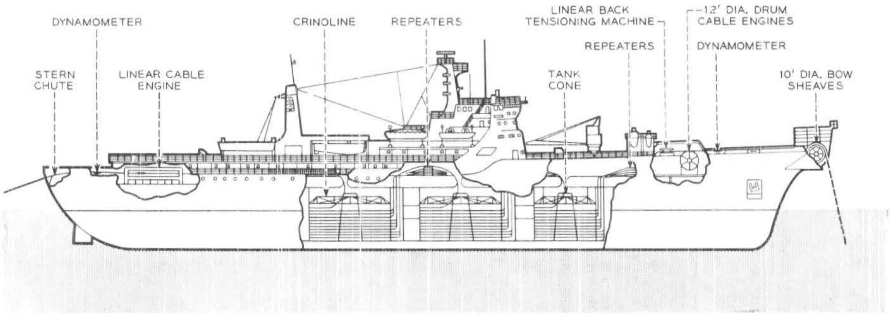
7.2 Cable Ship *Long Lines*

Original plans contemplated modification of the British cable ship *Monarch*, but the changes were so extensive and the projected cable-installation program so large that the modifications were never made. Instead, AT&T decided to build its own cable ship, CS *Long Lines*. The ship was designed by the eminent firm of naval architects Gibbs and Cox, Inc. and built in a German shipyard. The linear-cable engine was constructed by the Western Gear Company of Seattle, Washington. All final design work was carried on in close collaboration with the Bell Laboratories system developers. This was a particularly exciting venture since the Bell System had never been associated with anything quite like it before. The large scale of things and the massive, clanking, machines were in marked contrast with the miniaturization and absence of motion otherwise characteristic of electronic design. The full story of the design and construction of the cable ship and its machinery is a saga by itself, full of strikes, delays, and adaptation of plans, but somewhat outside the scope of this history. The technical story is treated in detail in the literature.[11] The accompanying pictures and diagrams should convey better than words the general nature of the ship and equipment [Fig. 12-13].

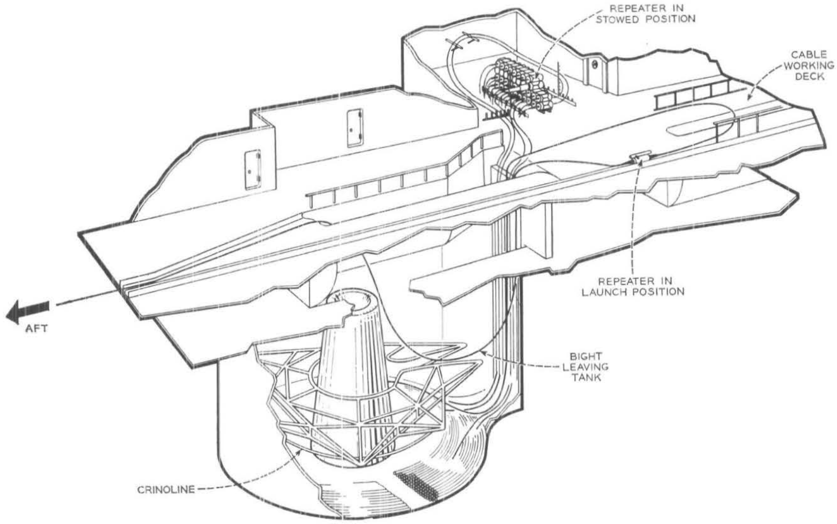
The CS *Long Lines* was launched in September 1961 and sailed to her first cable laying job in July 1963, too late for the very first installation of the new cable system, but in time for its new equipment to be used on the transatlantic installation of the radically redesigned cable and repeaters. There was some



(a)

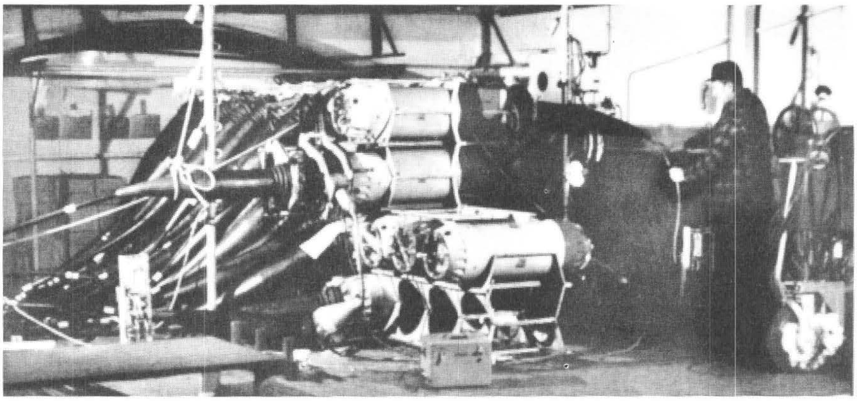


(b)

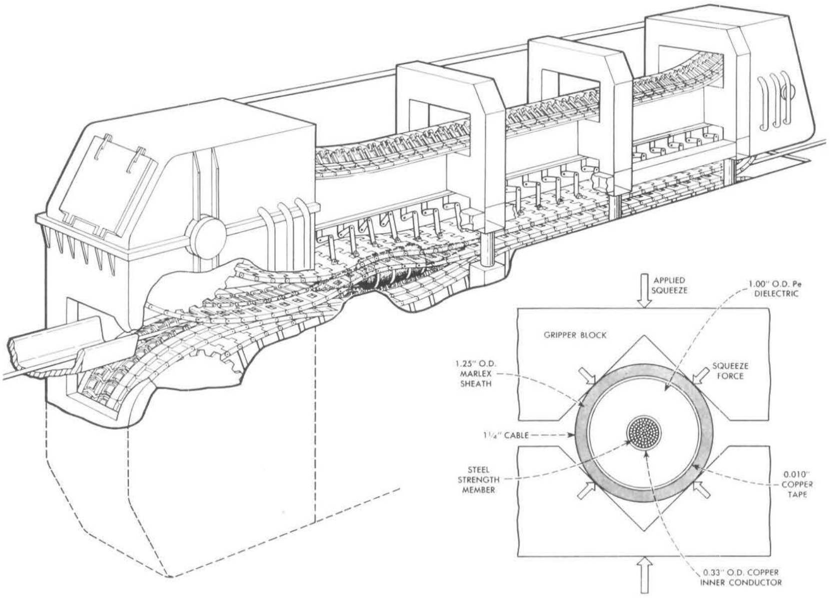


(c)

Fig. 12-13. (a) The CS *Long Lines*. (b) General layout of the cable storage tanks and the cable-laying system. (c) Cable and repeater payout arrangement on the CS *Long Lines*. Cable and repeater stowage and repeater launching.

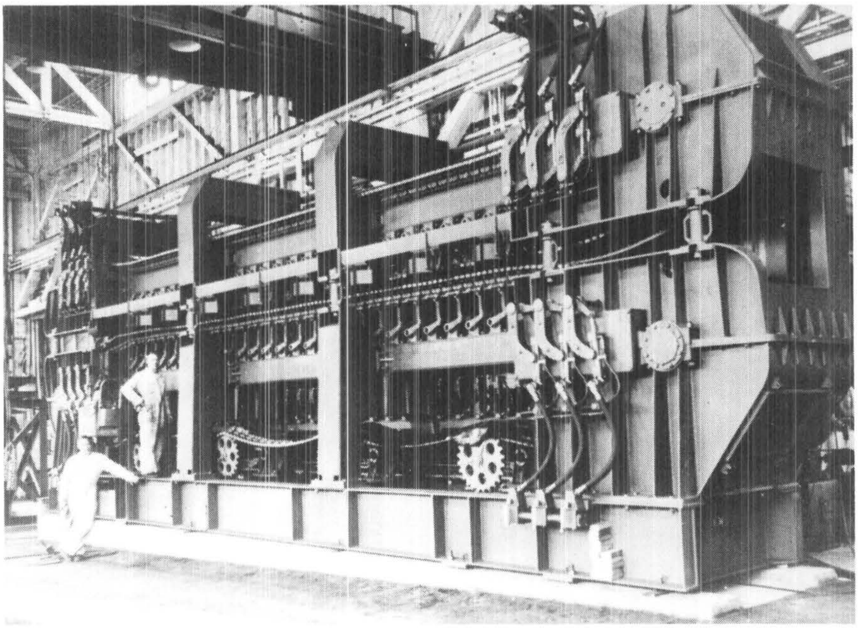


(d)



(e) (f)

Fig. 12-13. (d) Repeater stack on the cable working deck. The final assembly of the repeater into the cable was done on board prior to the laying operation. (e) Linear cable engine with rigid repeater passing through the engine's caterpillar tracks. (f) Cross section of the cable in a gripper block of the engine track showing the forces applied.



(g)

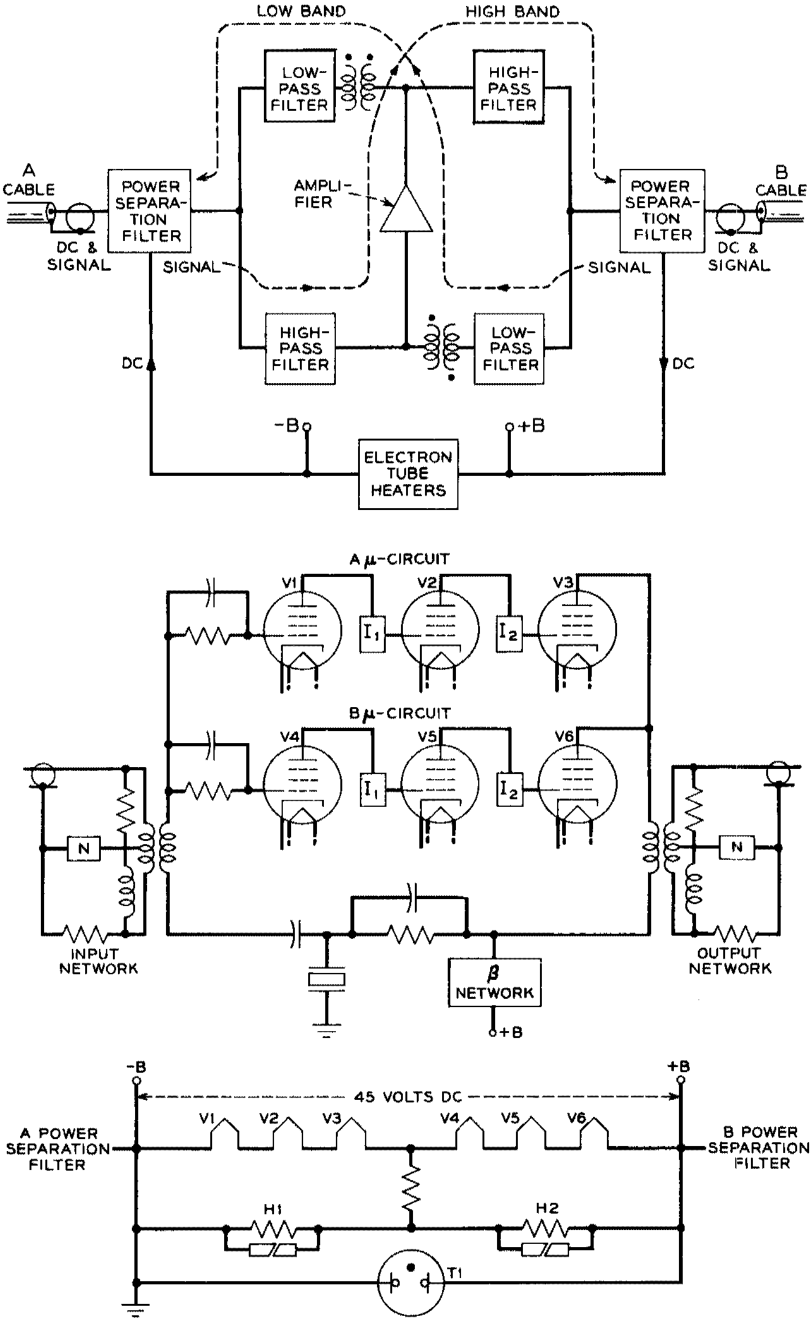
Fig. 12-13. (g) The linear cable engine prior to its installation on the cable ship.

trepidation about the large number of innovations and the complexity of her equipment, but the rigid repeater, armorless cable, and the ship's cable-payout system have been the framework for all subsequent Bell System submarine cable development. By 1984, the ship had laid over 45,000 nautical miles of cable without serious incident.

VIII. THE SECOND-GENERATION SD SYSTEM

By the middle and late 1950s, enough progress had been made in all essential areas to commit to a specific development of a second-generation system, designated the SD system. The goal was a single-cable system with a bandwidth of about 1 MHz. The system was to be capable of direct connection between the United States mainland and Europe. Repeater spacing was to be 20 nautical miles. The transatlantic system was to be in operation in 1963, with a shorter pilot system set for late 1962.

The 6000-micromho vacuum tubes were to be used as the active devices, and the same amplifier would be used for both directions of transmission. Despite the extensive experience with vacuum tubes, the new design was enough of a departure and life data too limited to trust to a single string of tubes. Parallel amplifiers were provided for redundancy to maintain transmission in the event of most types of tube failure [Fig. 12-14]. The repeater amplified a band from about 100 to 1100 kHz in which a minimum of 128 two-way 3-kHz channels was derived [Fig. 12-15]. One-half of the band was assigned to one direction of transmission and the other half to the opposite direction, with



(a)

(b)

Fig. 12-14. The SD submarine-cable repeater. (a) Repeater block schematic, showing transmission in both directions by means of low- and high-band frequency separation. (b) Circuit schematic, showing paralleled amplifiers, generation of the plate voltage by the drop across the heaters in series, and gas-tube surge protection.

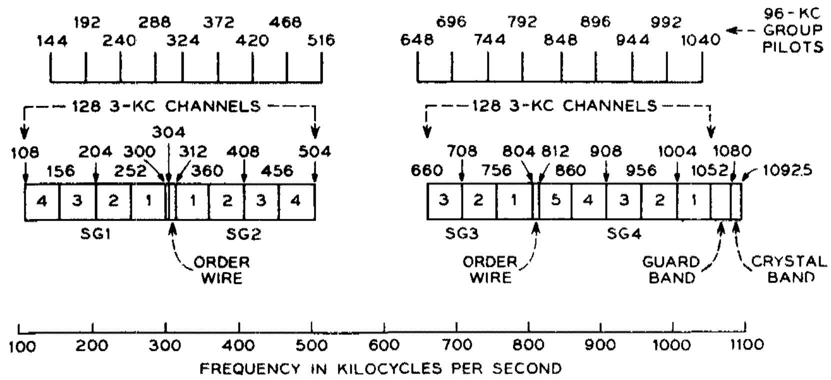


Fig. 12-15. The SD signal spectrum. Frequency allocations for 128 three-kHz message channels in 16-channel groups.

high-pass/low-pass filter combinations to separate the directional bands at the input and output. As in all coaxial submarine cables, DC was fed along the center conductor.

The repeater was contained in a rigid housing made from a special heat-treated copper beryllium alloy, chosen for its resistance to corrosion [Fig. 12-16].[11] Within the housing, the components were assembled in so-called cord-wood packaging in plastic mounts. This approach, though awkward in some respects and somewhat limiting by lengthening feedback paths, had the virtue of highly reproducible transmission and made the soldered connections accessible for inspection [Fig. 12-17].

A pilot system was planned from Florida to Jamaica (with a big detour around Castro's Cuba). The transatlantic system was to follow in 1963. Everything went according to plan, except that the completion of the *Long Lines* was delayed, and it was necessary to use the British cable ship *Alert* to lay the Florida–Jamaica link. This meant that the first operation of the *Long Lines* was a big one, the third transatlantic system, TAT-3, a 3500-nautical-mile (6480-km) system between Tuckerton, New Jersey and Cornwall, England [Fig. 12-18]. Laying the cable was completed in October 1963; and, after final equalization, 138 channels, ten more than the original objective, were found acceptable for service.

In SD submarine cable, as in all long analog systems, the equalization plan was of critical importance. It was necessary to obtain a signal-to-noise ratio consistent with the requirements of the network. Since a transatlantic link is invariably part of a long circuit, often with substantial overland extensions at each end, the signal-to-noise goals were set very high. The job in undersea cable systems was simplified by the fact that the environment, while extreme in some respects, is very stable. However, there is extremely limited access to any part of the system after installation, except the ends.

In the TAT-3 SD system, the total transmission loss in 3500 nautical miles was 9000 dB at the highest transmitted frequency. This was to be matched by

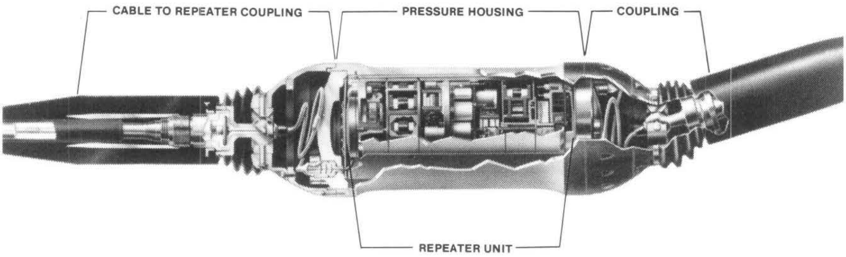


Fig. 12-16. The SD system rigid repeater.

the gain of 180 or more repeaters. Obviously, even small deviations had to be corrected at intervals along the system to prevent accumulation of very large signal misalignments. Basic equalization of the cable loss was accomplished by shaping the repeater gain to match the cable loss at standardized temperature and pressure. The individual section lengths of the manufactured cable were trimmed in terms of expected temperature and pressure of the specific site. The

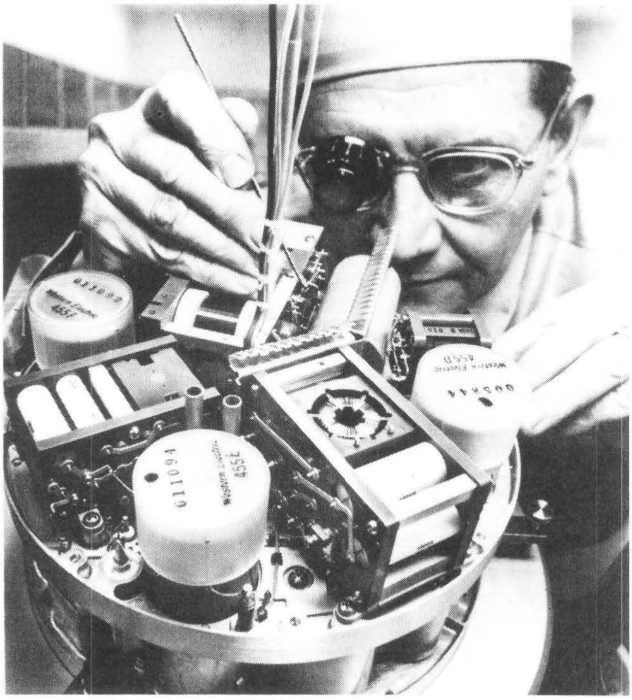


Fig. 12-17. Cordwood packaging of SD repeater components. The packaging arrangement made all soldered connections accessible for inspection.

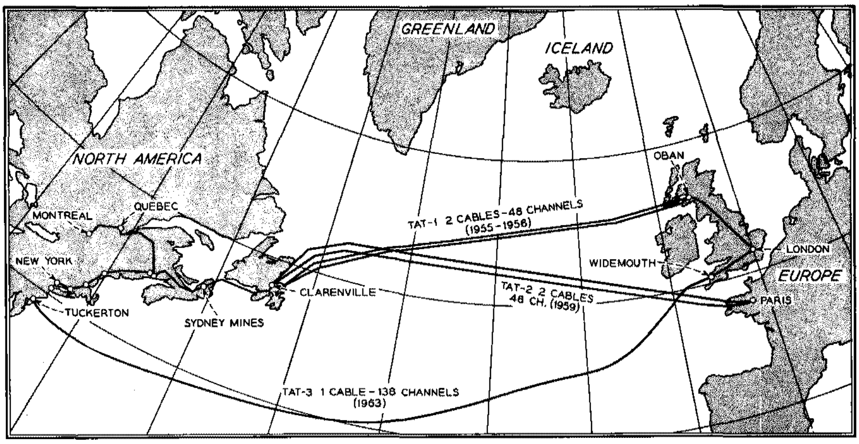


Fig. 12-18. Routes of the first three Bell Laboratories-designed transatlantic-cable systems.

objective was to obtain the correct loss at the top transmitted frequency. All residual deviations were then equalized at intervals of 200 nautical miles by an adjustable equalizer. The proper adjustment was determined from an analysis of transmission data taken continuously as the system was installed. In the SD system, the equalizer was adjusted just before it went overboard. After the system was installed, the signals at the terminals were adjusted in level across the frequency band so that the optimum overall signal-to-noise performance was obtained.

The installation of the first SD systems was not without incident. In a design measure to relieve an earlier problem, a section of the repeater external to the sealed canister, called the end bell, was flooded with sea water. During the second lay of TAT-3, the ship's crew occasionally discovered broken bolt ends in the repeater-stowage area. Investigation proved that sea water and oxygen could cause stress-corrosion cracking of the bolts that held sections of the end bell to the repeater body. Enough bolts held to avert catastrophe, and revised heat treatment alleviated the problem. Later, the designers went to a design in which the large cylindrical parts themselves screwed into each other. This eliminated the cracking problem and incidentally produced a significant cost reduction. Misfortunes occasionally did pay off.

Another problem was nonlinear sing, a phenomenon found in equivalent four-wire systems using a single amplifier for both directions of transmission. Under certain conditions, such systems can lock up in a so-called noise sing. This is generally the result of gross overload in the presence of large misalignment. It was encountered in testing the first transpacific SD system with much of the cable on shipboard. The problem was attacked methodically, and computer programs were developed to predict nonlinear-sing margin in a system for different misalignments.[12] These and other less serious problems were met and contained. By 1984, more than 19,000 nautical miles (35,000 km) of SD systems were in service. Also, by that date, the TAT-3 was over 20 years old and had experienced no repeater failures.

IX. THE SF SYSTEM

With the high-quality circuits provided by TAT-3 and subsequent SD systems, overseas traffic continued to expand. Exploratory efforts had started in the late 1950s on transistor characterization, life tests, and the development of amplifier circuits for a new wider-band undersea system, the SF design. By 1963, a detailed development plan for a system using solid-state amplifiers was completed. The first system was to be installed between Florida and the Virgin Islands in 1968. A transatlantic SF system, TAT-5, established a link between the United States and Spain in 1970.[13]

Except for the wider band and the use of transistors, the SF system followed the SD design, both physically and electrically. New transistors were required, and many design problems were associated with the wider bandwidth but none that were not solved in due course. Enough data was accumulated and life testing experience was sufficiently encouraging that the parallel amplifier used in SD was omitted. It was not clear that the additional transistor string would have helped much in any case, but the promised life and reliability of the solid-state circuitry was indeed delivered. The top frequency of the SF system was 6 MHz, and repeaters were spaced at intervals of 10 nautical miles. The cable diameter over the dielectric was increased to 1.5 inches, but otherwise the cable was similar to the SD design. The overall loss of a system 3500 nautical miles long at the top frequency was about 15,000 dB. Misalignment was controlled at intervals of about 200 nautical miles by so-called ocean-block equalizers. Initially, the capacity objective was 720 two-way channels spaced 3 kHz apart, but, with careful equalization, about 845 were ultimately achieved in practice [Fig. 12-19].

Like the SD system, the SF system was not without its initial problems. Five seals failed during the installation and early operation of the first SF link, a cable from Florida to St. Thomas in the Virgin Islands. A design defect was identified and the problem corrected in later production. Some high-voltage breakdowns in the polyethylene-enclosed pigtail lead-ins also occurred on this link, revealing a defect in the insulation-molding procedure during splicing at that point. This problem was also eliminated by improved molding methods. The experiences, while costly, demonstrated the wisdom of short trial systems for new designs.

A much more subtle problem developed over a long period in the deep water sections of the TAT-5 transatlantic SF system. Materials were carefully selected to avoid corrosion in sea water, but one situation escaped attention. Electric currents are generated by ocean-bottom sea-water currents traveling in the earth's magnetic field. In many areas of the ocean, particularly around the mid-Atlantic ridge and similar formations, the rate of the ocean currents is significant and varies from place to place. The current generates a potential between points and causes an electrical current to flow in the cable's outer conductor. After many years of operation, this caused corrosion and loss of the outer conductor of the repeater pigtail. In systems such as the SF, loss of the outer conductor changes the cable characteristic at high frequencies. The effect becomes serious after a number of repeaters have been attacked. This effect was first discovered on TAT-5 and was corrected by picking up and relaying a long section of the system with the pigtail isolated from the sea water

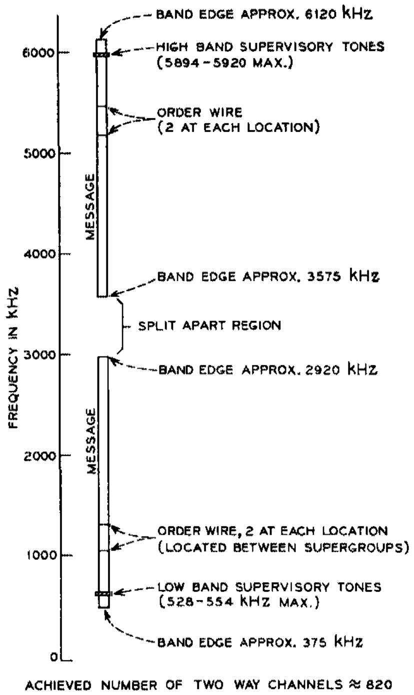


Fig. 12-19. The SF system frequency allocation.

in the new repeaters. As of 1984, the system continues to be used, and even installed, where its capacity and cost offer the greatest economy. By the end of 1983, over 17,000 nautical miles (31,000 km) were in operation.

X. THE SG SYSTEM

In the late 1960s, as traffic continued to grow at a rate in excess of 20 percent per year on the transatlantic route, it became apparent that the SF system would not provide adequate increments in capacity for very long. Despite the widespread use of satellites and the rapid growth of traffic on them, it was generally accepted that both cable and satellite technologies should be developed to their maximum potential. Each had favored fields of application. Cables were best used for large circuit cross sections between highly concentrated terminating points. Satellites were uniquely capable of direct connection to widely scattered terminations. Each had vulnerabilities of very different kinds. Over the period covered, traffic on both continued to grow at a very high rate, and it became evident that cables were not destined to go away. Even before the first SF system was installed in 1970, AT&T began to consider a higher-capacity system to follow.

To meet the growing traffic demands, AT&T had proposed to install an additional SF cable to France in 1973, to be followed by a larger-capacity system at a later date. The Federal Communications Commission ruled against the SF installation and urged that the advanced system be expedited to as early a date as possible. Some exploratory work on the new AT&T system had been started in Bell Laboratories about 1966 to determine what the latest technology would permit. In 1971, AT&T embarked on a development program to place a system with a capacity of about 4000 circuits in service in 1976. Some of the pressure was relieved by a decision of the British Post Office and the Canadian Overseas Telecommunications Corporation to install Cantat-2, an 1800-channel system, between Widemouth in Cornwall, England and Halifax, Nova Scotia in 1974.

The ownership of all telephone submarine cables terminating in the United States was shared by the Bell System and the participating foreign partners, generally in some proportion to usage. Early in 1970, it was decided that the next design, called the SG system, would be a shared development as well. The initial agreement was with the British Post Office, which was to be responsible for the development of a new, larger-diameter, armorless cable. The Bell System was to develop and furnish the repeaters. Later, the French Ministry of Postal Services and Telecommunications agreed to accept responsibility for the multiplex and other terminal development. A formal agreement for the shared development was signed on May 2, 1973. The first SG system was to be installed between Green Hill, Rhode Island and St. Hilaire de Riez, France in 1976.

Like the SD and SF design, the SG proposal was for a single-cable, equivalent four-wire system. The initial capacity objective was set at 4000 3-kHz-spaced two-way channels with a top frequency of 29.5 MHz [Fig. 12-20]. The channel-noise objective was 1 pw per kilometer average in all channels, with no channel greater than 2 pw per kilometer. The cable diameter over the dielectric was increased to 1.7 inches (43 mm) to reduce the cable loss. The repeaters were spaced at intervals of 5.1 nautical miles (9.5 km) [Fig. 12-21].[14]

Until the SG system development, submarine coaxial-cable systems had lagged at least a generation behind the land-based systems, but, with SG, they were catching up rapidly. The first systems, SA and SB, were far smaller in capacity than the contemporary overland designs. SD, at 1 MHz in 1962, bore some resemblance to the first L carrier experiments of the 1930s. SF, at 6 MHz in 1968, came 15 years behind the 8-MHz L3 coaxial system and was virtually contemporary with the 18-MHz L4 coaxial land system. But the SG system at 30 MHz followed the 60-MHz L5 system by only a few years and, in many ways, was as, or even more, demanding on the technology.

Externally, the SG repeater was very similar to the SF repeater, but newer techniques were used inside to meet the needs of the higher top frequency and greatly increased bandwidth. Compact circuitry, using glass-epoxy etched circuit boards, provided the high-frequency capability and reproducibility required. The wide band and large number of repeaters placed especially difficult requirements on the equalization methods. These factors complicated the design of the passive networks and required a high degree of sophistication in the development and measurement of the cable dielectric. In principle, it would have been possible in the earlier system designs to have furnished shore-controlled adjustable equalizers, but the submarine-cable designers preferred hard, permanent connections. They distrusted undersea switches, fearing they might

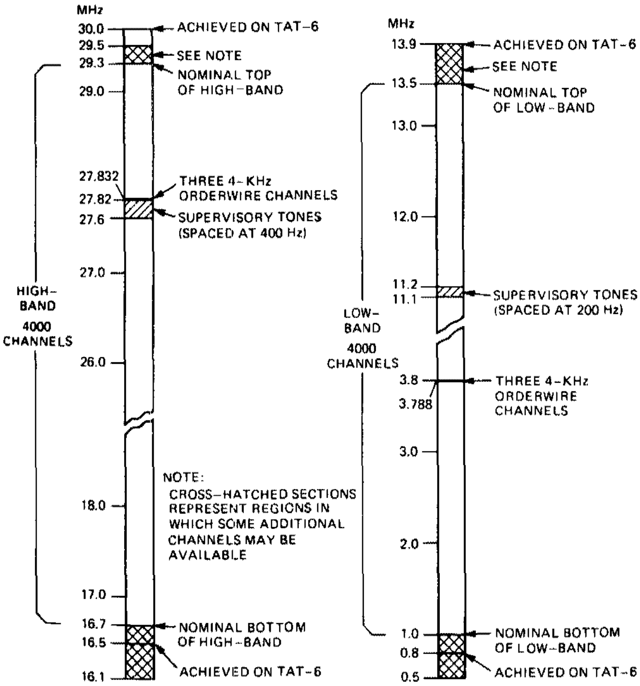


Fig. 12-20. The SG system frequency allocation.

do more harm than good if they failed to function properly. The equalization plan for SG followed the same pattern as SF, but a new equalizer was added late in the development to handle possible changes in loss over time. For the first time, this equalizer was controllable from a shore terminal station.

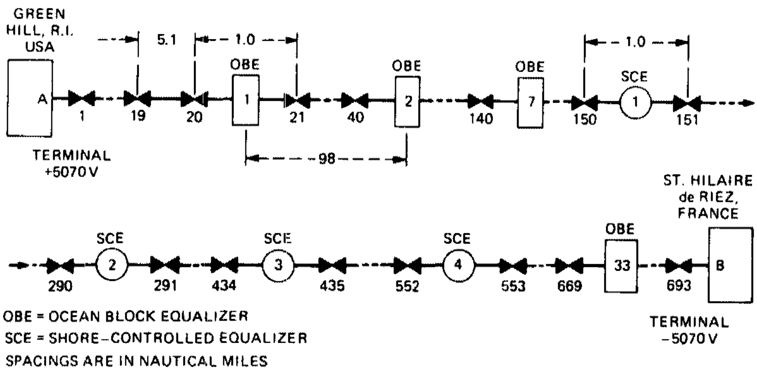


Fig. 12-21. Plan for the TAT-6 SG link.

The first SG system was installed between the United States and France on schedule and turned up for initial service as TAT-6 on July 27, 1976. A second SG system, TAT-7, was installed from the United States to England in 1983. By the end of that year, about 8000 nautical miles (14,800 km) of SG installations were in service. There was about a normal, but not excessive, number of problems in the development of the SG system. The first system did not quite meet noise objectives in the top third of the band. Attempts to improve the performance by circuitry at the terminals were only partially successful, but the deviation from objectives was not large, and the circuits were of quite high quality. The cause for the deviation from the design objective was analyzed and the problem eliminated in the design.

By 1984, it appeared almost certain that the SG system was to be the last major analog submarine-system design to be realized. However, before attention was turned to optical digital systems, some exploratory work was done on an analog successor, called the SH system. With a top frequency of 125 MHz, it would have been capable of transmitting about 16,000 channels. The repeater spacing would have been short (about 2-1/2 miles), and transmission problems were extremely tough, although perhaps not insuperable. The work on SH was dropped in favor of a lightwave program as optical technology advanced, and it became evident that a lightwave system was better matched to the needs of an increasingly digital communications world.

XI. PROTECTION AND SPECIAL FACILITIES

Up to this point, emphasis has been on the evolution of the sequence of cable systems of successively larger capacity to meet the increasing demand of overseas traffic around the world. The systems form a convenient framework on which to hang the story, but there were many other major technical problems to be faced in building up the worldwide undersea network.

11.1 Protecting the Cable

After the completion of TAT-1 in 1956, the cables worked very well and everyone was happy with their performance. However, in February 1959, a cable break off Newfoundland caused the first interruption of service. The nature of the break indicated that a fishing trawler had probably done the damage. After this initial incident, additional trawler breaks began to occur with appalling frequency. The intensive fishing activity in the shallow waters of the continental shelves and the impact of the rapid postwar development of large, powerful trawlers had not been fully appreciated. It was an expensive matter in at least three aspects: the direct cost and hazards associated with cable repair, the amount of time required for repair that resulted in lost revenue or the cost of leasing alternate circuits (this was by far the largest component of the loss), and the potential electrical damage to the system caused by rapid discharge of the cable. Repair time in particular became a major factor as the number of circuits increased. It did not take long to realize that a major offensive on the problem was required, and fast.

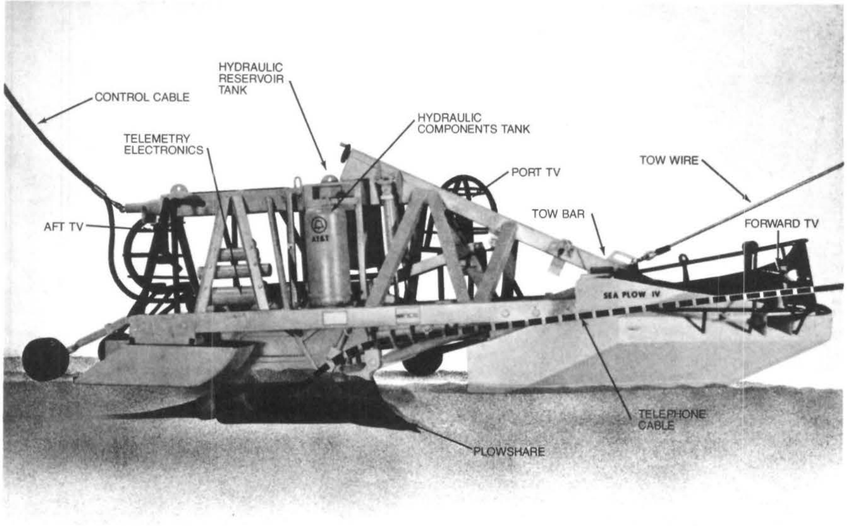
All options were reviewed. A plan to use redundant shore connections was given extensive study. This involved a Y in the cable and an underwater switching device about 200 miles offshore, but it was not implemented. Cable-laying procedures were reviewed to specify what had to be done to ensure that the cable lay on the bottom with no suspensions or tangles, especially following repairs. An underwater television survey of the cable in the shelf area was undertaken to see exactly how things stood. A large effort was also undertaken to educate fishermen by giving them cable locations and other information. Payment was offered if, when they snagged a cable, they would leave their trawling gear on the bottom. The compensation idea was not very popular. Fishermen are very conservative folk and regarded trawling gear in hand as worth more than compensation in the bush.

The most promising plan was to bury the cable beneath the ocean bottom in the hazardous areas, and an extensive program was undertaken to study methods for burying cables by water jetting or plowing. For a time in the 1960s, high-pressure jetting experiments produced some spectacular displays on the Bell Laboratories grounds, as spray and soil flew in all directions, but a limited amount of offshore testing soon indicated that the use of jetting was not a satisfactory solution to the burying problem. A number of bottom soils encountered were not easily moved by jets. Development work then concentrated on a plow. Starting in the early 1960s, several versions were built over the years as more was learned about the problems of ocean-bottom plowing.[15] One difficulty was that, as fishing trawlers grew larger and more powerful, breaks began to occur at greater and greater depths. A plow design, Sea Plow IV, completed in 1975, buried cable and rigid repeaters to a depth of two feet below the bottom during the laying operation [Fig. 12-22].[16] Sea Plow IV operated to a depth of 500 fathoms (914 m). In recent years, plowing has reduced trawler cable breaks to practically zero.

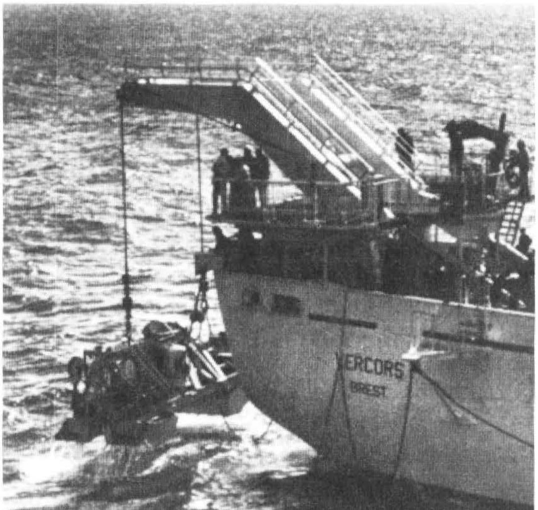
Long sections of existing cable could be recovered and relaid with the plow in shallow waters. At times, however, it was also necessary to recover a buried cable or bury a short length already on the bottom. For these and many other situations, a more versatile tethered, unmanned, submersible vehicle was developed. The Submersible Craft for Assisting Repair and Burial (SCARAB) was a highly maneuverable vehicle, equipped with television cameras and remotely controlled manipulators designed for these tasks [Fig. 12-23]. SCARAB also served as an excellent survey vehicle to examine the proposed routes for new cables within its operational depth of 500 fathoms.

11.2 Ocean-Bottom Simulation Facilities

Before the first transatlantic cable was laid, the designers knew that the ocean-bottom pressure and temperature would change the transmission properties of the cables slightly. These changes were offset in the repeater design and equalization plan, based on measurements from the same type of cable used in a military installation in the Bahamas and taken during deep-sea trials. In the two-cable system, the results were further refined, based on the experience with the first cable. However, there was still considerable apprehension over what would happen to the cable-transmission properties over a long time. With no access to intermediate points for adjustment, even a small change per unit length could accumulate to a degree where the signals would be out of the acceptable range.



(a)



(b)

Fig. 12-22. The Sea Plow IV. (a) The ocean cable and repeater canister passed through the plow and were buried about two feet below the sea bottom. (b) The Sea Plow IV being recovered from the sea by the French CS *Vercors*.

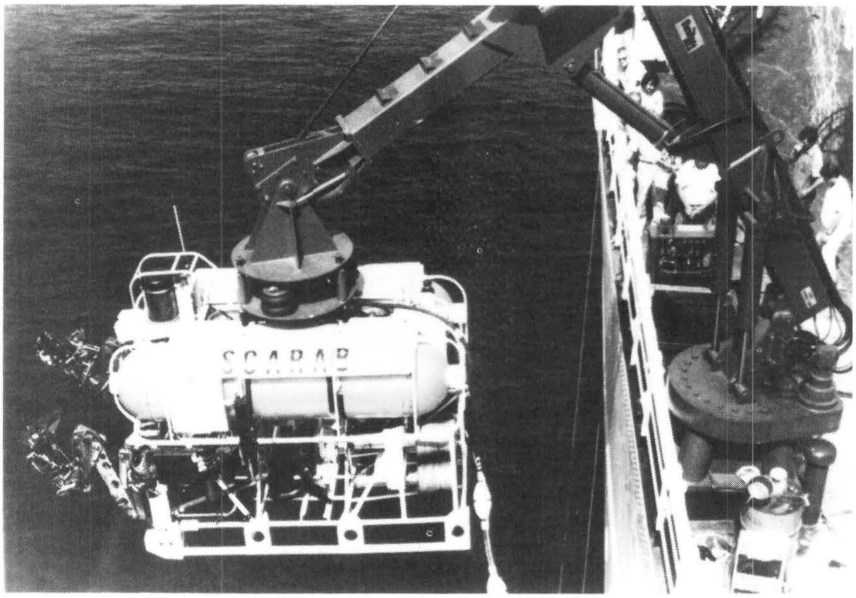


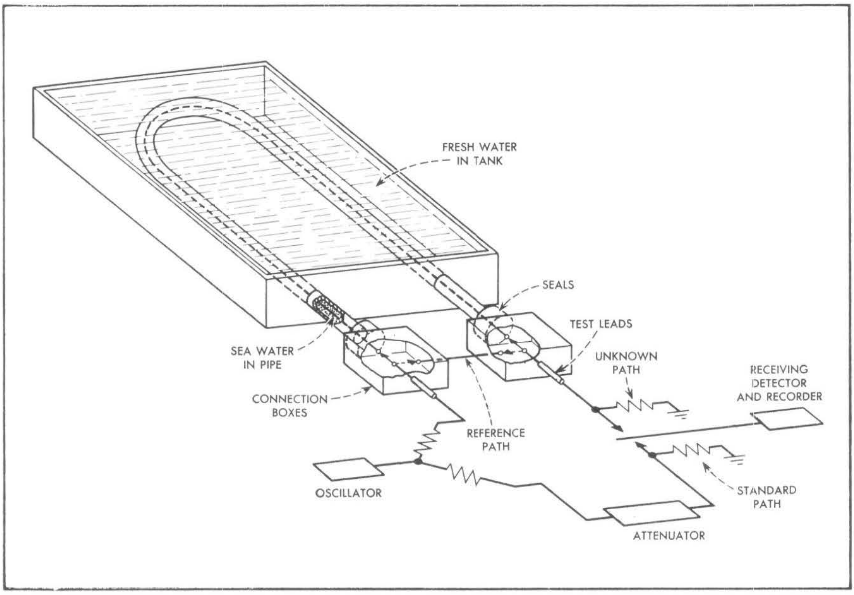
Fig. 12-23. The Submersible Craft for Assisting Repair And Burial (SCARAB). A maneuverable undersea tethered vehicle equipped with television cameras and remotely controlled robot arms and manipulators.

The initial effect of laying changed the loss by about two percent or over 70 dB for the total length. While this had been anticipated and offset to a large extent by adjusting cable lengths and terminal levels, there was not much margin left. Any appreciable further change in loss from the design value would hurt the signal-to-noise ratio.

The first system surprised and dismayed everyone by continuing to age toward lower loss. There was much head scratching, leading to the following hypothesis. The cable designers had been concerned that one-piece tubular copper conductors might break under the flexing and stresses of laying and recovery. They had therefore used composite inner and outer conductors. Under the constant high pressure, it was surmised that the resistance between the edges of the spiraled tapes was slowly decreasing, thus reducing the signal loss. The aging was very slow, but continuous. The designers held their breath, for a year or so, to see whether this would continue until some major corrective action would be necessary or if it would stabilize. Finally, the rate of change tapered off, and the loss stabilized at a value about 0.4 percent below the initial loss, within the range that could be handled by level adjustments at the terminals. With the assumption that the hypothesis was correct, all later-generation cables used a one-piece cylindrical inner conductor and a longitudinal-seam outer conductor.

These changes were successful, but the designers did not wait for a new cable to be installed to make certain they had cured the problem. A test facility

(a)



(b)

Fig. 12-24. Ocean-bottom simulation. (a) The artificial ocean at Chester, New Jersey. The facility consisted of two troughs, one 15 feet long, the other 315 feet long. Each trough contained pairs of pipes filled with salt water in which the temperature and pressure could be varied to simulate ocean-bottom conditions. (b) Testing arrangement for ocean-bottom simulation facility.

called an artificial ocean was built at Chester, New Jersey [Fig. 12-24].[17] The new armorless cable-loss was measured repeatedly to ±0.02-percent in the high-pressure and temperature-controlled pipes of this test bed for a period of over three years under sea-bottom temperature and pressure. No changes were detected. The development of a shore-controlled equalizer for the SD system was abandoned because of these excellent results. Later, the artificial ocean was used to measure pressure and temperature coefficients of the SF ocean cable. Because the SF cable design was similar to SD, with the outer conductor diameter scaled from 1.0 to 1.5 inches, the aging tests were not repeated. There were no appreciable aging problems with the armorless cable. With the Chester artificial ocean so successful, a more modern ocean-simulating facility was constructed at Holmdel, New Jersey to test SG coaxial and later lightwave ocean cables. In this ocean-bottom simulator, the instrumentation was more elaborate and computerized, and it was possible to vary the cable tension as well as the pressure and temperature to detect the effects, if any, of laying and recovery.

11.3 Summary

While this account has focused on the development and major initial installations of the Bell Laboratories designed systems, which were usually first

TABLE 12-1: Some Important Dates in American Undersea Cable History

1950	Florida–Cuba. First SA system.
1952	Negotiations start on TAT-1. Exploratory development of next generation.
1956	Scotland–Newfoundland (TAT-1) in service. First SB system. Final development of next system started.
1957–1960	France–Newfoundland (TAT-2), Washington–Alaska, Florida–Puerto Rico, California–Hawaii (HAW-1) (SB).
1959	Three-kHz-spaced channels used. TASI-A service. Cable-protection studies. First man-made cable break.
1961	United Kingdom–Canada (CANTAT-1). First transatlantic British system.
1962	Florida–Jamaica–Panama. First SD system.
1963	United States–United Kingdom (TAT-3, SD).
1964	California–Hawaii (HAW-2), Hawaii–Japan (TP-1), Florida–St. Thomas (SD).
1965	United States–France (TAT-4) (SD).
1967	First major Sea Plow system operated (Sea Plow-I).
1968	Florida–St. Thomas. First transistor system (SF).
1970	United States–Spain (TAT-5) (SF).
1974	California–Hawaii (HAW-3) (SF).
1975	Hawaii–Japan (TP-2) (SF). Sea Plow IV.
1975	TASI-B operational Vancouver–Sydney.
	United States–France (TAT-6). First SG system.
1980	St. Thomas–Venezuela, St. Thomas–Brazil (SF)
1981	Guam–Taiwan (SF). TASI-E operational.
1982	Florida–St. Thomas (SG). Greece–Egypt (SF). Light-wave demonstration—light-wave final development.
1983	United States–United Kingdom (TAT-7, SG).

installed across the Atlantic, many subsequent links were installed in the Pacific, the Caribbean, and to Central and South American points [Table 12-1].

In addition to the Bell System program, extensive development and manufacture was carried on in several other countries, including England, France, and Japan. With the exception of the first single-cable British systems, the Bell System projects set the pace in high capacity and advanced features. The increase in capacity by a factor of more than 100 depended critically on the progress in achieving high-frequency devices with the required reliability.

From an extremely cautious approach, with very conservatively designed vacuum tubes, designers moved to more advanced tubes, to germanium transistors without redundancy, and, finally, to high-performance silicon transistors with a common-emitter gain limit of 2.7 GHz. The 2.7-GHz figure was close to the limit of the state of the art in 1976, and not much different in performance from the highest-frequency devices used in land-based systems. In addition, it was designed with a reliability objective—one failure per 100,000 transistor years. No one expected any one transistor to operate that long, but the evidence so far, from the large numbers on life racks and in service, is that they are close to the objective.

REFERENCES

1. O. E. Buckley, "The Future of Transoceanic Telephony," *Bell Syst. Tech. J.* **21** (June 1942), pp. 1-19.
2. J. J. Gilbert, "Key West-Havana Cable No. 4," *Bell Lab. Rec.* **9** (May 1981), pp. 412-415.
3. J. J. Gilbert, "A Submarine Telephone Cable with Submerged Repeaters," *Bell Syst. Tech. J.* **30** (January 1951), pp. 65-87.
4. M. J. Kelly, Sir G. Radley, G. W. Gilman, and R. J. Halsey, "A Transatlantic Telephone Cable," *Proc. Inst. Elec. Eng. London* **102B** (March 1955), pp. 117-130.
5. See special issue on the TAT-1 SB Cable Design, *Bell Syst. Tech. J.* **36** (January 1957), pp. 1-326.
6. R. J. Halsey, "Cantat—A New Submarine Cable Telephone System to Canada," *Post Off. Elec. Eng. J.* **54** (January 1962), pp. 220-222.
7. R. S. Tucker, "Sixteen-Channel Banks for Submarine Cables," *Bell Lab. Rec.* **38** (July 1960), pp. 248-252.
8. E. F. O'Neill, "TASI," *Bell Lab. Rec.* **37** (March 1959), pp. 82-87.
9. R. L. Easton, P. T. Hutchinson, R. W. Kolor, R. C. Mondello, and R. W. Muise, "TASI-E Communication System," *IEEE Trans. Commun.* **COM-30** (April 1982), pp. 803-807.
10. R. D. Ehrbar, "Submarine Cable Laying System," *Bell Lab. Rec.* **42** (March 1964), pp. 86-93.
11. See special issue on SD Submarine Cable System, *Bell Syst. Tech. J.* **43** (July 1964), pp. 1367-1459.
12. C. D. Anderson, "Overload Stability Problem in Submarine Cable Systems," *Bell Syst. Tech. J.* **48** (July/August 1969), pp. 1853-1864.
13. See special issue on SF Submarine Cable System, *Bell Syst. Tech. J.* **49** (May/June 1970), pp. 601-798.
14. See special issue on SG Undersea Cable Systems, *Bell Syst. Tech. J.* **57** (September 1978), pp. 2313-2564.
15. "Burying Ocean Cables," *Bell Lab. Rec.* **45** (June 1967), pp. 186-191.
16. G. S. Cobb, D. L. Garren, and T. H. Rose, "Sea Plow IV: Digging-in the Newest Transatlantic Cable," *Bell Lab. Rec.* **54** (September 1976), pp. 220-224.
17. B. J. Kinsburg, "Sixteen Oceans at Chester, New Jersey," *Bell Lab. Rec.* **45** (October 1967), pp. 290-294.

Chapter 13

Satellites

I. INTRODUCTION

When Arthur C. Clarke, an engineer and scientist turned science-fiction writer, first proposed communication by satellites in a 1945 paper, no object, however small, had been placed in orbit, though Clarke was confident this was not far in the future.[1] He proposed that a radio-relay satellite be placed in an orbit about 22,000 miles above the earth's surface so that its orbital period exactly matched the daily rotation of the earth on its axis. If located in the plane of the equator, it would then appear fixed in position from the earth's surface. Transistors had not been invented, so it was necessary to assume electronics based on vacuum tubes. Microwave technology was in its infancy. Solar cells had not been invented; there was no efficient or convenient way to convert sunlight to electricity. Clarke thought a solar-powered steam plant might do the job. The station would be a complicated mechanism, and Clarke suggested that it would be prudent to have people in attendance. The primary market he addressed was television and radio broadcasting.

Nine years later, in 1954, J. R. Pierce* was asked to talk to an Institute of Radio Engineers (IRE) group in Princeton, New Jersey. Without having seen the Clarke paper, he wrote a remarkably similar paper, differing from its predecessor in two respects.[2] First, the needed technology had come to pass. Transistors, solar cells, and traveling-wave tubes had been invented and were in use. Second, Pierce's main interest was in point-to-point transoceanic transmission of both telephony and television. He considered both passive and active satellites and both low-altitude and geostationary orbits. His system parameters were remarkably close to those of the early communications satellites, and, as a matter of fact, his work guided the early development.

For Pierce, like Clarke, there was no launch-vehicle experience, but, on October 4, 1957, *Sputnik I* was launched from Tyuratam, USSR, and the space

---

* Pierce came to Bell Laboratories after receiving his PhD from California Institute of Technology in 1936. He was one of an array of Cal Tech and Stanford University graduates who were to transform the postwar world of electronics. In the 1950s, he became director of electronic research at Bell Laboratories. He was at the forefront of almost every field of research in telecommunication sciences, from acoustics and information theory to antennas and traveling-wave tubes.

age had arrived. Just four months later, on January 31, 1958, *Explorer I* went into orbit, and launch capability came closer to home. Pierce, like Clarke, wrote science fiction. Were their works more fiction than science or more science than fiction? As it has turned out, they were prophecy.[3]

II. PROJECT ECHO

Despite what would seem to be a head start in planning, the Bell Laboratories entry into satellite communications came about essentially by chance. In Pierce's paper, he had discussed both active and passive satellites, the latter being simple large objects with no electronics intended to reflect signals transmitted from the earth. In 1958, he and his colleague, R. W. Kompfner, saw a picture of a large inflated plastic sphere that the National Aeronautics and Space Administration (NASA) was proposing to place in orbit 1000 miles above the earth [Fig. 13-1]. NASA's purpose was to measure the density of the upper atmosphere by observing the rate of decay of the orbit, but Pierce and Kompfner recognized at once that the sphere was precisely what was needed for passive satellite communications. It could reflect (that is, scatter or reradiate) radio waves to such an extent that communication would be possible between any two stations on the earth's surface that could see the satellite at the same time. At a technical meeting, they called the balloon satellite to the attention of W. H. Pickering of the Jet Propulsion Laboratory (JPL), then a part of NASA. Pickering enthusiastically supported the idea of extending the tests to include communications. A proposal to use the 100-foot sphere of metallized plastic film for this purpose was written for JPL and was accepted as the basis for a joint NASA, JPL, and Bell Laboratories experiment. Bell Laboratories paid for the equipment it provided, and NASA funding paid for the operational phase.

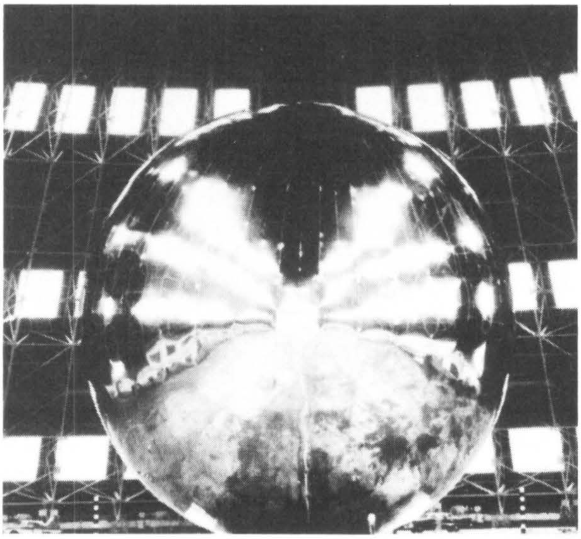


Fig. 13-1. NASA's *Echo I* satellite in a ground inflation test.

After one launch failure, the sphere, named *Echo I*, was launched in a circular orbit by a Delta rocket on August 12, 1960. Its altitude was about 1000 miles, giving it an orbital period of 118 minutes. Both JPL at Goldstone, California and Bell Laboratories at Crawford Hill, New Jersey were ready; and, on the very first pass through the area of mutual visibility, JPL transmitted a message from President Eisenhower that was successfully received at Crawford Hill. (See another volume in this series, *Communications Sciences (1925–1980)*, Chapter 5, Section 5.2.)

Ground-based equipment was provided for an east-to-west link at 960 MHz and a west-to-east link at 2390 MHz [Fig. 13-2]. The transmitting equipment at Crawford Hill consisted of a 10-kw klystron amplifier connected to a focal-point-fed, parabolic antenna 60 feet in diameter, arranged to radiate a circularly polarized signal. (The electric and magnetic fields of an electromagnetic wave are always orthogonal. In terrestrial transmission, the orientation is usually fixed, that is, the electric vector is either vertical or horizontal. By appropriate launching arrangements, however, the field vectors can be made to rotate through space in circular polarization. This was essential with the moving satellite, since transmission to it and reception from it would be from constantly changing aspects. Circularly polarized transmitters and receivers avoided the difficult alternative of polarization tracking.)

Separate transmitter exciters were provided for FM, AM, and phase modulation, and stable crystal oscillators generated continuous-wave test signals. Wide-deviation FM was normally used.

Receiving equipment at Crawford Hill consisted of a steerable horn-reflector antenna with an aperture of 20 by 20 feet [Fig. 13-3]. This was followed by a preamplifier, either a liquid helium cooled traveling-wave maser (maser was an acronym for microwave amplification by stimulated emission of radiation) or, less often, a parametric amplifier. (See another volume in this series, *Physical*

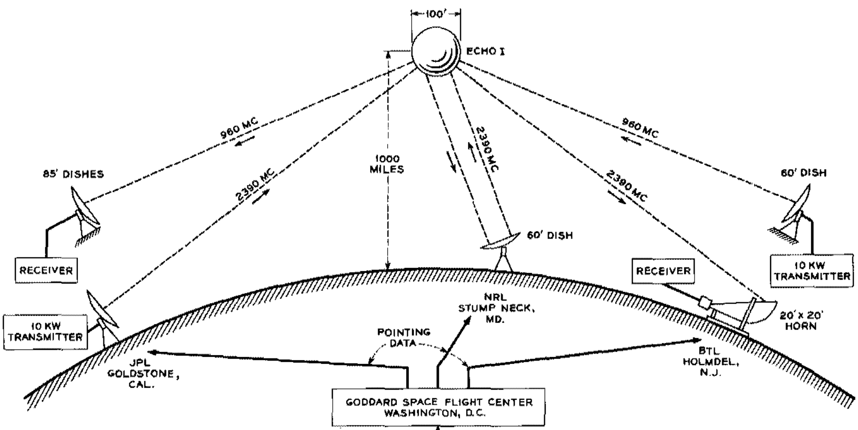


Fig. 13-2. General features of the Project Echo experiment.

Fig. 13-3. The Crawford Hill antennas for Project Echo. *Right,* the horn-reflector antenna was the antenna later used by A. A. Penzias and R. W. Wilson to detect the 3°K background radiation from the origin of the universe.

Sciences (1925–1980), Chapter 5, Section I.) The ultralow-noise receivers were the one element not included in Pierce's 1954 analysis; they greatly reduced the size required for the receiving antenna. The low-noise receiving equipment was mounted in a cab on the antenna, and the signal was carried to a fixed building for further amplification and detection. For FM signals, a threshold-extension receiver, invented at Bell Laboratories more than 20 years earlier by J. G. Chaffee, was used for the first time.[4,5] This detector, also called an FM-feedback receiver, tracked the instantaneous frequency of the incoming signal with a relatively narrow-band slot and thereby made possible the detection of FM signals several decibels lower in level than would have been possible with a full-band detector. Satellite positions were calculated by NASA and transmitted to Bell Laboratories for antenna pointing.

The noise temperature of the entire receiving system with the antenna at the zenith was predicted to be $18.90 \pm 3.00°K$ but was measured at $22.2 \pm 2.2°K$. At the time, it was thought that most of the measuring errors must have been in the positive direction. But a few years later, A. A. Penzias and R. W. Wilson, using the same antenna, made very careful noise measurements that led them to conclude that the temperature difference was real. The difference was interpreted to be radiation remaining from the big bang at the creation of the universe. This epic observation won them the Nobel prize in 1978. (See *Physical Sciences (1925–1980)*, Chapter 7, Section 1.2.)

Echo was the most visible of all satellites and was certainly in the public eye; predictions for viewing were printed in the daily newspapers. The communications experiments were successful and caused some public excitement, but the stir was not prolonged, because the satellite really did not do anything that the existing landline network could not do. (*Telstar* was to be different.) In general, the objectives of the experiment were met. Voice, music, and pictures were transmitted during 120 passes, which ranged in duration from 15 to 25 minutes. Some thought was given to a commercial system, using a larger satellite, larger antennas, and higher-power transmitters, but the tremendous penalty of two path losses in tandem directed attention to active satellites, where the path losses could be addressed one at a time.

Personnel for the Echo project came largely from the Electronics and Radio Research Department of Bell Laboratories, although specialists were enlisted from other areas, particularly for the maser, the parametric amplifier, and antenna steering. The experience of the Echo experiment was solid background for system planning and earth-station design for later satellite systems, including *Telstar*, which followed immediately.[6]

III. TELSTAR

All through the period when Bell Laboratories was involved with the Echo project (early 1959 through early 1961), studies of active satellites had continued.[7] Pierce and his associates concluded early that an orderly progression for research and exploratory development would involve first a passive satellite, then a low-altitude active repeater, and finally, an active repeater in a synchronous orbit. The first step would confirm the magnitude and stability of the loss of the radio path. (Clarke had noted that, while optical frequencies from the sun and stars were received with little loss, and very-high-frequency [VHF] and microwave signals were not returned to earth, there was no direct evidence that the radio signals would propagate through the atmosphere and ionosphere to a satellite.) A passive satellite experiment would advance the development of low-noise receivers and narrow-beam, low-noise antennas. The second step would bring in the problems of space-borne electronics and power. It would not require great sophistication in attitude stabilization and station keeping, which would be reserved for the third step.

When the Echo experiment concluded, Bell Laboratories felt obliged to collaborate with NASA on passive satellites; but, as early as May 1960, even before the launch of *Echo*, Kompfner wrote to NASA describing research toward an active satellite already under way in Bell Laboratories. The first proposal for an active system was issued on August 30, 1960, less than three weeks after *Echo* had been launched. This proposal called for the launch of an active satellite in mid-1962 to demonstrate real-time television and telephone transmission between the United States and Europe.

Following the success of *Echo*, imagination caught fire at AT&T headquarters, and, almost overnight, what had been little more than wishful thinking about active satellites at Bell Laboratories became company policy at AT&T. Financing and staff were set up quickly, and the development was off and away. The work began in the Research Area of Bell Laboratories under L. C. Tillotson and Kompfner. When it became apparent that a major project was materializing, development people were brought in, first under A. C. Dickieson, then a director in transmission development* and, after he had been promoted, under E. F. O'Neill. Specialized skills were drawn upon from almost every part of Bell Laboratories, some obvious (solar-cell people worked on solar cells) and some not obvious at all (the structural and heat-transfer designs of the satellite were done largely by vacuum-tube experts). A directory of those who were affiliated with the Telstar project contained 450 names. There is little doubt that the project at its peak occupied more people than any other project in the history of Bell Laboratories up to that time. But measured in technical staff-years, it was far down the list, because the whole thing was over so quickly.

The ground-based equipment in the United States was conceived essentially as a scaled-up version of the Echo station at Crawford Hill. Research by D. C. Hogg had identified a low-noise window in the atmosphere in the frequency range between 1 and 10 GHz [Fig. 13-4].[8] Hardware availability and the problems of coordination with other radio services led to a decision to operate in the common-carrier bands from 3700 to 4200 MHz and from 5925 to 6425 MHz, transmitting at 6 GHz and receiving at 4 GHz on the same antenna. The main antenna was to be a fully steerable horn reflector with an aperture of 3600 square feet.

Consideration of the availability of launch vehicles led to planning around the capabilities of the Delta rocket, the same vehicle that had launched *Echo*. With Delta, it appeared possible to put a satellite of about 150 pounds into an elliptical orbit, inclined to the equator, ranging in altitude from 500 to about 3000 miles. For the orbit planned, periods of mutual visibility of up to 30 minutes were projected between Eastern United States and Western Europe for at least some of the orbits. During these intervals, the object was to transmit either a full-band television signal or the equivalent of several hundred telephone channels over the Atlantic Ocean.

As a secondary but significant benefit, the Delta rockets were launched under a Bell Laboratories-designed guidance system run by a resident Bell Laboratories group. It was going to be necessary to have a rather complex installation at Cape Canaveral to test and certify the health of the satellite, right up to the launch moment. It was helpful to have an experienced team at the site who knew the terrain and the launch routines.[9,10]

Up to early 1960, NASA had shown no inclination to develop an active satellite of its own. AT&T was prepared to pay for the satellite development

---

* Dickieson joined the Western Electric Engineering Department as a technician in transmission systems development in 1923 and retired as Bell Laboratories vice president of Transmission Development in 1970. He contributed inventions and designs to an extraordinary range of developments, from the first feedback amplifiers and early overseas radio to satellites. As an executive in the post–World War II era, he directed a wide range of transmission projects, with especially important influence on the microwave radio program and the realization of the first pulse-code-modulation digital system.

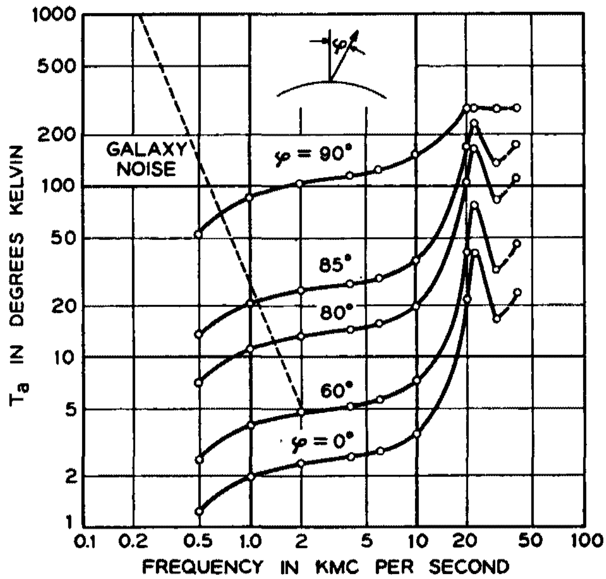


Fig. 13-4. Noise temperatures due to oxygen and water vapor in the atmosphere as measured at Bell Laboratories, Holmdel, New Jersey. The favorable frequencies are in the 1- to 10-GHz range. [Hogg, *J. Appl. Phys.* **30** (1959): 1417.]

and launch, as well as the earth station in the United States. The risk appeared to be high. There was no long line of candidates arguing that their satellite should come first; that was to come much later. Despite this, NASA regarded the AT&T initiative with something less than unrestrained enthusiasm. Up to this time, all nonmilitary space ventures had been its exclusive domain. Neither NASA nor anyone else in the government had figured out what role, if any, private enterprise was to have in space.

3.1 The Satellite

Telstar I was spherical, 34 inches in diameter, and weighed 175 pounds [Fig. 13-5]. The satellite was stabilized in space by spinning about its vertical axis at about 180 rpm. The axis was kept perpendicular to the plane of the ecliptic for optimum output from the solar cell array and favorable heat balance. The sequence of launch provided both the spin (from the spinning last stage of the launch vehicle) and the initial axis orientation. Later attitude corrections were possible by switching current through a magnetic torquing coil to interact with the earth's magnetic field.

The surface of *Telstar* carried 3600 silicon solar cells that were the source of its primary power. The cells were of a new n-on-p-type silicon variety, for longer life than the older type p-on-n cells when exposed to the radiation of the Van Allen belts. (See another volume in this series, *Electronics Technology,*

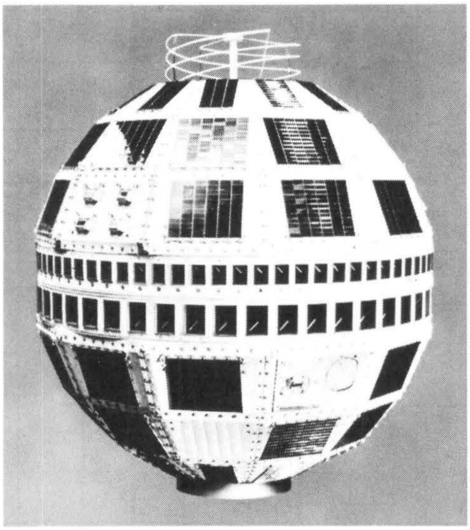


Fig. 13-5. The Bell System's experimental communications satellite, *Telstar I*, as it appeared before launch. Dark areas are solar cell arrays.

(1925–1975), Chapter 1, Section IX.) The solar cells covered 35 percent of *Telstar*'s surface to preserve the heat balance of the satellite, which was carefully designed to keep the internal electronics at 70 degrees F and to keep the solar cells near an optimum 30 degrees F. The solar cells generated 14 w of unregulated power when new and fully illuminated. Power was stored in 19 rechargeable nickel-cadmium cells to carry the peak demand of 35 w with the traveling-wave tube turned on. The power plant permitted half-hour operations and a total of 90 minutes per day.

Transmission to the satellite was at 6390 MHz and at 4170 MHz from the satellite to the earth. Since the satellite would be viewed from every aspect, the desired antenna pattern at both frequencies was made as nearly omnidirectional as possible; transmission in both directions was circularly polarized for the same reason. Separate antennas for receiving at 6 GHz and transmitting at 4 GHz consisted of equatorial belts of radiating boxes, each fed by a probe in one corner to achieve the circular polarization. The boxes were interconnected by a series of printed-circuit hybrids to a single 6-GHz input and a single 4-GHz output of the repeater in the electronics package.

In the repeater, the signal traversed a 6390 to 4170 GHz cross-band repeater (a transponder in satellite parlance), similar to a terrestrial microwave repeater with up- and down-converters, a 90-MHz intermediate-frequency (IF) amplifier, and a two-watt traveling-wave-tube amplifier at 4170 MHz [Fig. 13-6]. All locally generated modulation tones were crystal controlled. The traveling-wave tube was used in a reflex arrangement, in which it amplified a 4180-MHz beacon signal that was radiated for tracking and the two beat oscillator fre-

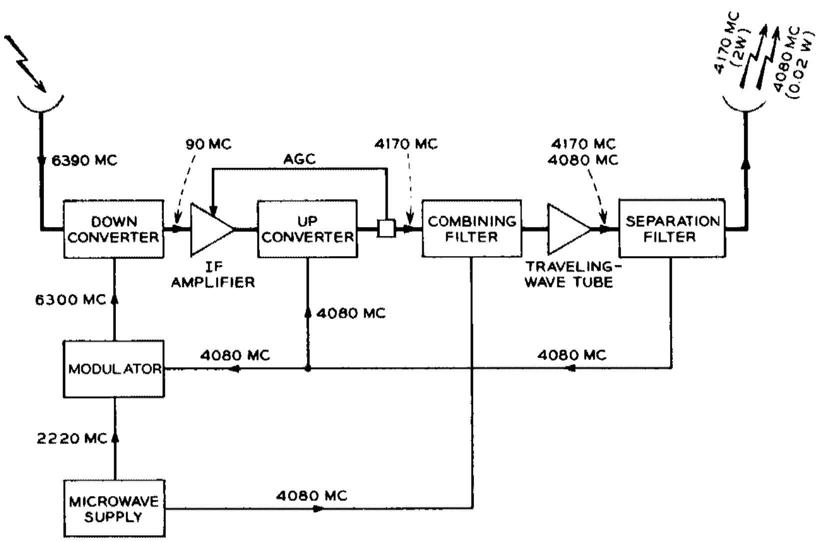


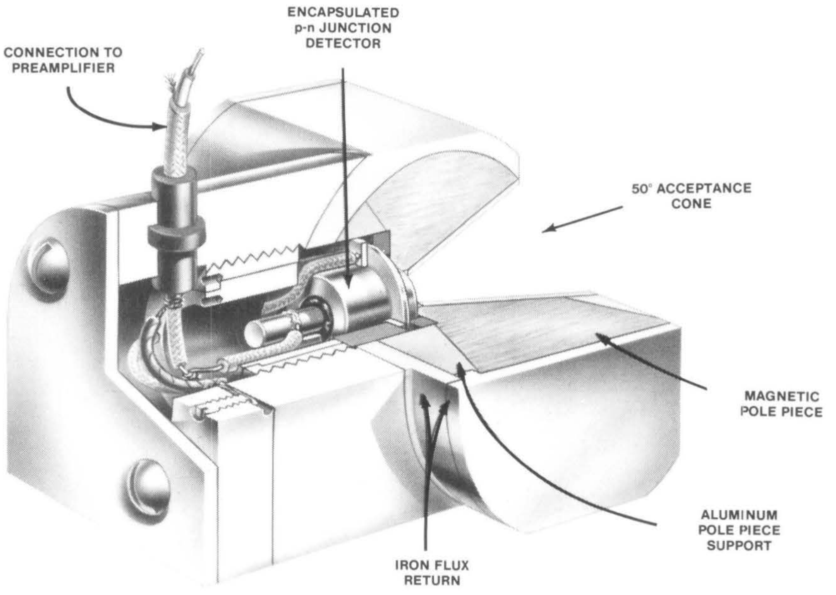
Fig. 13-6. *Telstar I* communications repeater.

quencies for down- and up-conversion, as well as the radio signal received from the transmitting earth station.

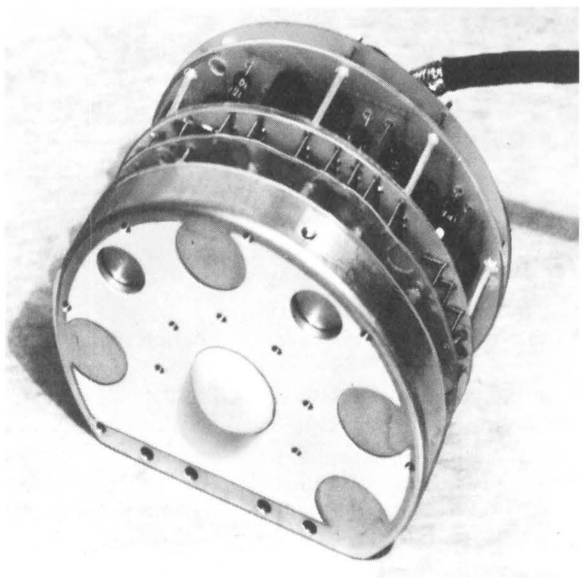
In addition to its purely communications equipment, the satellite carried a number of sensors and special devices to measure the electron and proton distributions at several energy levels in the region of space explored by the satellite and to give information on the cumulative damage to semiconductors caused by these particles [Fig. 13-7]. A near-omnidirectional helical antenna on top of the satellite received commands transmitted from the ground at 123 MHz to activate the various satellite functions. The same antenna transmitted a continuous 136-MHz signal from the satellite for initial acquisition and tracking. The 136-MHz signal was modulated for telemetry transmission of data from the radiation experiments, as well as for such housekeeping items as signal levels, temperature, and battery charge.

Except for the antennas, solar cells, and radiation experiment components, all of which were mounted on the outside surface of the sphere, the entire repeater was contained in a hermetically sealed cylindrical canister about 20 inches in diameter and 17 inches high [Fig. 13-8]. The canister was suspended by a web of nylon lacing to isolate it from high-frequency vibration during launch and from temperature variations caused by solar effects when in orbit. The primary structure of the satellite, on which the outer skin was mounted and from which the canister was suspended, was a series of trusses made of rectangular magnesium tubing [Fig. 13-9].

The waveguide components of the repeater would have been 13 feet long if straight and placed end to end. Obviously, many turns and convolutions were required to fit these components into the canister along with the traveling-wave tube, its power supply, converters, IF amplifiers, nickel-cadmium storage



(a)



(b)

Fig. 13-7. High-energy particle experiments on *Telstar I*. (a) Cutaway view of the low-energy proton detector mount. (b) The radiation-damage transistor assembly.

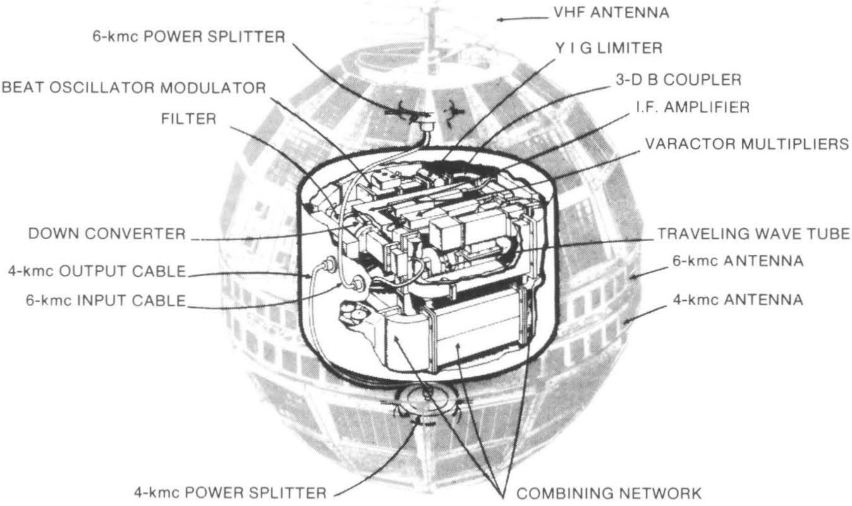


Fig. 13-8. The *Telstar I* electronics canister. All the active components of the repeater were in a hermetically sealed compartment 20 inches in diameter by 17 inches in height.

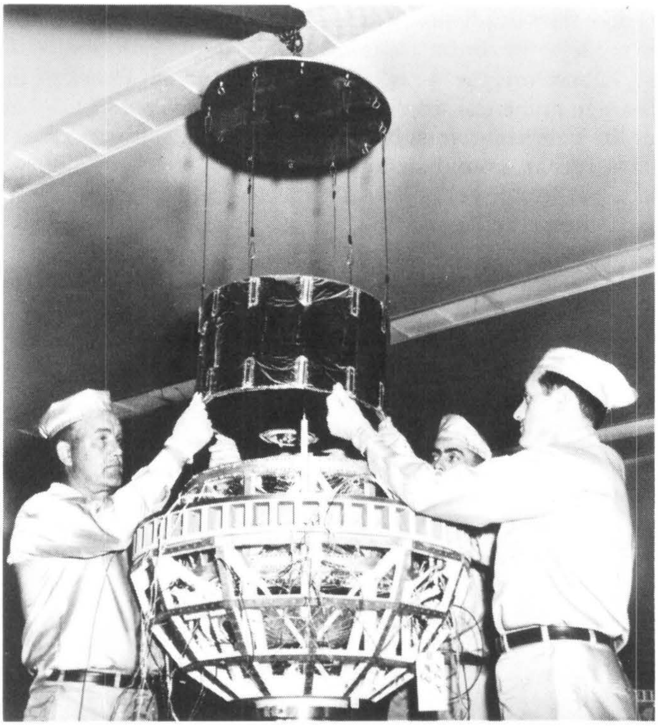


Fig. 13-9. Mounting of the electronics canister into the frame.

batteries, and the telemetry encoder. All told, over 6000 transistors and diodes were used (the solid-state era had arrived, but the integrated-circuit era was barely dawning); the only vacuum tube was the traveling-wave tube. Individual subassemblies were filled with polyurethane foam, and the entire canister was similarly treated. This gave the structural strength required to withstand the shock and vibration of a rocket launching, but reduced heat conduction, which had to be reinforced by metal straps.

3.2 Project Relay

On November 4, 1960, just as the *Telstar* development was gathering momentum, NASA announced that it was planning four launchings of communication satellites of its own, and it was clear NASA expected Bell Laboratories to bid. Considering the somewhat delicate state of relations between NASA and AT&T and considering NASA's monopoly on launching, Bell Laboratories decided to bid. When the NASA Request for Proposal was issued on January 4, 1961, a substantial part of the *Telstar* effort was temporarily diverted to proposal writing. NASA called for a satellite similar in size and orbit to *Telstar* (they used the same launch vehicle) but proposed that communications be at 400 to 500 MHz upward and at 2200 to 2300 MHz from the satellite to earth. Bell Laboratories was skeptical of the frequencies, so submitted three proposals, ranging from one fully responsive, to one based on *Telstar I* frequencies and using Bell Laboratories radiation-experiment package instead of NASA's. There were seven bidders; the contract was awarded to Radio Corporation of America (RCA). Bell Laboratories was ranked fourth, but much closer to the bottom than to the top in numerical score. "The best of the amateurs," Bell Laboratories was told at the debriefing. In subsequent negotiations, it was agreed that the United States Telstar ground station, as well as foreign ground stations that had agreed to participate in the Telstar project, would be adapted to work with the Relay satellites as well.

Relay I was launched on December 13, 1962, five months after *Telstar I*, and *Relay II* was launched on January 21, 1964. They lasted well, partly because the devices intended to turn them off automatically after six months failed, but they added little to the state of the art after *Telstar* stole the show.

3.3 *Telstar* Earth Stations

For transmission across the Atlantic Ocean, European partners were needed. Representatives of AT&T and Bell Laboratories visited prospective correspondents in England, France, and Germany and found them receptive. Informal agreements called for England to build an earth station at Goonhilly Downs in Cornwall and for France to build at Pleumeur-Bodou in Brittany. France elected to follow the Bell Laboratories earth station design and to secure key components from Bell Laboratories; the English decided to go on their own. There was some stir in NASA when it learned of this, because it was felt that all such negotiations should be handled by the government, despite the decades of bilateral dealings between AT&T and foreign communications entities on high-frequency radio and submarine-cable links.

AT&T immediately looked for an earth-station site as close to Europe as feasible to maximize mutual visibility, preferably shielded by hills, not near existing or planned radio-relay routes but not so far from centers of population that long end links would be required. Attention quickly focused on a bowl in the hills near Andover, Maine, and an 1100-acre site was acquired. The Andover station was perceived as a prototype to be expanded, if needed, into a commercial installation for a low-orbit satellite system. In the low-orbit concept, there would be many satellites, perhaps as many as 50, but few earth stations. The earth station would maintain continuous transmission by tracking with steerable antennas and switching from one satellite as it set to another via a second antenna as it rose.

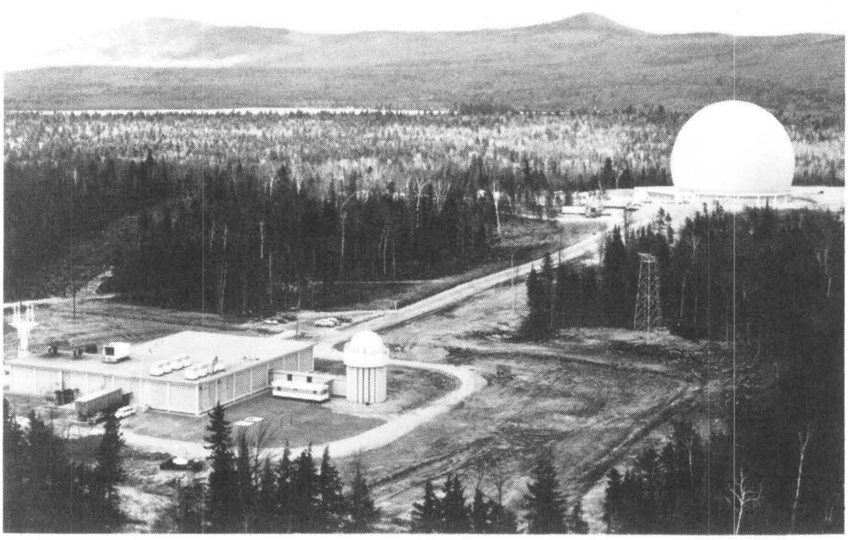
3.3.1 The Communications Antenna and Radome

The most conspicuous feature of the earth station was a huge inflated radome housing the large horn-reflector antenna [Fig. 13-10]. The antenna was designed to rotate in azimuth around a pintle bearing carried down to bedrock. Two circular tracks of precisely ground, heavy-duty crane rail carried the rotating azimuth structure via four trucks. The antenna proper consisted of a conical horn, a segment of parabolic reflector, and an aperture shield, all made of aluminum honeycomb, formed and glued to the desired contour. The elevation structure consisted of a 10-foot-diameter bearing at the small end of the horn and a 70-foot-diameter wheel at its large end [Fig. 13-11]. The antenna was rotated in azimuth and elevation by hydraulic drives. Its overall length was 177 feet and, with the equipment housed on it, weighed about 380 tons. Extreme mechanical precision was maintained throughout the structure. The absolute tolerance on the reflecting surface was only 1/8 wavelength at 6 GHz—about ±1/8 inch—the same as it would have been on a much smaller antenna. Since programmed tracking might have been required, the mountings had to provide accurate pointing of a radio beam only 0.2 degree wide.

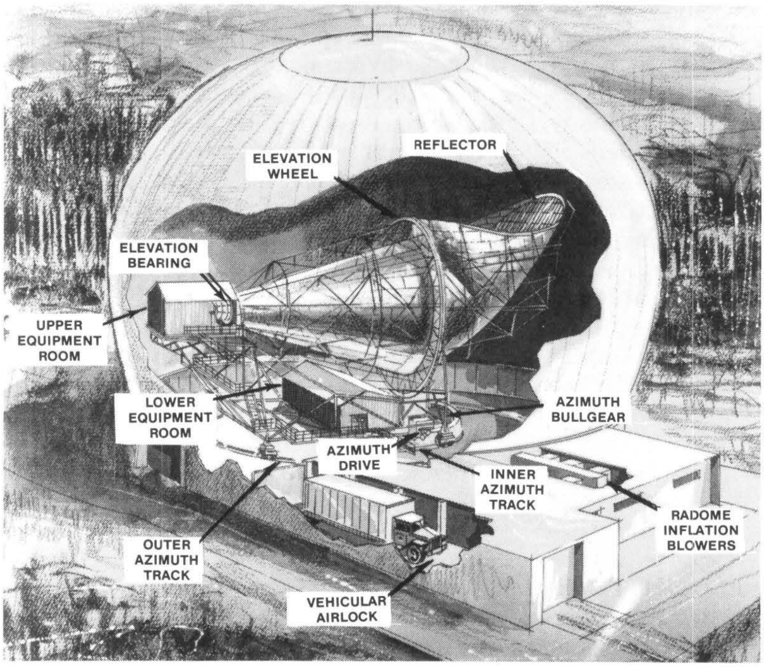
The horn-reflector antenna was a highly successful component in the Telstar project that turned out to be extraordinarily versatile. It was chosen for its low noise and because it was inherently broadband. Over the first few years, it was equipped to work not only with *Telstar* and *Relay* but with *Syncom* and *Intelsat I* (nicknamed *Early Bird*) as well. When it no longer was convenient to use the separate VHF command antenna, the horn was fitted for transmission and reception at 123 and 136 MHz. In all, it was ultimately fitted with transmission and tracking feeds at seven different frequencies ranging from 123 to 7360 MHz. It was even tested at 11 GHz, where its gain was found to be only 3 dB below the theoretical level.

The radome, 210 feet in diameter, was the largest purely air-supported structure built to that date. It was made of two layers of Dacron* coated with a synthetic rubber. There were six acres of single-ply fabric, 31,600 feet of hand-cemented joints, and a total weight of 60,000 pounds. An internal pressure of only 0.1 to 0.2 psi was sufficient to sustain it, even in a high wind. (At normal inflation, the fiber stress was about the same as in a 30 psi automobile tire.) The radome was to protect the antenna from distortion due to wind loading,

* Trademark of E. I. DuPont de Nemours and Co.



(a)



(b)

Fig. 13-10. (a) The Andover, Maine earth station. *Right,* the radome housed the communications antenna. *In the foreground* are the control building and the acquisition, tracking, and telemetry antennas. (b) The Andover horn-reflector antenna in its inflated radome.

Fig. 13-11. Placing the 50-ton elevation wheel into position during construction of the Andover antenna.

the accumulation of ice and snow, and uneven solar heating. Perhaps most importantly, it facilitated erection of the antenna on a short schedule in a northern New England winter.

The McKiernan-Terry Corporation, working under the supervision of Bell Laboratories people experienced in military antenna design, was prime contractor for the detailed design, fabrication, and erection of the antenna. Advanced Structures Corporation was responsible for the fabrication and alignment of the antenna surfaces, and Birdair Structures, Inc. furnished both a preliminary construction shelter (a multisection dome assembled with snap fasteners) and the permanent radome.

In the design of the electronics for the Andover station, the detectability of the received signal was the greatest concern, and every technique was pushed to the limit to assure successful reception. The receiver preamplifier was a traveling-wave ruby maser immersed in liquid helium at atmospheric pressure. Initially, the helium was in a batch-loaded Dewar flask, from which it was permitted to evaporate. But because helium is a scarce commodity (at least on Earth), a continuous closed-cycle cryostat was used later to limit the amount needed. The maser provided 42 dB of gain over a 16-MHz band, which was widened to 25 MHz by equalization in the following IF stages, and had a noise temperature of only 3.5°K. The noise temperature of the entire station, including the effect of the radome, was 33°K, with the antenna at a 30-degree elevation. Following the preamplifier and further amplification at IF, detection was by an FM-feedback detector, similar in principle to the one used for *Echo*. The Andover transmitters were of conventional design except that the final stage of the communications transmitter for *Telstar* was a specially designed 2-kw traveling-wave tube.

3.3.2 Acquisition and Tracking

The programmed tracking of *Echo* from predicted positions had not been completely satisfactory, so the antenna pointing and tracking for *Telstar* was designed to be self-sufficient. Several modes were provided [Fig. 13-12]. (1) A command tracker received the continuously transmitted 136 MHz beacon and could track to about ±1 degree. This was sufficiently accurate to turn on the microwave electronics by commands at 123-MHz. (2) A precision tracker was slaved to the command tracker. With the command tracker locked on to the 136-MHz beacon and the satellite electronics turned on, the precision tracker could acquire the 4080-MHz beacon and track to about ±0.01 degree. (3) The main antenna could be slaved to the precision tracker and kept on target close

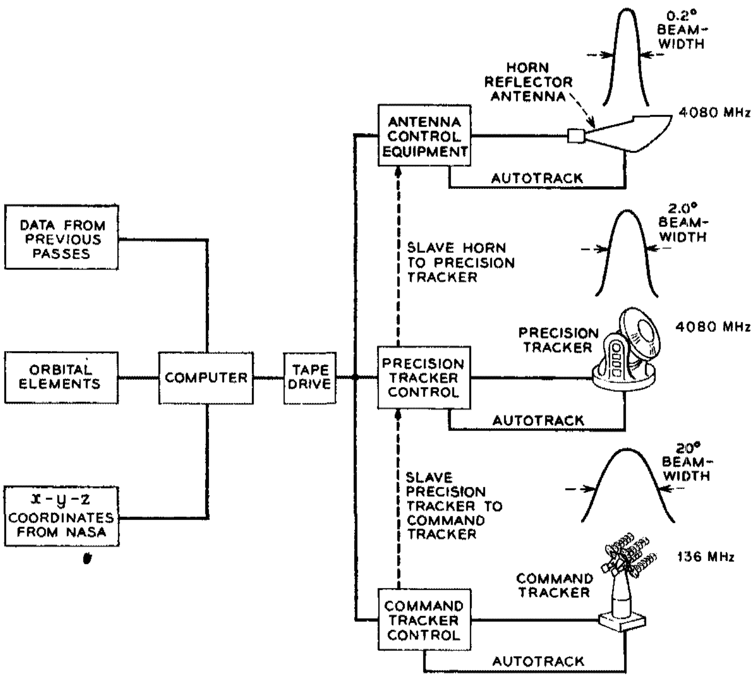


Fig. 13-12. The Telstar acquisition and tracking sequence. Initial acquisition was of the 136-MHz continuously transmitted beacon. The narrow-beam precision tracker and horn-reflector antenna then tracked the 4080 MHz microwave beacon switched on by the command tracker.

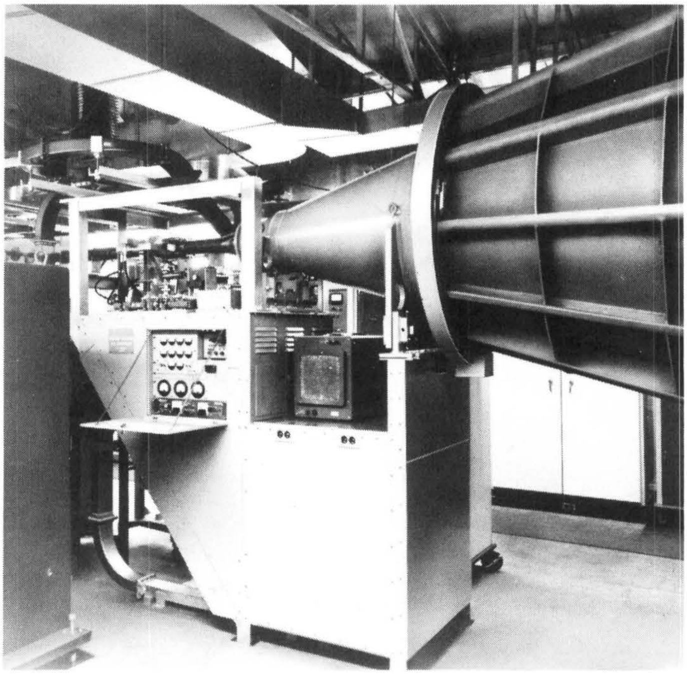


Fig. 13-13. The autotrack microwave-frequency unit in place at the horn apex, Andover, Maine.

enough for normal operation, but was also equipped for autotrack. (4) Autotrack of the 4080-MHz beacon by the big antenna was provided, using the principles of monopulse radar, in which a signal off-beam generated a different mode in the throat of the horn. This mode could be detected and used as an error signal. Nulling the error by means of servodrives brought the antenna to an exact bearing, within 0.005 degree [Fig. 13-13]. Except for a few seconds of acquisition at the start of each pass, the horn-reflector-antenna autotrack mode was used in practice. Design of the pointing system, including its digital and servofeatures, made use of Bell System experience in guided missile systems, computers, and microwave communications systems.

3.4 Telstar's Triumph

After an uneventful countdown, *Telstar I* was launched from Cape Canaveral by a Delta rocket at 08:35 Greenwich time (4:35 AM eastern daylight savings time [EDST]) on July 10, 1962 [Fig. 13-14]. The first five orbits did not offer usable mutual visibility between Andover and Europe, or, indeed, any significant visibility from Andover alone. There were snatches of telemetry from NASA stations around the world indicating that all was well, but not until the sixth orbit was there mutual visibility at a safe elevation.

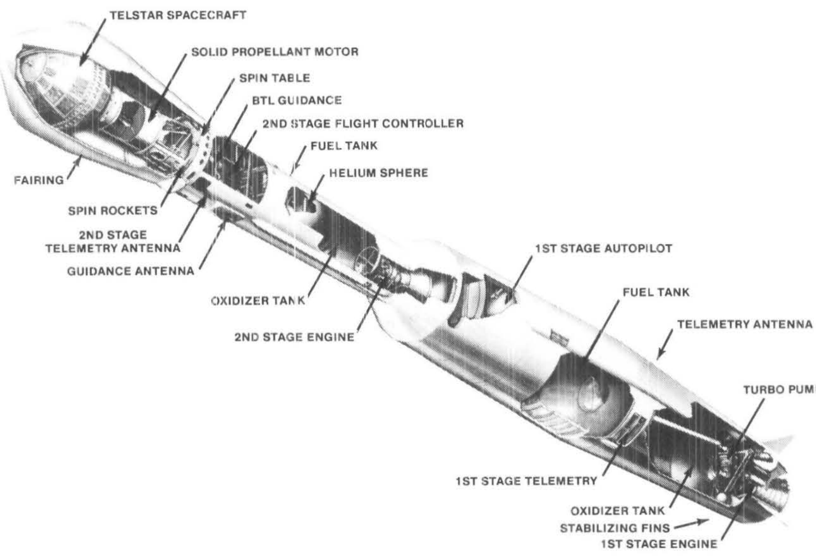


Fig. 13-14. Cutaway of the Delta rocket with the *Telstar I* spacecraft on board.

The reason for the long delay was that injection into orbit was toward the southeast, at the lowest orbit altitude (the 500-mile perigee) over the Atlantic near the equator. The satellite proceeded over the southern hemisphere, gaining altitude, then recrossed the equator to reach the 3000-mile apogee in the northern hemisphere, but over the Philippines on the opposite side of the world. The orbital period was about two and a half hours. A half orbit later, when it returned to the Atlantic Ocean, it was again close to perigee near the equator and could not be seen. It was not until the Earth slowly made a half turn under the orbit that the high-altitude apogee was over the Atlantic side. It was a long day for the Bell System crew at Andover, making last-minute adjustments, exercising their equipment, and watching the preparations to broadcast their success or failure to the world [Fig. 13-15].

Interest in the Telstar project was intense, especially in Washington, DC, where proponents of the Kerr-Magnuson bill and the Kefauver bill (pro and anti private enterprise in space, respectively) were debating the future of satellite communications at that very moment. Everyone wanted to be at Andover for the historic moment. But one thing had been overlooked there: no facilities were available to receive visitors. AT&T had no choice but to bring the show to Washington via closed-circuit television, and a large party of dignitaries was sitting in the Daughters of the American Revolution auditorium waiting for the button to be pushed. The event was also made available to the television networks via a pooled service, since there was not room for more than one network's cameras in the Andover control building. Never before had a privately sponsored technical event of such high risk been exposed to such intense public view.

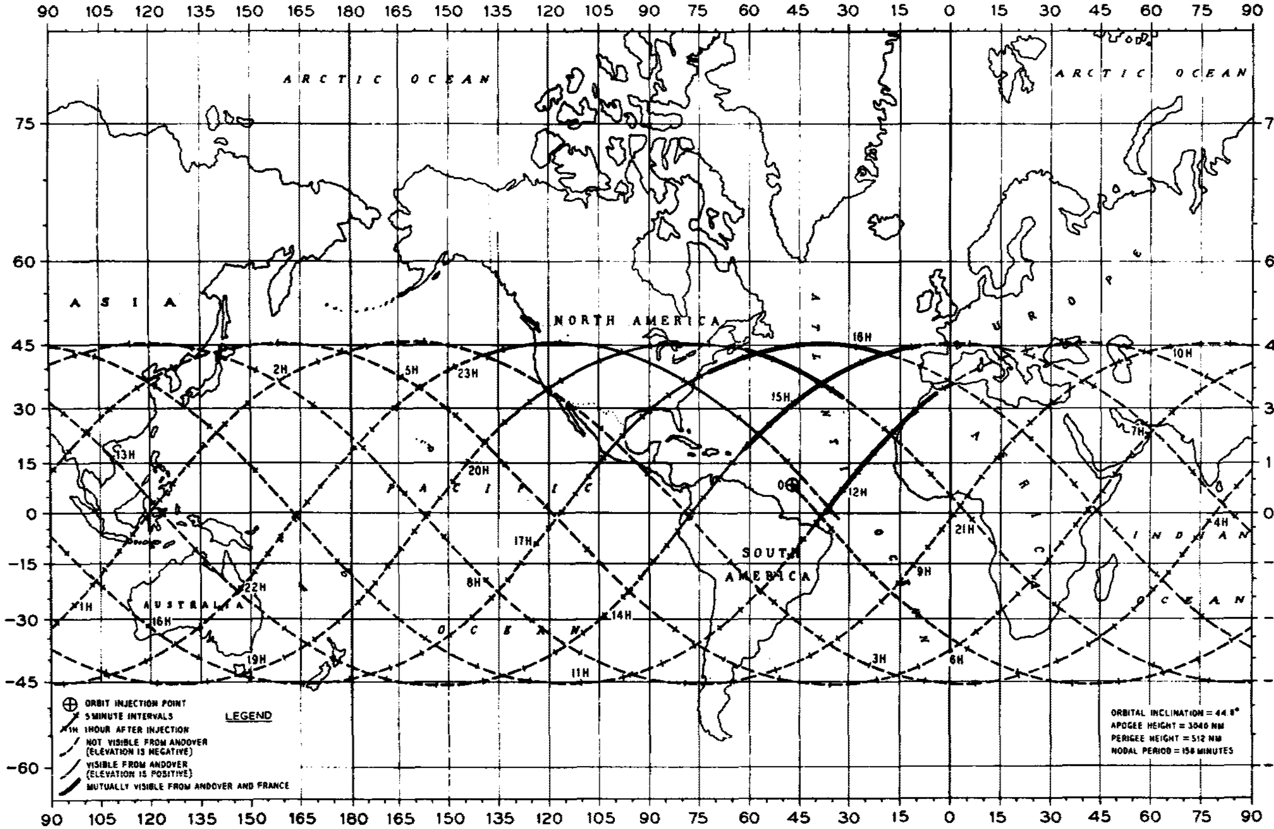
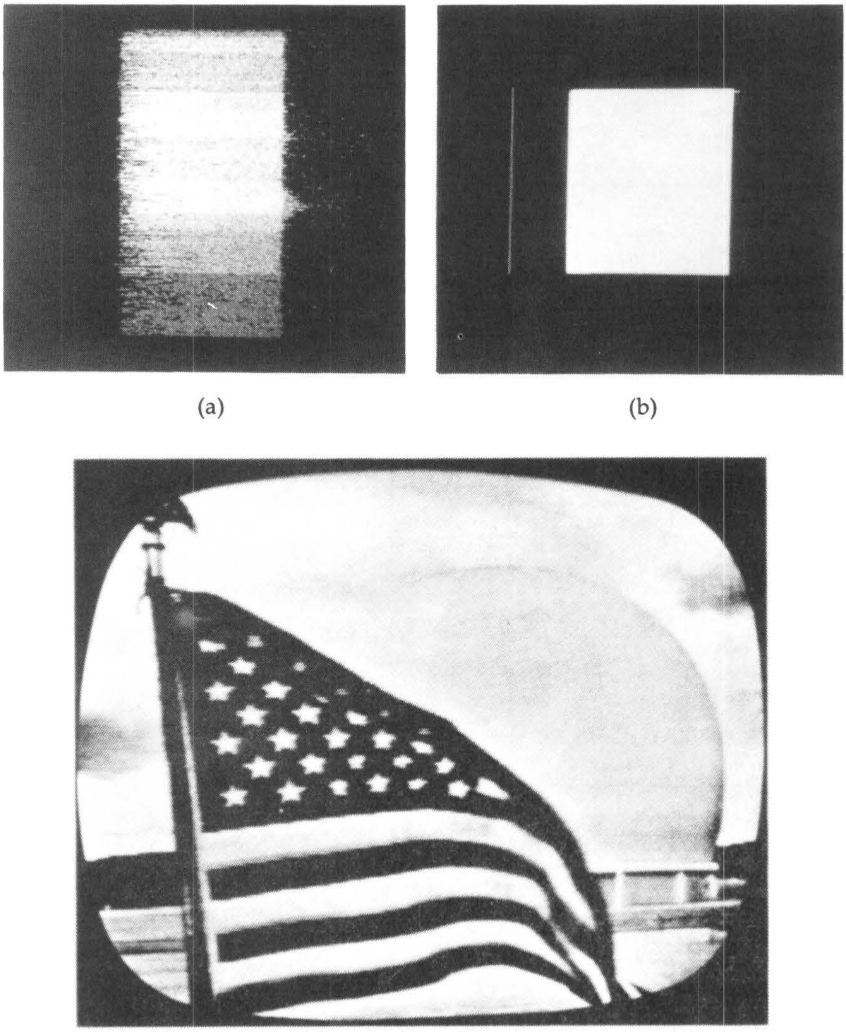


Fig. 13-15. Suborbital tracks of *Telstar I* for the first 24 hours. Portions with mutual visibility are shown by solid lines.

At 9:18 PM EDST, the Andover command tracker acquired the telemetry signal, and, six minutes later, the satellite was on, the precision tracker was in autotrack on the microwave beacon, and even the big antenna was in autotrack. At 9:25, there was a brief look at a television test signal, called sine square and window. It came through perfectly, just a vertical white line and a white square. At 9:30, F. R. Kappel, chairman of the board of AT&T, placed a telephone call



(a) (b)

(c)

Fig. 13-16. First video signals from Telstar satellite; pass 6, July 10, 1962. (a) Test signals, first synchronized frames. (b) Seconds later, a clear pulse and window. (c) The American flag in front of the Andover radome.

through *Telstar* to Lyndon Johnson, then vice president of the United States. At 9:33, television transmission started, with the results transmitted live to the Washington audience and the television networks. First there was a direct picture, not via satellite, of the American flag waving in front of the Andover radome. After a few seconds, this faded down and the picture via satellite faded up, appearing about a half frame to the right. When the picture transmitted by *Telstar* dominated, the television monitor synchronized on it, and it filled the screen. All engineers' eyes were on the *Telstar* picture. It looked exactly like the direct picture! [Fig. 13-16]

The French were ready and received the first pictures perfectly. The English received them poorly, due to a misunderstanding of the sense of the circular polarization (there can be a left-handed or right-handed helix in the electromagnetic corkscrew). But the next night, the English station was on line, and television signals from England and France were received at Andover and fed to the United States networks. In contrast to *Echo*, *Telstar* did something that had never been done before—the transmission of real-time television across the Atlantic. It is now hard to remember what barrier seemed to be broken, but, in those first minutes at Andover, there was no doubt that a communications milestone had been passed.

In the three months that *Telstar I* was available, extensive tests were made of the spacecraft and of the system. Telephone, television, and data transmission were measured. *Telstar I* was made available to all interested parties, and hundreds of different tests were made. *Telstar I* operated under an experimental Federal Communications Commission (FCC) license, so never earned a nickel. But it reaped immeasurable rewards in public interest and good will.

3.5 The Name

It is generally agreed that the name *Telstar* had more than a touch of genius. In the early stages, the satellite was referred to as *TSX-1*, but *Telstar* captured the romance of space and tied it to telecommunications. Who coined it? Nobody knows for sure. J. E. Dingman, then executive vice president of AT&T, remembered that someone from public relations gave him a list of several names and asked him to pick one, which he did. The success and publicity of the tests created a great public stir, perhaps because it was one of the first beneficial returns from the enormous space program and the final step to worldwide television coverage, or perhaps because it was the first space venture that required close international cooperation and did not appear to carry any element of threat. In any case, the name caught on and was used for everything from newstands in Rome to popular songs in the United States. For years, communications satellites were commonly known as Telstars. This brought a rueful smile to the Telstar crew, who had been pretty much legislated to the sidelines, while others carried the technology forward.

3.6 Starfish—The Effect of a High-Altitude Nuclear Explosion

On July 9, 1962, the day before *Telstar I* was launched, the United States exploded a nuclear device called Starfish at a high altitude over the Pacific. The Telstar people were not overly concerned at first, because the explosion was expected to be relatively clean, and products dangerous to *Telstar* were expected to disperse very quickly. In fact, however, a great many high-energy particles from the explosion were captured in the Van Allen belts, where they decayed very slowly. *Telstar,* which traversed part of the belts on every orbit, received in days the radiation damage anticipated for a year.

To *Telstar's* credit, it was on the spot to measure these products, and, with its radiation detectors, was equipped to do so. It also was able to measure the effects of Soviet explosions in space in the fall that same year. The heavy radiation damage to *Telstar* was ultimately fatal although not right away. Transistors in the command decoder were damaged, and *Telstar I* began to falter in its response to commands on November 18, 1962, four months after launch. For a while, it was possible to subvert the failed commands, but, by February 21, 1963, the satellite failed permanently to respond to commands.

A second satellite, *Telstar II,* was launched on May 7, 1963, and operated until May 1965; but, by then, AT&T had been denied a place in the international satellite business, and synchronous satellites were in orbit.

3.7 Some Hard Knocks

Was all this accomplished with no untoward incidents? By no means! Some account of a few of the many perils may convey some of the flavor of those hectic months.

When construction of the first satellite was almost complete, the communications repeater-transponder was found to be working fine when upright, but would not function when turned upside down for balance tests. Surgery into the already foamed and sealed canister revealed a tiny wire clipping in a microwave cavity. It was awesome to consider the possible consequences, if the satellite had been launched with this defect.

At Andover, during the winter preceding the launch, a heavy wet snow began to accumulate and depress the top of the radome. The heating, designed to melt off just such an accumulation, merely caused it to run into the center, depressing the dome further. The solution finally lay in several shots from a rifle, permitting the dimple to drain into the interior. At a critical point, the internal air pressure regained the upper hand and the dome popped back into position. When this happened, several tons of wet slush avalanched off, demolishing a trailer parked alongside!

The French were less fortunate. Partway through the construction, their temporary radome blew away in a gale. By good luck, the identical construction shelter at Andover had just been replaced by a permanent radome, and the Andover shelter was flown to France to complete their job. The French had other woes. Only a few days before the launch, as the antenna was being maneuvered for the first time, a panel fell from the parabolic reflector. And, even closer to the moment of truth, only minutes before the first transmission, a test meter poised on a cabinet toppled into an open equipment drawer, smashing the vacuum tubes of the antenna-position control circuit. It is widely believed that the French broke the world's record for electronic repair in the moments that followed. All in all, however, the Telstar team had reason to feel that the goddess of fortune smiled on their efforts.

IV. COMSAT, LOW ORBITS, AND SYNCOM

On February 7, 1962, President Kennedy sent a proposed bill to Congress that eventually became the Communications Satellite Act of 1962. After intensive debate between pro- and anti-government ownership factions, the bill was passed on August 17, 1962, just over a month after many members of Congress had watched *Telstar I*'s activation. The bill established the Communications Satellite Corporation (Comsat) as a quasi-private corporation with a United States monopoly in international communication by satellite. Comsat carried out its function as systems manager for the International Telecommunications Satellite Consortium (Intelsat). AT&T was thus out of the satellite business and rapidly wound down its satellite staff. The Andover earth station was sold to Comsat.

In the Washington debate, the main issues were the ownership and operation of this exciting new technology, but this had become entwined with questions of which engineering approach would produce the best system at the earliest date. The Bell System position had been stated succinctly in an internal memo of July 7, 1961 by W. A. MacNair, who was then vice president of Switching and Transmission Development at Bell Laboratories.

> Broadly, we have been discussing some of the pros and cons of 24-hour vs. random orbit satellite systems. In this argument, we would favor taking a position that allows us complete freedom to change our emphasis from the random orbit system to a 24-hour system. The results of current tests of the effects of long delays on telephone customers, results of other analyses being carried out, our own experimental results and the growth of space technology generally could cause us to reverse our emphasis in the future.

However, in the heat of a political debate, the public statements of the contending parties were stripped down to a simple advocacy of a position. The engineers' desires to search in the gray areas for the best engineering compromise did not adapt well to the adversary-advocate confrontation. The calamities attending the space program in the late 1950s and early 1960s led Bell System planners to underestimate the rate of progress in rocket capability and spacecraft control and to a feeling that the achievement of a long-lived satellite in synchronous orbit lay many years in the future. (They were not the first to misjudge the progress of space technology. In his 1945 paper, Clarke noted that in the 1940s German engineers, aware of the V-2 rocket program, had foreseen the possibility of a synchronous satellite, but estimated it would take 50 to 100 years to achieve.)[11]

The Bell System position became firmly established in Washington minds as espousing the low-altitude system, not merely for initial operation but also as the preferred long-term solution. Bell Laboratories entered into a contract, jointly with RCA Aerospace, for a preliminary study of a low-orbit satellite system for Comsat; but, right in the middle of the work, *Syncom II* and *III*, developed by the Hughes Aircraft Co., were launched to pioneer the stationary orbit. The joint contract with RCA was carried through to completion, but the low-orbit/synchronous issue was settled.

V. DOMESTIC SYSTEMS

On March 3, 1966, the FCC issued a Notice of Inquiry, designated Docket 16495, which was to occupy satellite experts, both within and outside the Bell System, for many years. The Commission asked various questions—legal, technical, and sociological—concerning possible domestic satellite communications systems. They did not ask for system proposals, but two unsolicited proposals came in, whereupon the FCC invited proposals from all who were interested. Several proposals were made, but the FCC did not act until March 4, 1970, when a Report and Order was issued, finally winding up Docket 16495 and adopting essentially an open-skies policy for domestic systems.

By that time, the state of the art in satellites was exemplified by *Intelsat IV*, designed by Hughes Aircraft Co. for Comsat. *Intelsat IV* carried 12 transponders, each with a 5-watt traveling-wave tube capable of transmitting 1200

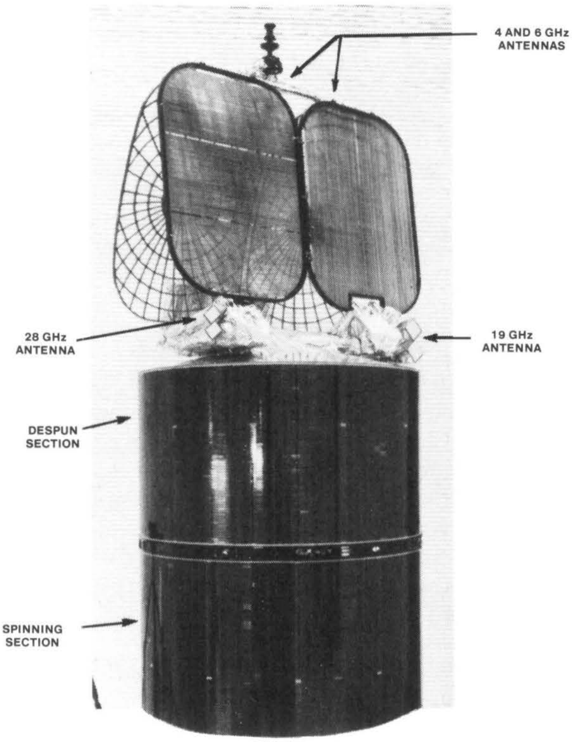


Fig. 13-17. The *Comstar* satellite. The lower solar-cell-covered cylindrical portion was 9 feet high by 8 feet diameter. The overall height was 20 feet. Screens in front of the parabolic reflectors improved polarization discrimination.

telephone channels or a television signal, and each duplicated for reliability. Vigorous work was under way, on both satellite electronics and launch-vehicle capability, to raise the transmission capacity significantly.

In early 1970, Comsat and Hughes approached AT&T, who agreed, in August 1970, to rent from Comsat three satellites in orbit and a spare on the ground, for just over one million dollars per satellite in orbit per month. The satellites were to be launched by an improved Delta rocket. Like their predecessors, they were to operate in the 4- and 6-GHz bands, but beacons were added at 19 and 28 GHz to test the transmission at those frequencies. Twenty-four 1200-channel transponders were provided, 12 with vertical and 12 with horizontal polarization. Switches were included to manipulate front ends in case of failure [Fig. 13-17]. Finally a contract was agreed upon, to be effective as soon as the FCC granted AT&T's application, which was filed on October 19, 1970.

The time for domestic satellites had arrived. There were several applicants, and there was much pushing and tugging among them. RCA Communications, Inc. proposed to give service to RCA Alascom, which RCA then owned. AT&T felt that their satellite should also be able to serve Alaska. Before the debate was over, the Bell System's satellite had antenna feeds for Hawaii, Puerto Rico, and Alaska, in addition to those covering the 48 contiguous states. Halfway through the debate, General Telephone and Electronics (GTE) proposed that AT&T and GTE build a joint system, and AT&T agreed to amend their plan to include GTE. Comsat would own and operate the satellites, while AT&T and GTE would build and own earth stations in their areas of operation.

5.1 Comstar

On September 12, 1973, the FCC granted several applications for satellite systems, including the AT&T/GTE/Comsat proposal. The contract with Comsat went into effect, and it was possible to contract for earth stations. Ultimately, eight earth stations worked with the satellites, named *Comstar* by Comsat, their owner. All stations used 30-meter parabolic antennas, met a 70°K receiving-

TABLE 13-1: *Comstar* Earth Stations

Location	Area Served	Owner
Hawley, Pennsylvania	New York City	AT&T
Hanover, Illinois	Chicago	AT&T
Woodbury, Georgia	Atlanta	AT&T
Three Peaks, California	San Francisco	AT&T
Homosassa, Florida	Tampa/Miami	GTE
Triumfo Pass, California	Los Angeles	GTE
Sunset, Hawaii	Honolulu	Hawaiian Telephone*
Cayey, Puerto Rico	Puerto Rico	American Cable & Radio**

* Owned by GTE.

** Owned by ITT but expropriated before service began.

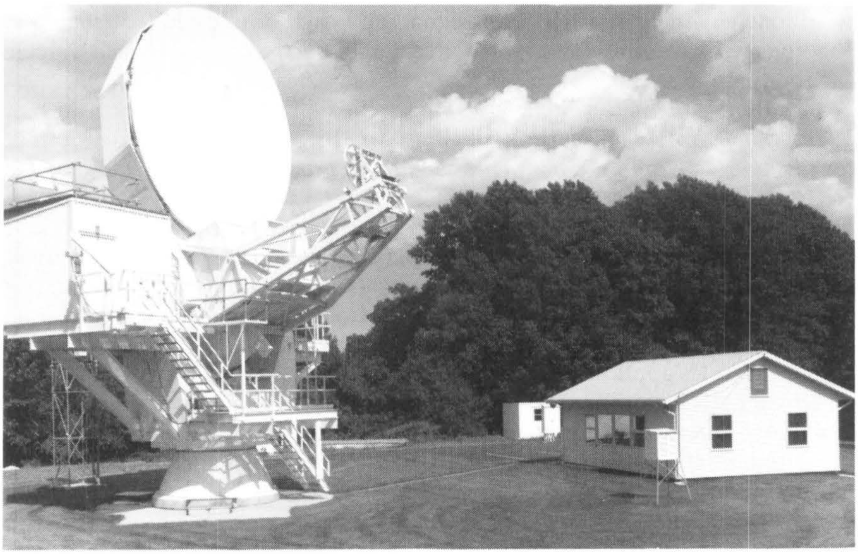


Fig. 13-18. The Bell Laboratories 7-m diameter offset Cassegrainian antenna at Crawford Hill, New Jersey, used to measure 19- and 28-GHz beacons from *Comstar*.

noise temperature requirement, and were equipped with 3-kw klystron transmitting amplifiers [Table 13-1].

Monitoring of the 19- and 28-GHz beacons was done at Bell Laboratories in Crawford Hill, New Jersey and with field receivers dispatched to especially rainy areas [Fig. 13-18].

The first *Comstar* satellite was launched successfully from Kennedy Space Center on May 13, 1976. A second satellite was launched on July 23, 1976, and the system was put into service on that day. A third satellite was launched in June 1978.

The *Comstar* satellites were designed to carry positioning fuel to keep them on station for seven years. Experience with *Comstar's* ancestors, *Intelsat IV*, led to an expectation that the electronics would also last as long, although there was some doubt about the batteries. The satellites performed about as expected, but, because of transponder failures, Comsat was asked to launch the fourth *Comstar*, which went up on February 21, 1981. It replaced the first *Comstar*, and the old satellite was collocated with the second to operate as a composite spare.

REFERENCES

1. A. C. Clarke, "Extra-Terrestrial Relays," *Wireless World* **51** (October 1945), pp. 305-308.
2. J. R. Pierce, "Orbital Radio Relays," *Jet Propulsion* **25** (April 1955), pp. 153-157.
3. J. R. Pierce, *The Beginnings of Satellite Communications* (San Francisco, CA: San Francisco Press, 1968).

4. J. G. Chaffee, U.S. Patent No. 2,075,503; filed March 26, 1936; issued March 30, 1937.
5. J. G. Chaffee, "The Application of Negative Feedback to Frequency Modulation Systems," *Proc. IRE* **27** (May 1939), pp. 317-331.
6. See special issue on Project Echo, *Bell Syst. Tech. J.* **40** (July 1961), pp. 975-1233.
7. J. R. Pierce and R. W. Kompfner, "Transoceanic Communication by Means of Satellites," *Proc. IRE* **47** (March 1959), pp. 372-380.
8. D. C. Hogg, "Effective Antenna Temperatures Due to Oxygen and Water Vapor in the Atmosphere," *J. Appl. Phys.* **30** (September 1959), pp. 1417-1419.
9. See special issue on Telstar, *Bell Lab. Rec.* **40** (April 1963), pp. 116-181.
10. See special issue on The Telstar Experiment, *Bell Syst. Tech. J.* **42** (July 1963), pp. 739-1908.
11. See reference 1.

Chapter 14

Mobile Radio

I. THE EARLY YEARS (1925–1945)

By the mid-1920s, the long-distance wireline network had evolved to the point where it was reliable, linked most cities and towns, and had good voice quality. Broadcast radio was well established, and overseas radio was in advanced development. It seemed only reasonable to extend telephone service to customers on the move by mobile-radio service.

1.1 Ship to Shore

For several reasons, ship communications were the first application of mobile radio. This was the heyday of the transatlantic ocean liner, and on each ship was a large body of (usually affluent) prospective users who would be otherwise incommunicado for a week. Furthermore, each ship was large enough to have a full-time radio operator on board. The scale of the ship environment was also suitable for the practical constraints of size, weight, power dissipation, and antenna length, which the early vacuum-tube technology and the low radio-frequency operating band imposed. Finally, the fundamental work to establish transatlantic point-to-point communications was providing the technological base required for ship-to-shore radio.

The earliest experimental service was to coastal steamers between Boston and Baltimore starting in 1919.[1] The ships *Ontario* and *Gloucester*, and later the *America*, were outfitted as mobile stations able to communicate with base stations at Green Harbor, Massachusetts and Deal, New Jersey [Fig. 14-1]. Service began at 750 kHz, and testing continued until 1922. Uncertainties as to the market, cost, and long-term frequency allocations created a seven-year hiatus, but, in 1929, commercial service to ships on the high seas was inaugurated. This first commercial service used channels at 4.2 and 8.7 MHz, with a transmitter power of 1000 w at the shore base and 250 w from the ship. Modulation was double-sideband AM. The first ship equipped was the SS *Leviathan*, the largest ship in the world at the time, and the first shore stations were at Ocean Gate and Forked River, New Jersey.[2]

Soon after the high-seas mobile-radio service was begun, a service using similar equipment was inaugurated for coverage of busy harbors and major coastal areas.[3,4] Frequencies allocated were in the band from 2 to 3 MHz (the so-called shortwave range). This coastal and harbor service developed over the

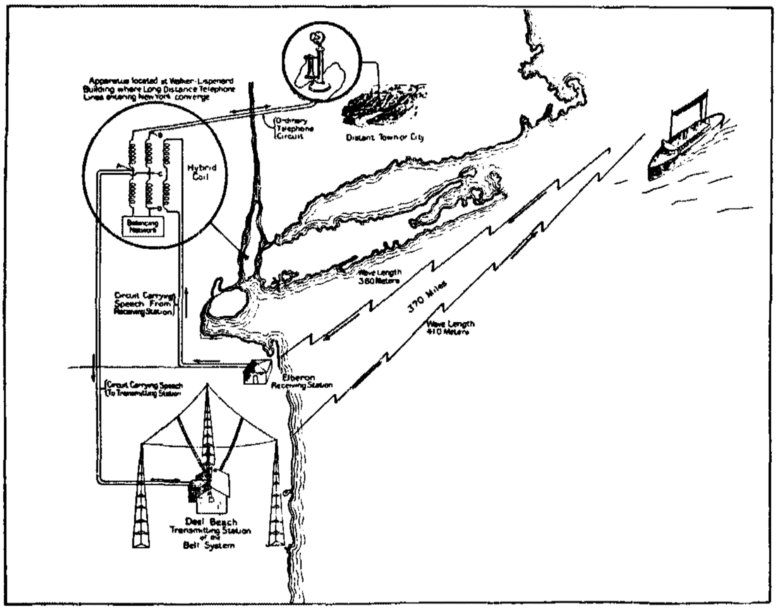


Fig. 14-1. Experimental ship-to-shore radio, 1919.

years prior to World War II until it covered all major harbors and a large portion of the Atlantic, Pacific, and Gulf coasts; even Alaska had service along its lower coastline.[5]

Much less space was available on the smaller coastal vessels, however, than on the large ocean liners. By the mid-1930s, Western Electric had designed new equipment for this application, including a 50-w No. 13A transmitter for the ship station in a 34-pound package that occupied only 1-1/2 cubic feet. The companion receiver, coded the No. 12C, weighed only 16 pounds. A separate motor-generator was required to provide the plate voltage. Included in the design was a radiolocation feature, operating at frequencies in the range from 242 to 515 kHz. By turning a rotatable loop antenna, navigators could determine their positions, at least roughly, by establishing fixes on known land transmitter locations. By this time, the coastal and harbor service was mainly used by boat operators for their own communications; passenger service was non-existent. Weather reports and safety information were important features of the service [Fig. 14-2].

In later years, coastal and harbor service improved with advancing technology but slowly. In the 1960s, single-sideband modulation and narrower channels were required by the Federal Communications Commission (FCC). Minimum special-operator attention became a major goal. Maritime service at 150 MHz, with its more predictable propagation and benefits from advances in land-based systems, became more popular, and short-wave traffic declined.

High-seas mobile-radio service developed along similar lines. However, be-

Fig. 14-2. Coastal and harbor radio equipment on a fireboat in New York harbor, 1930. The receiver is on the shelf above a loudspeaker on the deck.

cause of the greater distances involved, higher power and more complicated antenna structures for the base stations had to be used [Fig. 14-3]. The diurnal effects, especially when a sunrise or sunset occurred between ship and shore, also caused severe fading and noise impairments. On the high seas, calling to and from passengers continued as the predominant traffic.

During this period, the auxiliary circuits developed to improve overseas radio were applied to ship-to-shore radio as well. A CODAN (carrier-operated device, antinoise) relieved the radio-operating crew of the need to listen to the channel all the time, waiting to hear something other than noise. Only when a carrier signal was received was the audio output enabled. A selector circuit, developed in 1926, provided selective signaling, although it was not commonly used. Calling a ship, even at present, is done by voice using the name of the vessel. A VODAS (voice-operated device, antisinging), a version of an echo suppressor, was developed in 1937 to couple the four-wire radio path to the two-wire telephone plant. A compandor was developed in 1934 for the transatlantic radiotelephone, but it seems not to have been applied to the ship-to-shore service because of the lack of an international agreement on standard characteristics and because of the alignment and calibration problems of the early years.

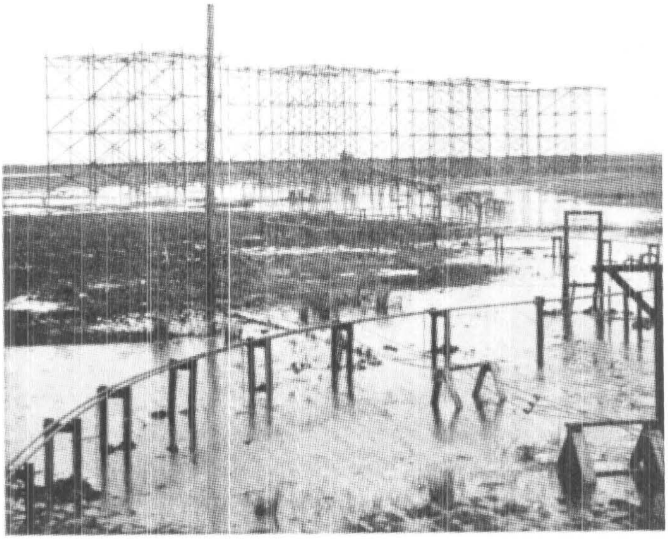
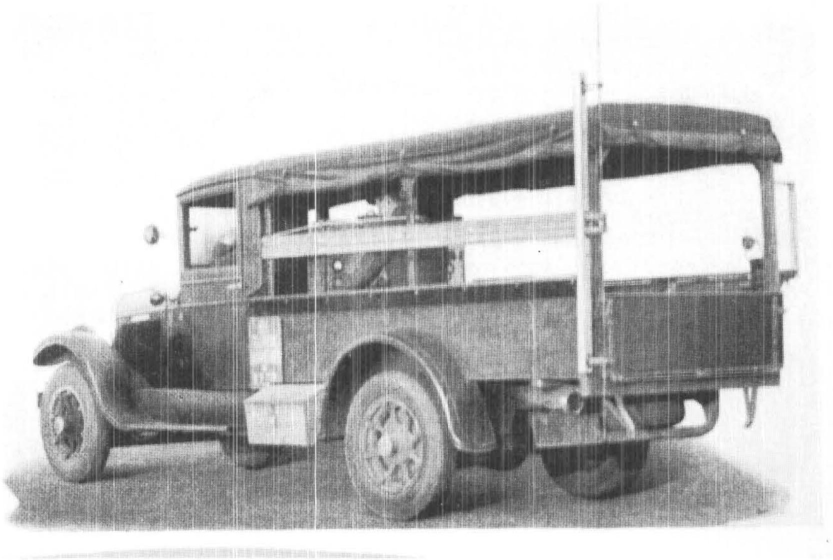


Fig. 14-3. A receiving-antenna structure of the ship-to-shore radio system at Forked River, New Jersey, 1930.



(a)

Fig. 14-4. (a) Mobile receiving equipment for the Boston propagation survey at 35 MHz.

(b)

Fig. 14-4. (b) Field-strength contour map of Boston, Massachusetts at 35 MHz from tests carried out in 1935.

Along with each land-transmitter location were two or more widely separated receiver locations to provide a diversity that compensated for the lower power of the ship transmitter. It was the responsibility of the marine-service operator to monitor the signal at each of these receivers and select the best one for the call. Manual receiver selection was commonplace for a number of years.

1.2 Land-Mobile Service

Satisfactory transmission to automobiles in city streets was more difficult in many ways than to ships over water because of high noise levels and the shielding effects of large buildings. The earliest application of two-way radio to land-mobile service was in 1933 to police and fire vehicles, although experimental installations had been made much earlier. There was no specific allocation for common carriers at first, and land-mobile tests remained a research activity through the 1930s. It was expected, however, that a band near 35 MHz would eventually be made available. During this time, some special applications were designed as a public service and to gain experience in this field.

1.3 Propagation Studies

Early propagation work centered on long-range transatlantic radio communications, and ship-to-shore development benefited from these studies. Radio noise was correlated with the sunset cycle. The effects of the ionosphere were investigated with regard to the diurnal and seasonal variation of signal strength. The theory of propagation over a smooth spherical Earth with specific conductive properties was developed. The well-known four-thirds radio-equivalent Earth radius was derived to take into account the atmospheric refraction of radio waves. Later, the efficacy of higher frequencies for land-mobile use was investigated. Work at 40 MHz began in 1926 and at 70 MHz in 1933. In 1935, there was an extensive test program in the Boston area at 35 MHz [Fig. 14-4].[6] In 1938, multipath fading at 75 to 150 MHz was noted to be a problem. (See Chapter 7, Section 2.2.)

II. THE POSTWAR EXPLOSION (1945–1963)

In many areas at Bell Laboratories, there was an explosion of new discoveries and new services after World War II. Landmark advances were made in solid-state physics, microwave transmission, and toll switching; and mobile-radio transmission was no exception. Ship-to-shore service was well established, and the focus shifted to land-mobile work.

2.1 Two-Way Land-Mobile Service

The early mobile police and fire systems were below 40 MHz. As a result of the research on urban propagation during the 1930s, enough had been learned about the 150-MHz band to indicate that transmission was reliable, with manageable interference problems and diurnal effects. With phase or frequency modulation, the achievable noise floor and the practicable base-transmitter power combined to yield a coverage radius comparable to the size of a major city. Bell Laboratories moved quickly to develop the necessary hardware for a commercial service, as soon as the FCC opened up the 150-MHz band. By early 1946, the FCC granted a license for the first land-mobile system in St. Louis, Missouri; Green Bay, Wisconsin followed soon after; and, by year-end, 25 American cities had begun service [Fig. 14-5].[7]

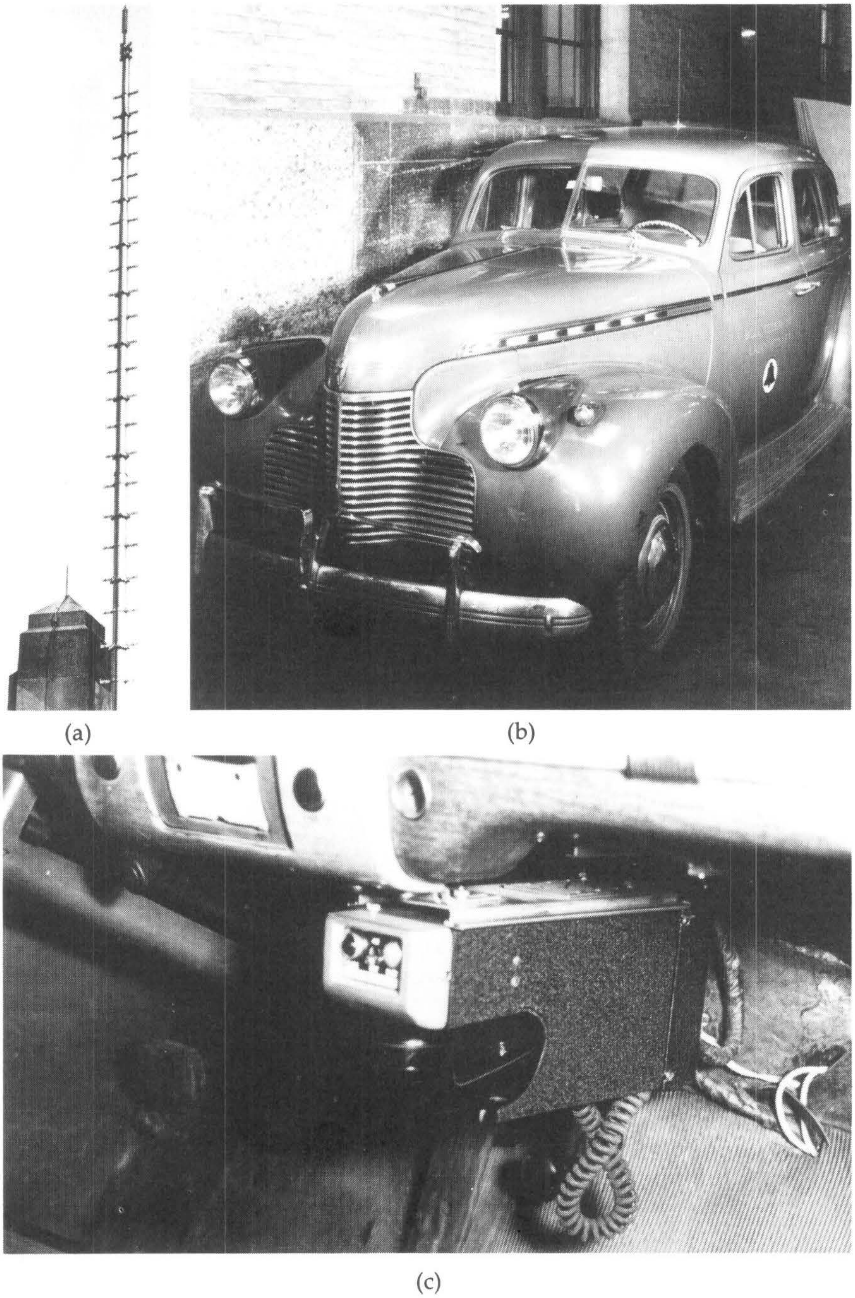


Fig. 14-5. The first land-mobile radio service, 1946. (a) The half-wave antenna array for the main transmitter of the St. Louis mobile-radio system. (b) A short length of piano wire served as the antenna for the mobile unit. (c) The vehicle control unit mounted under the dashboard.

Originally, the channels were allocated on a 120-kHz spacing. By 1950, the channels were split (i.e., the spacing was halved, to 60 kHz); however, early receivers could not discriminate between these closely spaced channels and a given city could make use of only every other frequency assignment. By the 1960s, however, filter technology had progressed to the point so that another split to 30 kHz spacing was made, and adjacent channels could be used in the same city. This split also required that the previous 15-kHz peak deviation be reduced to 5 kHz; it had been determined that the wide deviation did not yield enough benefit to justify the added use of the spectrum.

Early land-mobile operation followed the practice established in the earlier public-safety services. Since the mobile equipment could not transmit and receive simultaneously, the users had to operate "push-to-talk" switches whenever they wished to transmit. Operation was thus half duplex, i.e., one direction at a time, over the radio link. However, since the base and mobile stations transmitted on different paired frequencies, the land transmitter and the land receiver could be coupled to the two-wire telephone plant without automatic voice-signal-sensing circuitry. Frequent reuse of channels could be made, since the numerous 20-w mobile transmitters did not interfere with the 250-w fixed land transmitters. Each mobile unit was permanently tuned to one specific channel. Little mobile-to-mobile calling was expected.

From the start, the Bell System contemplated three services on the common-carrier channels: (1) public service fully interconnected with the land network via mobile-service operators, (2) dispatch service between the members of a fleet of vehicles and a central non-Bell dispatcher who controlled the fleet,[8] and (3) one-way paging. These services were usually offered on separate channels. Typical charges in the 1940s for the public two-way mobile radiotelephone service were 15 dollars per month, plus 15 cents per minute of usage.

As with the ship-to-shore system of the 1930s, various auxiliary circuits were used as adjuncts to the overall service. A high modulation index was maintained, in spite of talker-volume variations and different trunk losses, by means of a VOGAD (voice-operated, gain-adjusting device). Selection of one of a number of fixed receivers was required, since several receivers scattered over the service area were needed to compensate for the 10-dB difference between the fixed-base and mobile transmitter powers. The receiver with the best signal-to-noise ratio had to be selected automatically, with no switching transients. Mobile identification was still done on a manual, verbal basis. In fact, the mobile stations were quite simple, consisting of a transmitter, receiver, handset, and selector for base-to-mobile calling. All special circuits were at the base station side of the connection.

By the mid-1950s, the FCC reached some decisions as to the allocation of the ultrahigh-frequency (UHF) band from 450 to 890 MHz and common carriers were assigned more channels for mobile use at 452 to 457 MHz. These started with 50-kHz spacings but eventually were split to 25 kHz. At about the same time, the FCC set aside other channels, both at 150 MHz and 450 MHz for use by what came to be called radio common carriers. This common-carrier category was created to provide competition to the established wireline carriers already engaged in mobile service. A channel resource equivalent to that of the wireline carriers was deemed appropriate to allocate to them. (Some air-to-ground work was also done later by the Bell System at 450 MHz, but this did not amount to anything commercially.)

Soon after the 150-MHz service (called *VHF* for *very high frequency* or *high band* by the industry) was started in 1946, a second service in the band from 35 to 42 MHz (called *low band*) was also initiated. This served a wider coverage area than VHF and was intended to be a highway service. However, it never achieved the widespread use of the urban VHF service. Coverage was often unpredictable, and, during the peaks of the sunspot cycle, interference between widely separated systems occurred. This meant that the channels could not be reassigned in nearby cities to the same extent as those at 150 MHz.

Train service gained attention soon after manual 150-MHz systems had proliferated along the railroads' rights-of-way.[9] By 1950, service had been initiated on several Baltimore and Ohio, Pennsylvania, and New York Central Railroads intercity trains. Because the users would not be trained as to the calling procedure, including manual channel selection, an on-board telephone attendant was required [Fig. 14-6]. This, of course, was a major expense; and, soon, an unattended, coin-operated, train mobile set was designed. However,

Fig. 14-6. Telephone service to railroad trains, 1948. Train telephone operator at the control point for the mobile-radio telephone service. The window to the right of the handset mounting opened to the telephone booth on the other side of the wall.

this was not a commercial success either, and train mobile service died away, even before the trains themselves succumbed to the airlines.

The 1956 consent decree, ending the government's 1949 antitrust suit, changed the role of the Bell System in mobile radio. Prior to that time, the Bell System could (and did) participate in a wide range of work relating to mobile service. They custom designed special installations for public safety service, even though these did not interconnect to the switched public network. They manufactured a wide variety of base-station and mobile equipment to complement their radio-broadcast, studio-sound, and motion-picture products. After 1956, Western Electric dropped out of the business of providing radio equipment and special system design, although they continued to build the control terminals that interconnected the radio gear to the land network.

2.2 Paging Service

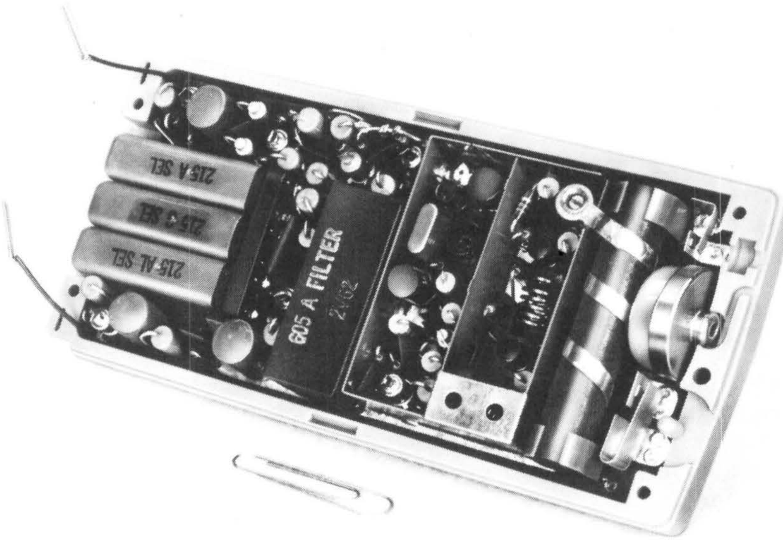
Paging service, known in the Bell System as BELLBOY,* was one of the first services to exploit the transistor's capabilities. In the vacuum-tube era, paging was seen to be merely a one-way dispatch service, using rather bulky vehicular-mounted receivers, with only a limited demand. However, by the late 1950s, a truly personal paging service, independent of vehicular two-way radio, was anticipated. After preliminary tests at 35 MHz in the Allentown, Pennsylvania area, building-penetration capabilities at higher frequencies were studied, and developers worked on solid-state designs for lightweight small receivers.[10,11,12] In 1962, the Western Electric No. 55A BELLBOY unit was born [Fig. 14-7]. Signaling was accomplished by a combination of 3 out of 32 closely spaced audio tones, filtered by means of tuned reeds.

After a series of tests in New York City, a full-blown service test took place in Seattle during the 1962 World's Fair.[13] This was followed up by a second developmental system in Washington, DC to test the system's ability to operate on a channel adjacent to an operating mobile channel. Unfortunately, the nationwide commercial use of BELLBOY became mired in the regulatory process for nearly a decade. By the time the log jam was broken, the No. 55A receiver and the base terminal were obsolete, and the three-tone signaling plan had been superseded. Western Electric had dropped out of the paging-equipment market, and the radio common carriers jumped ahead when the FCC finally began to grant licenses. The Bell System's pioneering effort was exploited by other participants in later years.

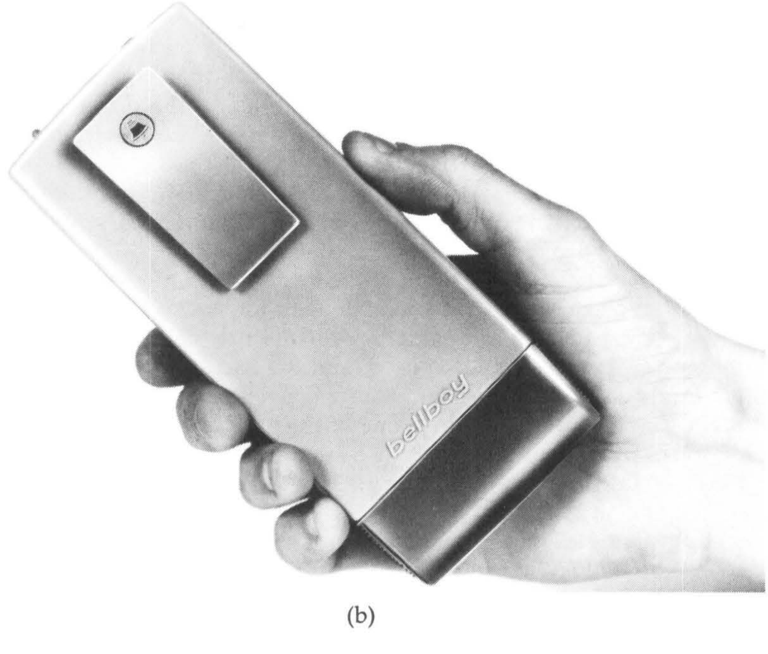
2.3 Research and Long-Range Planning

With commercial service initiated at 35 and 150 MHz, research effort turned to higher frequencies. Research was initiated on propagation at frequencies as high as 3.7 GHz in urban and suburban environments.[14] K. Bullington summarized what was then known about coverage prediction.[15,16] The planners also realized that the few channels at 150 MHz were woefully inadequate for the long term. In 1960, W. D. Lewis, H. J. Schulte, and W. A. Cornell showed

* Trademark of AT&T.



(a)



(b)

Fig. 14-7. Pocket-sized personal paging receivers, 1962. (a) BELLBOY circuit layout. (b) BELLBOY receiver—finished product.

how a coordinated, centrally controlled system could work to provide greatly increased service capacity.[17,18] These studies were of basic importance to future widespread high-quality service to really large numbers of mobile subscribers. (See another volume in this series, *Communications Sciences (1925–1980)*, Chapter 5, Section VI.) In these studies, the planners considered the space-frequency separation possibilities, and service area cells and cell splitting to increase capacity as traffic grew, anticipating in these concepts the 1983 commercial service of advanced mobile-telephone service (AMPS) by more than 20 years. At the same time, advances in solid-state technology made possible low-power, low-cost, reliable circuits operating in the upper VHF band.

III. MOBILE RADIO COMES OF AGE (1963–1975)

Manually controlled land-mobile service was in place in several hundred cities and towns nationwide, but it offered service to only to an extremely small fraction of the driving public. Research and exploratory effort therefore turned increasingly to expanding total capacity, which already was saturated in major cities, and to making calling easier. The objective was to make calling as convenient and quality as good as direct-dialed landline calling.

3.1 Improved Mobile-Telephone Service

One method of increasing the number of channels and improving spectrum utilization was to narrow the channels. It was soon realized that the 15-kHz FM deviation used at first was unnecessarily wide. As noted previously, the FCC required that mobile radio convert to a 5-kHz maximum deviation by 1963; and this, plus a general tightening of specifications, permitted the split to 30-kHz channel spacing. Since the change had to be made, this was an opportune time to add other major improvements. As a result, in 1964, the improved mobile-telephone service (IMTS) was developed and put into service. It furnished automatic channel selection for each call and full duplex service for simultaneous talking and listening, eliminating the need for push-to-talk operation, and allowed customers to do their own dialing.

The improvements were a result of the low cost, low power, small volume, and more complex circuits made possible with the new semiconductors. Very small, automatic pulsing and counting circuits had become practical; many more processing stages could be used; and the use of multiple audio-control tones became easy and convenient. Automatic channel selection made use of the marked-idle concept, in which one of the available channels was marked with a tone. The mobile station could search for this tone automatically, and the next call, in either direction, was made to use this channel. Each user was assigned a line and a telephone number in a local office, and thus the mobile numbering plan became compatible with standard numbering. As a result, except for the higher blocking probability and longer call setup times, mobile calling became as easy as normal land calling.[19,20]

Development of IMTS took place from about 1962 to 1964, with a field trial in Harrisburg, Pennsylvania. By 1965, IMTS began to spread to other cities, both within the Bell System and in those served by other companies. In the latter case, Western Electric went to outside vendors and procured equipment on specification for mobile, base-station, and control-terminal equipment. Mobile equipment was fully transistorized except in some cases where a vacuum-tube final-transmitter stage was still used. The unit size of the combined transmitter, receiver, logic, and power supply shrank to under one cubic foot. The control terminal, which predated the introduction of stored-program-controlled electronic switching by a short time, was one of the last switching systems to use relays and crossbar switches.

3.2 IMTS Derivatives

The initial application of IMTS was at 150 MHz. Manual systems at 450 MHz were still few in number by the mid-1960s, and, by 1968, IMTS was also adapted for that band. Channel spacing was halved to 25 kHz and Touch-Tone capability was also designed into the system.[21]

An interesting one-of-a-kind system, also an offshoot of IMTS, was for the Metroliner high-speed train service between New York and Washington, DC. This service was to be a demonstration project, sponsored by the United States Department of Transportation, to help revitalize railroad transportation on the short- to medium-length run.[22] The linear nature of the coverage area permitted a trial of the cellular concept, with planned and orderly channel reuse in the 450-MHz band. Otherwise, to minimize extra development, designers borrowed the marked-idle technique and control-circuit designs from IMTS. The service was designed for coin operation with control provided by operators at consoles of the traffic service position system (TSPS). Coverage in tunnels was provided by antennas directed into them [Fig. 14-8]. A minimum talking time of 15 minutes was guaranteed, and the telephone booth was integrated into the club car's design. In general, the service was a technical success but a maintenance and administrative failure and fell into disuse after about a 15-year life (from 1968 to 1983).

3.3 Further Research Activity

Despite the improvements made by IMTS, service in most areas continued to be far from satisfactory. The primary factor hampering the spread of public mobile-telephone service was the lack of available spectrum. In most areas of the United States, public mobile-telephone systems were saturated with customers and new customers for service could not be accommodated, since only a few dozen channels were available. The existing systems, with their limited channels, typically had an extremely high blocking probability. The few channels available were spread across several frequency bands and partitioned among different classes of companies offering service, making the number of channels available in any specific mobile-telephone system inadequate for more than 800 to 1000 customers. Waiting lists ranged up to 25,000, with many thousands more undoubtedly wanting service, but not even bothering to add their names to the list.

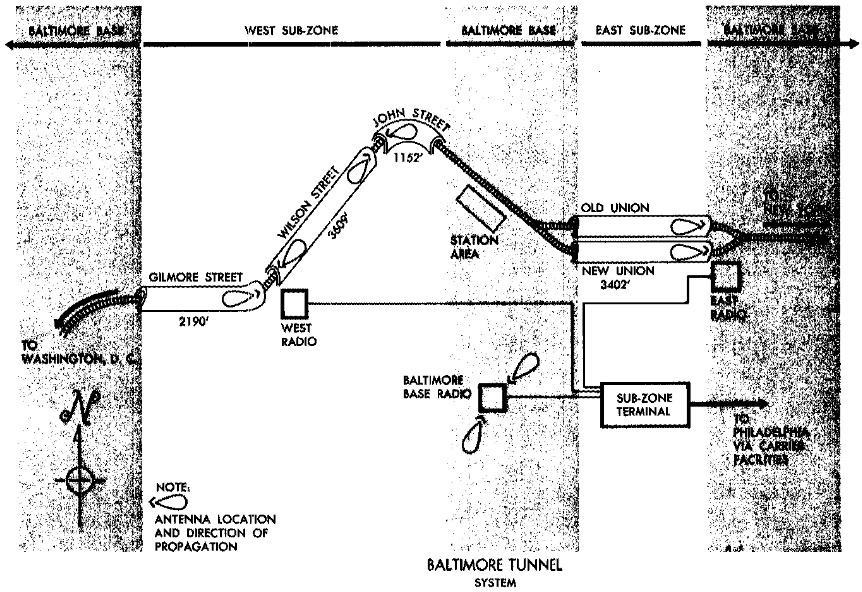


Fig. 14-8. Radiotelephone service to the Metroliner high-speed train. Special antenna arrangements were made to provide coverage inside a complex of five tunnels in Baltimore.

The unsatisfactory state of mobile service provided a strong stimulus to find a solution. Research activity continued to expand, again moving to higher frequencies.[23] The statistical properties of a radio channel propagation in urban areas and its correlations became better understood. Fading simulators for laboratory use were designed, and the efficacy of space, polarization, and other forms of diversity was demonstrated. This work was both theoretical and experimental, with much testing taking place around the Murray Hill and Holmdel Bell Laboratories locales, using specially instrumented vehicles. Diversity gains of up to 10 dB were achieved. (See *Communications Sciences (1925–1980)*, Chapter 5, Section 6.2.2.) Echoes and delay distributions were measured. Experimental radios at 850 MHz were built, and other tests were made at frequencies as high as 11 GHz. The results of this research were published in a definitive text that provided the solid base of knowledge on which practical high-capacity systems could be designed.[24]

3.4 Plans for Cellular Radio

The major obstacle continued to be the divided and woefully inadequate spectrum allocation. As early as 1947 and again in 1958, the Bell System pointed out that the systems at 35, 150, and 450 MHz could only be considered as stopgap services and petitioned for a continuous block of frequencies, first of 40 and later of 75 MHz, somewhere in the VHF/UHF band, but no action was taken, and mobile-telephone spectrum assignments continued on a piecemeal basis.*

The concepts evolved in the late 1950s for combined space and frequency planning, and the research on propagation continued to be pursued seriously in the late 1960s. By then, it was becoming evident that channel synthesizers could be built to generate any of hundreds of channels; antenna combiners for many channels at a base station could be designed; stored-program electronic switching made handing off a call from one base to another practical; and signaling could also be fast and reliable.

Finally, on July 26, 1968, the FCC issued a Notice of Inquiry and Notice of Proposed Rule Making (Docket 18262) in which it proposed to allocate 40 MHz for private mobile-radio systems and 75 MHz for common-carrier, high-capacity systems. In 1974, after much contention and shifting, the FCC allocated 40 MHz to the wireline common carriers for public mobile-radio service. (The contention and adjustments were to continue throughout the 1970s and later, but, as of 1984, it appeared certain that at least a major fraction of the allocation would survive.)

The docket asked interested parties to discuss how a high-capacity system could best be realized. After a quick feasibility study and more measurements comparing 850-MHz propagation to that at 450 MHz, a more comprehensive study was made in Bell Laboratories in 1970 and 1971. A major report was submitted to the FCC in December 1971, covering in considerable detail what was to evolve into the AMPS system.[25] The report covered not only the technology of a cellular system but service features, coverage, capacity growth, customer opinions on quality, and costs as well. An air-to-ground system at 850 MHz for passengers of the burgeoning United States jet passenger service was also included. This report was well received in general by the FCC, and the work was further enlarged to look toward field trials for both the equipment and the market. By 1975, detailed specification had been drawn up for equipping up to 2000 or more mobile units and creating a working cellular system. Work on air-to-ground service was deferred.

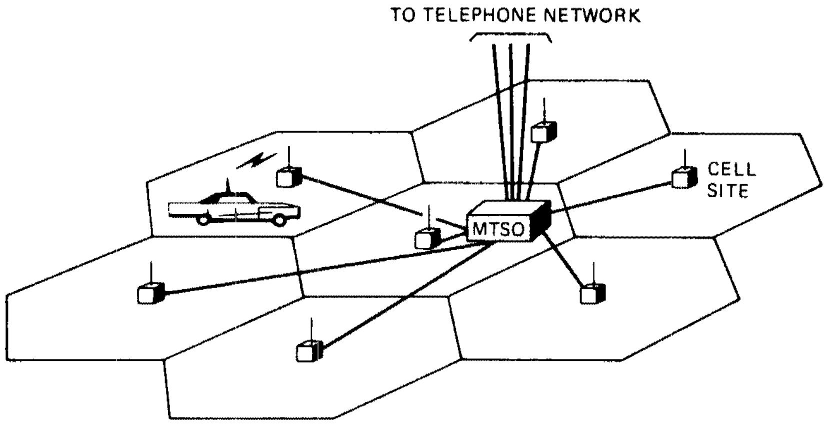
Further tests of urban propagation and interference in a tightly packed, cellular-channel assignment plan were made in Philadelphia, Pennsylvania and Camden, New Jersey, and a field laboratory was set up in the Newark, New Jersey area. These tests were meant to demonstrate cellular operation for a cell as small as 2 km in radius in a realistic environment. The full-scale technical and market AMPS trial was equipped and carried out in Chicago in 1978 and continued for several years.

3.5 Cellular Operation

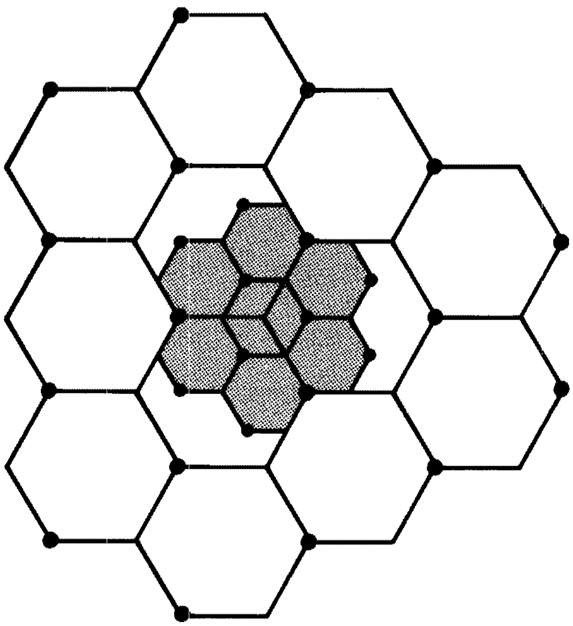
The AMPS system used FM channels 30 kHz wide. If 40 MHz were allocated, there would be 666 two-way channels available. In a conventional noncellular

---

* The frequency spectrum is an extremely valuable resource, and the FCC has the difficult and thankless job of allocating it among the contending interests for the maximum public benefit. The technology for mobile telephony was in a struggling early stage, and the potential users, while numerous, were few compared to television audiences that were growing to tens of millions. Nevertheless, it is startling to note that, in 1951, the entire 470- to 890-MHz band was allocated for additional broadcast television channels.



(a)



(b)

Fig. 14-9. The cellular mobile radiotelephone system. (a) Each cell is served by a relatively low-power transmitter-receiver. Cell-to-cell handoffs and other system functions are controlled from a mobile telecommunications switching office (MTSO). (b) When the initial cell capacity is reached, area capacity can be increased by cell splitting, reusing the same frequencies in smaller cells.

system, this might serve 10,000 to 20,000 customers. In such a system, a single high-power transmitter would cover about a 20-mile radius. Beyond this range, the signal would not be strong enough for reliable communications, but the same frequency could not be used by any other system within about 100 miles.

In a cellular system, the area served is divided into a number of cells, each served by a relatively low-power cell-site transmitter-receiver. Each cell site station is connected to a mobile-telecommunications switching office (MTSO) by standard telephone lines [Fig. 14-9]. Channel frequencies used in one cell can be reused in another only a short distance away, typically equivalent to about five cell radii. With cells as small as one mile in radius, a given channel frequency can thus be used simultaneously for many conversations in a single service area. The use of low power, possible for the short ranges, and interference-resistant FM make this possible.

A variety of cell sizes can be used. In Chicago, the initial cells were about eight miles in radius, and transmission was from omnidirectional antennas near their centers. As service demand grows and all available channels in any area are in use, the capacity can be increased by cell splitting. The fixed cell-site radio transmitter-receivers are multiplied in smaller cells, each of which can have as much capacity as the original large cell. This process can, in principle, be repeated several times, yielding the capacity to handle hundreds of thousands of customers in any one service area and millions nationally.

While the cellular structure makes high capacity possible, it also presents complex problems in systems operation. A mobile unit initiating a call or to which a call is directed must be located and identified. After a call is set up, the mobile unit may move from one cell to another in the course of the conversation, requiring a handoff from one cell site to another. Since adjacent cells use different frequencies, this, in turn, requires retuning the mobile unit to the new channel, while the switching office simultaneously reroutes the call from the old cell site to the new one. All this is accomplished by radio control channels, over which the mobile units can be scanned for call-initiation activity or addressed to receive a call directed toward them. A considerable amount of data must be exchanged over the control channels in both directions. All control is exercised from the MTSO by means of computer-like programmed equipment. In addition to controlling the operation of the cellular system, the MTSO provides the interconnection to the fixed-telephone network. As far as the users are concerned, calls from a mobile telephone are placed exactly as they would be from any fixed telephone, including calls to other mobile units.

The trial system in Chicago was a typical start-up system with ten cell sites, covering about 2000 square miles of the metropolitan area. The central controller was a specially programmed 1ESS* electronic switching system located at Oak Park, Illinois. Both the technical and market trials were resounding successes. While the skirmishing and politicking continued and corporate structures changed dramatically, cellular radio, based on Bell Laboratories research and development, appeared certain to play a significant role in the communications scene during the late years of the 20th century.[26]

* Trademark of AT&T Technologies, Inc.

REFERENCES

1. H. W. Nichols and L. Espenschied, "Radio Extension of the Telephone System to Ships at Sea," *Bell Syst. Tech. J.* **2** (July 1923), pp. 141-185.
2. R. A. Heising, "Telephony Between Ship and Shore," *Bell Lab. Rec.* **8** (January 1930), pp. 204-209.
3. W. K. St. Clair, "A Telephone System for Harbor Craft," *Bell Lab. Rec.* **11** (November 1932), pp. 62-65.
4. F. A. Gifford and R. B. Meader, "Marine-Radio Telephone Service for Boston Harbor," *Bell Syst. Tech. J.* **14** (October 1935), pp. 702-707.
5. H. M. Pruden, "Coastal Radio Telephone Systems," *Bell Lab. Rec.* **21** (March 1943), pp. 183-187.
6. C. R. Burrows, L. E. Hunt, and A. Decino, "Ultra-Short-Wave Propagation: Mobile Urban Transmission Characteristics," *Bell Syst. Tech. J.* **14** (April 1935), pp. 253-272.
7. A. C. Peterson, Jr., "Vehicle Radiotelephony Becomes a Bell System Practice," *Bell Lab. Rec.* **25** (April 1947), pp. 137-141.
8. A. E. Ruppel, "Direct Dispatching for Mobile Telephones," *Bell Lab. Rec.* **34** (March 1956), pp. 86-89.
9. N. Monk, "Telephone Service for Trains," *Bell Lab. Rec.* **26** (January 1948), pp. 9-12.
10. W. Strack, "Pocket-Radio Signaling," *Bell Lab. Rec.* **36** (January 1958), pp. 9-12.
11. L. P. Rice, "Radio Transmission into Buildings at 35 and 150 mc," *Bell Syst. Tech. J.* **38** (January 1959), pp. 197-210.
12. A. E. Kerwien and L. H. Steiff, "Design of a 150-megacycle Pocket Receiver for the BELLBOY Personal Signaling System," *Bell Syst. Tech. J.* **42** (May 1963), pp. 527-565.
13. D. Mitchell and K. G. Van Wynen, "A 150-mc Personal Radio Signaling System," *Bell Syst. Tech. J.* **40** (September 1961), pp. 1239-1257.
14. W. R. Young, Jr., "Comparison of Mobile Radio Transmission at 150, 450, 900, and 3700 MHz," *Bell Syst. Tech. J.* **31** (November 1952), pp. 1068-1085.
15. K. Bullington, "Radio Propagation at Frequencies Above 30 Megacycles," *Proc. IRE* **35** (October 1947), pp. 1122-1136.
16. K. Bullington, "Radio Propagation Fundamentals," *Bell Syst. Tech. J.* **36** (May 1957), pp. 593-626.
17. W. D. Lewis, "Coordinated Broadband Mobile Telephone System," *IRE Trans. Veh. Commun.* **VC-9** (May 1960), pp. 43-48.
18. H. J. Schulte, Jr. and W. A. Cornell, "Multi-Area Mobile Telephone System," *IRE Trans. Veh. Commun.* **VC-9** (May 1960), pp. 49-53.
19. H. W. Nylund and R. M. Swanson, "An Improved Mobile Dial Telephone System," *IEEE Trans. Veh. Commun.* **VC-12** (September 1963), pp. 32-38.
20. V. A. Douglas, "The MJ Mobile Radio Telephone System," *Bell Lab. Rec.* **42** (December 1964), pp. 382-389.
21. R. J. Cormier and H. M. Owendoff, "The MK (450 MHz IMTS) Mobile Radio Telephone System," *20th Annual IEEE Conference on Vehicular Technology*, Columbus, OH (December 4–5, 1969), p. 20.
22. C. E. Paul, "Telephones Aboard the 'Metroliner,'" *Bell Lab. Rec.* **47** (March 1969), pp. 76-84.
23. W. C. Jakes, Jr. and D. O. Reudink, "Comparison of Mobile Radio Transmission at UHF and X Band," *IEEE Trans. Veh. Technol.* **VT-16** (October 1967), pp. 10-14.
24. W. C. Jakes, ed., *Microwave Mobile Communication* (New York: John Wiley, 1974).
25. Bell Laboratories, "High-Capacity Mobile Telephone System—Technical Report," (December 1971), submitted to FCC.
26. See special issue on Advanced Mobile Phone Service, *Bell Syst. Tech. J.* **58** (January 1979), pp. 1-269.

Chapter 15

Terminals, Multiplexing, Interconnection

I. CHANNEL BANKS

1.1 A1 Through A5 Channel Banks (1938–1960)

The first carrier system on cable, Type K carrier, was limited to about 12 channels by technical and economic considerations. (See Chapter 5.) At about the same level of technical difficulty, Type J carrier could handle 12 channels in both directions over an open-wire pair. Of necessity, the two systems used different line frequencies. The coaxial system, designed for several hundred channels, required a much wider line-frequency band. Despite the differences, it became evident early in the development that a common terminal building block, usable for all three systems, would have important advantages. A generally applicable terminal unit would reduce the total amount of development and manufacturing preparation and would facilitate interconnection between different types of systems. Numerous frequency allocations and groupings of channels were studied, and a scheme was worked out that met the requirements of the three systems.[1]

In the common terminal unit, called a channel bank, each of 12 voice signals was modulated by mixing with a carrier tone from 64 to 108 kHz in a highly nonlinear rectifying bridge, passed through a filter to remove the upper sideband and combined with the other translated and filtered channels to form a standard basic group of 12 adjacent 4-kHz slots in the band from 60 to 108 kHz. The same bank received the reverse direction of the 12 channels at group frequency, separated them with band-pass filters, and demodulated them to recover physically separated channels at voice frequency. (See Fig. 5-12.) It consisted of 12 modulator-demodulators (modems) and a separate carrier supply to furnish the tones required for the modems. The channel bank was common to all three systems, each using it with different additional types of group modulators, demodulators, filters, and appropriate carriers to shift the 12-channel groups to the line frequencies of the specific system. (It was thus that channels, like eggs, came by the dozen. Transmission groupings were always determined by engineering and economic considerations but showed a strange decimal-duodecimal predilection. Later blocks of 60, 600, 1800, and 3600 evolved. The first digital system had 24 channels, and the second had 96.)

The channel terminal equipment developed by 1936 was designated the A1 channel bank [Fig. 15-1]. The modulating and demodulating circuits for the same two-way channel were paired, since the same carrier frequency was used. Similarly, in the A1 channel bank design, two modems (odd and even) were packaged together for convenient use of the carrier-supply outputs [Fig. 15-2]. Six of the double units were required for a complete 12-channel bank. Each two-channel unit occupied about 12 inches on a standard relay-rack bay, and 18 (three groups) were mounted on two 11-foot, 6-inch bays [Fig. 15-3]. Program circuits could be operated over Types J, K, and L carrier systems by removing two or three message channels from the channel bank and replacing them with the appropriate 5- or 8-kHz program modems. An adjacent bay included the carrier-supply equipment, with sufficient capacity to supply ten banks.

Modulation and demodulation were by means of copper-oxide disk rectifiers in a balanced bridge network. (See Fig. 5-15.) The rectifiers were carefully matched so that practically none of the carrier leaked to the voice circuits; and, for the same reason, the voice circuits were balanced out of the carrier supply, eliminating crosstalk between channels via this path. The disks were further selected to have substantially the same transmission loss, so that all channels were at approximately the same level at the output.

The filters were of the quartz-crystal type and provided a passband about 3000 Hz wide with very steep sides. Each filter required one side of a 3-1/2 inch panel for the full width of the bay. The 12 transmitting filters of a group

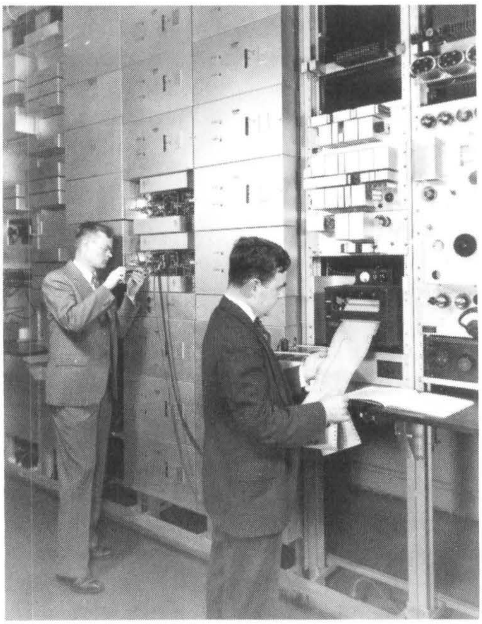


Fig. 15-1. Installation of A1 channel bank equipment at 32 Sixth Avenue, New York, 1938.

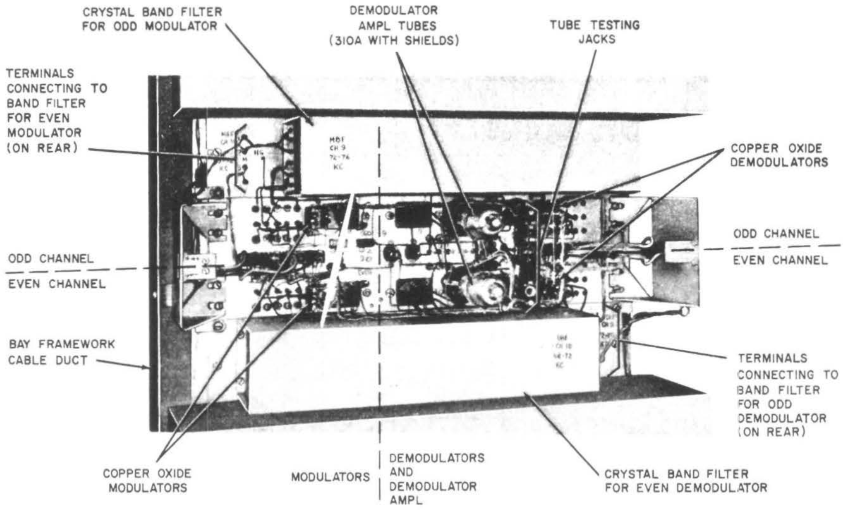


Fig. 15-2. Front view of the channel modem unit of A1 channel bank.

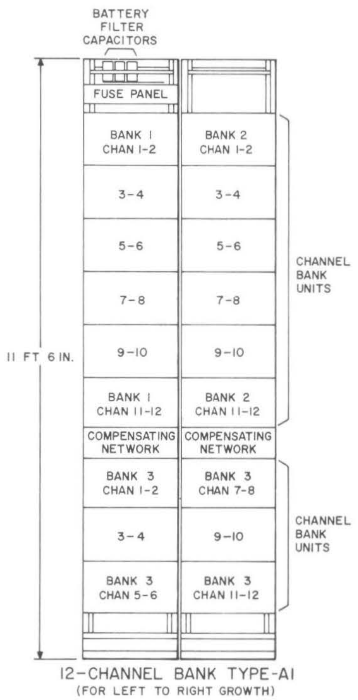


Fig. 15-3. Bay layout for the A1 channel bank.

were connected in parallel at their output terminals, together with a compensating network to improve the terminating impedance. A similar network was connected to the paralleled input terminals of the receiving filters. These compensating networks were housed on narrow panels in the same bay with the modems. (See Chapter 5, Section 3.3 for more detail on the first A channel bank.)

The actual channel signal amplitude on any carrier circuit depended on the loss of the connecting circuits and the volume level of the talkers. It was necessary, however, to have standardized gains and losses between circuit-access points to permit uniform testing and maintenance and to provide for cross connecting at patch bays. For most of the years from 1925 to 1975, the circuit-access point at the toll test board was taken as the reference (zero) transmission-level point (0TLP) from which circuit gains and losses were measured. Gains or losses with respect to 0TLP were defined as the transmission level (not to be confused with signal power).

To take care of losses between the test board and the four-wire circuit at the channel-bank terminal, the receiving level at the output of the channel bank had to be higher than the sending level. Prior to modulation, the outgoing circuit was entirely passive. A low power level for the transmitted signal was also desirable for low-distortion modulation, and the entire group was amplified as a single combined signal before application to the outgoing line. On the receiving side, the signal was received at a fairly low level and was further attenuated by losses in the demodulator. A voice-frequency amplifier was therefore necessary in each line so that the signals would reach the test board at the correct level. All the channels in a circuit group, often from more than one carrier system, were terminated in jacks grouped at one location (the four-wire patch bay). Gain-adjustment potentiometers for the receiving amplifiers were mounted above their corresponding jacks. This made a convenient arrangement for maintenance, since the level of all channels at the output of the carrier equipment could be measured and adjusted in rapid succession at the same place in the office. At first, the transmitted side was 13 dB below and the received side 4 dB above the test board. In later years, the levels at the carrier system four-wire terminals became standardized at +7 dB for the receiving side and −16 dB for the transmitting side.

About 1500 A1 channel banks were placed in service from 1936 through 1943. In 1944, a new design, the A2 channel bank, with smaller filters, went into production. The filter for the A1 channel bank was a double (cascaded) lattice, and the entire filter was hermetically sealed. In the A2 channel bank, the use of a single lattice of paralleled crystals with only the crystals sealed eliminated some large coils and achieved a size reduction of about 25 percent. Because of a shortage of natural quartz in the middle and late 1940s, a slight variant, the A3 channel bank, used synthetic crystals of ethylene diamine tartrate (EDT) as the piezoelectric element for the four lowest frequency channels. During and immediately following the war, a design in which each 4-kHz channel was split into two was also produced. The emergency-band (EB) channels were phased out as rapidly as the growth of line facilities permitted. The A2 and A3 channel banks furnished the terminals for the great surge of carrier expansion in the early postwar years. About 10,000 channel banks were placed in service from 1944 to 1953.

The A4 channel bank, which superseded the A2 design in 1954, differed from the earlier versions mainly in size. This was due to the first use of ferrite coils in the filter. (Ferrites are remarkable ceramic-like magnetic materials first developed by the N. V. Philips Company in The Netherlands in the 1940s.[2] Their outstanding property is resistivity higher than the metallic magnetic materials by a factor of a million or more, making possible low-dissipation magnetic-core coils.) Until magnetic ferrites were available, the low dissipation (high Q) needed for the sharp band-edge cutoff required bulky, air-core coils for use with the crystals. With ferrite coils and repackaging, it was possible to fit three complete 12-channel banks into an 11-foot, 6-inch bay, with a little room to spare.[3] A4 channel banks could be mounted in a single line, back to back or against a wall with single-side access. For the first time, a complete bank was factory wired and shipped as a unit [Fig. 15-4]. The jack field with the gain-control potentiometers could be located up to 500 feet from the banks.

Power for the banks was derived from 24- and 130-v office batteries or from the 152-v battery normally found in Type K carrier auxiliary stations. The

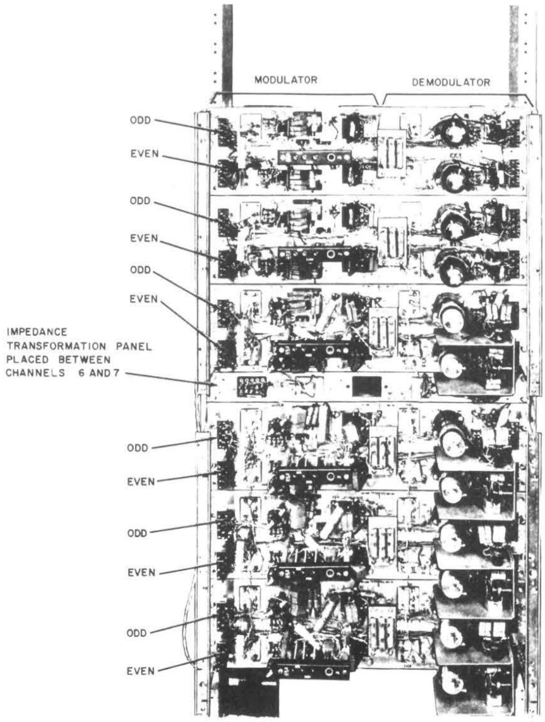


Fig. 15-4. Front view of the channel modem unit of a factory assembled and wired A4 channel bank.

amplifier following the demodulator for the A1 through A4 designs was a single No. 310A or No. 328A vacuum tube, a survivor from the early designs, and beginning to look old-fashioned, even to the vacuum-tube-accustomed eyes of the designers. Gain was adjusted over a range of 10 dB by varying the grid bias and feedback simultaneously. All A1 through A4 channel bank designs had about the same bandwidth, output limitation, noise, interchannel modulation, and adjustments in the transmitting and receiving circuits. The principal evolution was to smaller size and lower cost.

Up to the A4 design, the principal changes had been in the design and packaging of the filters. During the late 1950s, the spread of direct distance dialing meant that talkers were more and more likely to get machine-selected trunks, with no prior talk test by an operator. Trunks were routinely tested and adjusted, but loss variations were too high, especially on multilink connections. A channel bank with improved gain stability was essential.

Fortunately, the underlying technology was keeping pace with the need. By 1960, reliable, high-performance, low-cost transistors were becoming generally available. The major circuit change in the A5 channel bank design was the use of a three-stage transistor feedback amplifier following the demodulator [Fig. 15-5]. This provided 30 to 40 dB of gain, adjustable from the patchboard as in the earlier versions, but with improved impedance matching and a minimum of 30 dB of feedback at maximum gain. With this much feedback, it was expected that the channel transmission loss would not change by more than ±0.1 dB over a period of years, even with ±10-percent variation in the office battery voltage. The earlier channel amplifiers, with little feedback and aging vacuum tubes, were subject to as much as 2- or 3-dB gain changes.

Other devices, such as silicon diodes, were studied as possible replacements for the copper oxide modulators. However, problems of incompatibility with the carrier power available in existing offices and other considerations made it advisable to continue to use the copper oxide varistors. The transmitting and

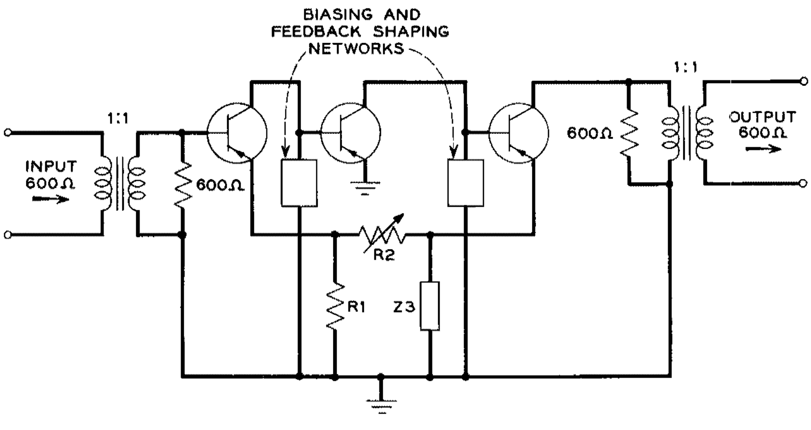


Fig. 15-5. The solid-state three-stage emitter-coupled feedback amplifier for the A5 channel bank.

receiving filters were combined in one container to obtain minimum volume. By careful placement of wiring and the shielding of individual inductors, acceptable crosstalk coupling was achieved [Fig. 15-6].

In the A4 channel bank, the physical design had been based on the use of flat mounting plates with the vacuum-tube amplifiers on one side and the crystal band filters on the opposite side. The A5 design incorporated a door-type arrangement that resulted in significant size reduction and better maintenance access [Fig. 15-7]. The 12 band filters were located on a fixed rear plate, extending back between the uprights when mounted in the bay. A few other frequency-determining components were located on the door assembly, but all modem and amplifier units were identical and plugged into connectors on the door.

The transistor amplifiers, compact filter design, and other miniaturized components resulted in a channel bank that required only 12-1/2 inches of vertical space in a bay (the entire bank now occupied the space of a single channel modem in the A1 design). Ten banks for 120 channels were accommodated in an 11-foot, 6-inch bay. Compared to the A4 channel bank, the required space, weight, and power consumption were all reduced by a factor of more than three. The only power required was from the standard 24-v office battery.[4]

Transistor circuitry had an even more dramatic impact on the design of the tone generators of the carrier supply. In the first 12-channel carrier terminals, more than a dozen bays were required to multiplex the signals, with much of the equipment in the carrier supplies. With the A5 design, the carrier supply for 30 banks had been reduced to one-half of one bay. Costs had been reduced by an even greater factor.

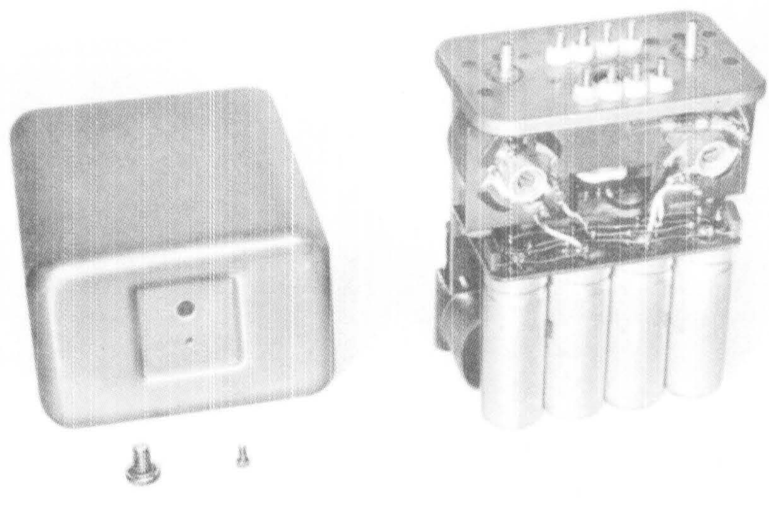
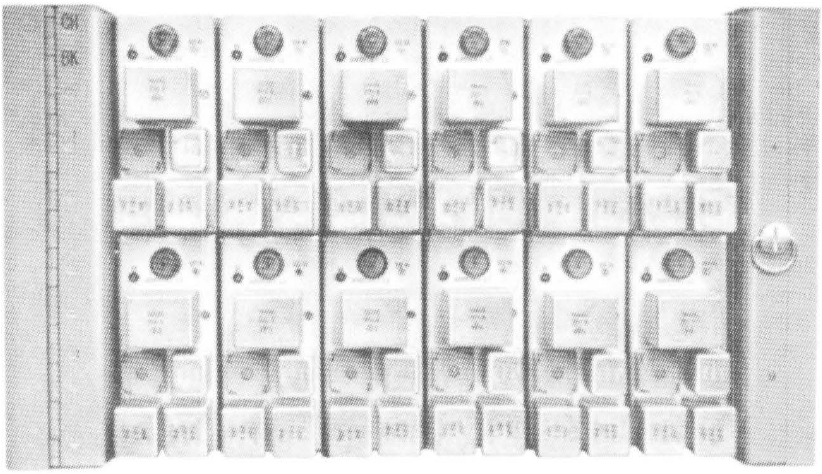
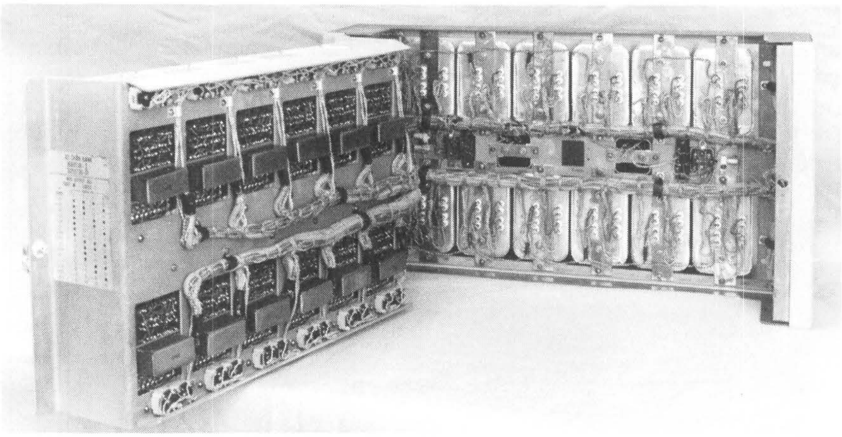


Fig. 15-6. The crystal channel filter assembly for the A5 channel bank. Careful shielding and component placement made it possible to place the transmitting and receiving unit in the same container.



(a)



(b)

Fig. 15-7. The A5 channel bank. (a) The bank with the hinged panel closed, showing the 12 modem-amplifier plug-in units. (b) With the hinged door open, the filters mounted on the rear frame were accessible.

1.2 The A6 Channel Bank

The carrier network continued to expand at a rapid rate. By the early 1970s, the number of channels in service was approaching one million, and the rate of growth was about 200,000 channel ends per year. Enormous economies had been achieved by handling circuits in bulk over carrier-system lines of ever-increasing capacity, but, as the growth continued, the pressure for further reductions in costs remained very high. The channel bank received exceptional scrutiny because it remained as the point where channels had to be handled one at a time. There was no economy of scale; the only route lay in further miniaturization, with the expectation that smaller would prove cheaper. The introduction of transistor amplifiers in the A5 channel bank left the filters as the bulkiest and most costly item; progress in integrated circuits promised even

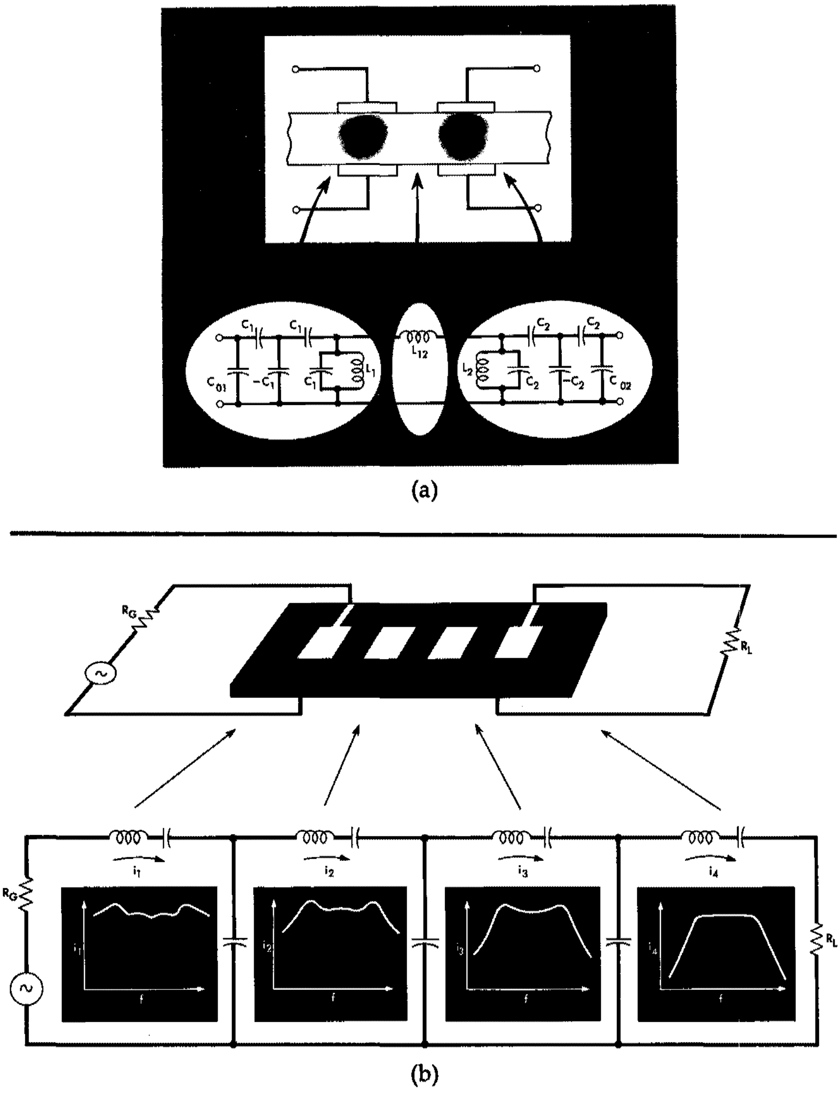


Fig. 15-8. The monolithic crystal channel filter. (a) Trapped resonances and coupling between them on a single quartz plate. (b) Selective band pass realized by a multielectrode filter.

further reduction in the active-circuit size and cost. As a result, attention was directed more and more to the filtering function.

Early research by W. P. Mason, and later work by R. A. Sykes, C. E. Lane, and others, had led to a deep understanding of the electroacoustic properties of quartz.[5,6] By the mid-1960s, it appeared possible to achieve a major reduction in channel-filter size by the use of a single quartz element in a monolithic filter, a radically different approach from the earlier designs in which the quartz crystals were used as discrete circuit elements.

In a monolithic filter, several pairs of electrodes were deposited on a single quartz wafer using photolithographic techniques. Because of the piezoelectric effect, signals applied to one electrode pair created an electromechanical (acoustic) resonance that was, in effect, trapped in the immediate vicinity of the electrodes. Some energy, however, was transmitted through the adjacent quartz. If an adjacent pair of electrodes and their local plate area had the same dimensions, a strong mechanical resonance under the second pair could be excited by the energy transmitted through the intervening nonresonant quartz. With a series of properly sized and spaced electrodes, the equivalent of a multisection, highly selective ladder network could be realized [Fig. 15-8]. The promise for miniaturization lay in the fact that no transformers, inductors, or other bulky discrete components were needed. When an experimental monolithic filter with a flat passband and excellent out-band rejection was demonstrated, development of a new channel bank, designated A6, was started in the late 1960s.

Although the monolithic approach promised a reduction in the filter size by a factor of ten, it had characteristics that made the channel bank much more complex than the earlier designs. A bandwidth of only 0.1 percent was available, compared to bandwidths ranging from 4 to 6 percent in the multicrystal filters. For this narrow percentage band to accommodate a 4-kHz voice channel, it was necessary to filter the channels at a frequency of several megahertz. Since compatibility with the existing banks with their output band at 60 to 108 kHz was essential (in operation, banks of any generation could be associated at the opposite ends of a line), a two-step modulation sequence was necessary to prepare the channels for transmission as a group. In the A6 channel bank, each channel was first modulated to about 8 MHz. At that frequency, a monolithic filter selected one sideband, which was then combined with the output of 11 other channel modulators and filters. The 12 single-sideband signals, spaced at 4-kHz intervals at 8 MHz, were then shifted as a group in a second stage of modulation down to the standard interval at 60 to 108 kHz [Fig. 15-9].[7]

Where the assembly (and disassembly) of the 12-channel group had been accomplished with a single vacuum tube or a simple transistor amplifier per channel, the bank now required channel-carrier amplifiers, balanced transistor modulators, and four 8-MHz amplifiers, in addition to the voice-frequency receiving-channel amplifier. Size and cost reductions were possible only by using the rapidly developing silicon technology and thin-film tantalum circuits. Four new silicon devices were designed to provide the carrier-tone amplifiers, the modems, and the voice amplifier. These, in turn, were interconnected with precision thin-film tantalum resistors and ceramic capacitors on alumina sub-

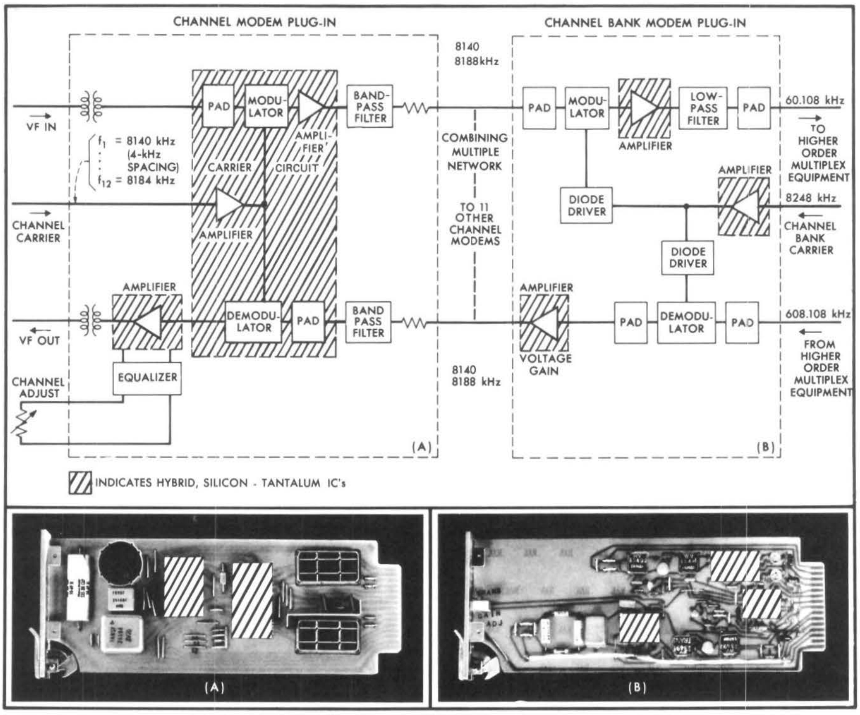
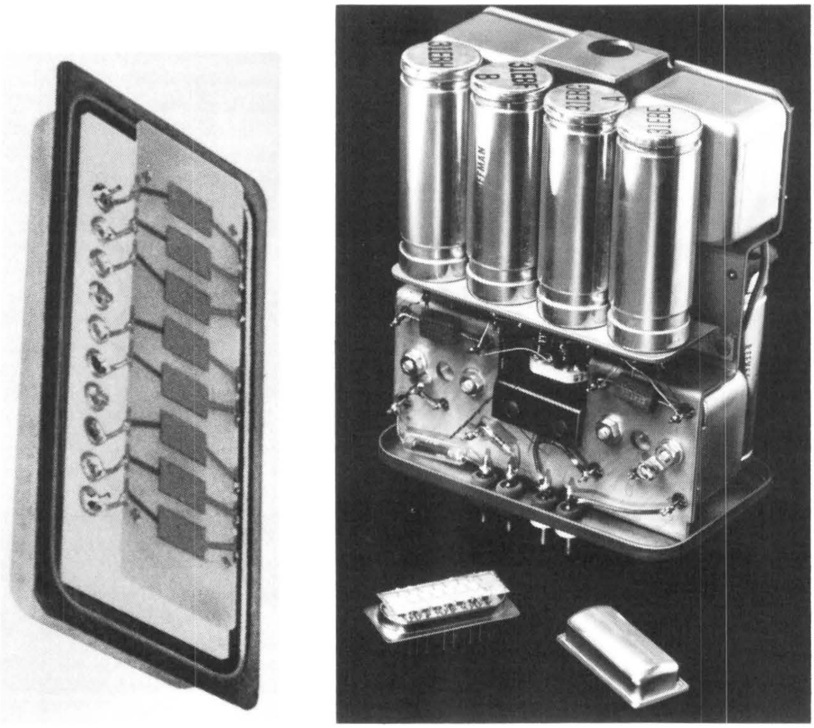


Fig. 15-9. The A6 channel bank schematic showing the two-step modulation to the standard 60-108 kHz group frequency. The individual channel modem (A) and the 12-channel group modem (B) were each realized on a single plug-in circuit board.

strates, forming hybrid integrated circuits (HICs). The filters were realized in an eight-electrode design on a quartz plate, only 0.5 inch wide by 1.5 inches long [Fig. 15-10]. Their performance was comparable in almost every respect to their bulky predecessors.

Filters, HICs, and all other A6 channel bank components were mounted on printed-circuit boards that plugged into a die-cast mounting shelf, a method of assembly that was almost universal by the time of the A6 development. A single card, about four inches high, served to translate each channel to and from the 8-MHz band, with the filters on the same card. Another card translated the entire group to and from the band at 60 to 108 kHz. By this approach, 20 channel banks (240 channels) were packaged in a single bay 23 inches wide and 11 feet, 6 inches high.

As more and more channels were packed in a bay, interbay wiring became a major problem; cable racks grew congested and overflowed. To relieve this, the A6 design associated the channel banks, carrier supplies, voice-frequency patching-jack fields, and gain-adjustment potentiometers into a single-bay assembly [Fig. 15-11]. The resulting space savings over the A5 channel bank were thus much greater than two to one, as suggested by the channel count.



(a) (b)

Fig. 15-10. The A6 channel bank monolithic quartz filter. (a) An eight-pole resonator for the A6 channel filter. (b) Monolithic A6 filters, *below*, compared to A5 filter, *above*.

II. LMX—L MULTIPLEX

2.1 Overview

In the Types K and J carrier systems, a single channel bank, followed by modulators specific to the system, provided the translation of the voice channels to the line frequencies. As the coaxial and microwave radio systems became available, additional modulation arrangements were required to multiplex many groups onto these high capacity facilities in a standard way. For the earliest coaxial system, a single-stage approach in which channel groups were stacked side by side up to the top line frequency, was briefly considered, but, by 1941, a two-stage plan was adopted for several reasons: (1) selectivity requirements on the band-pass filters were eased and fewer types of band filters were required, (2) fewer carrier frequencies were needed, and (3) more than 12 channels could be provided in a second common block in a standard-frequency range. This provided increased flexibility and economy for the interconnection of larger future systems.

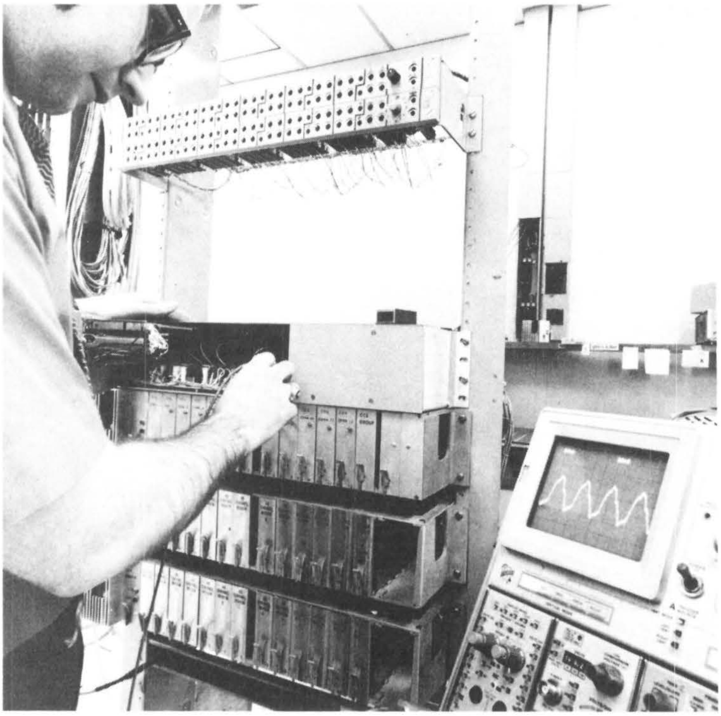


Fig. 15-11. A laboratory prototype A6 channel bank assembly. The patch panel is at the top of the frame.

Studies of traffic conditions and equipment feasibility led to the conclusion that the output of five channel banks should be combined into a supergroup of 60 channels. A basic supergroup, in the 240-kHz band from 312 to 552 kHz, became standard for analog multiplexes in the Bell System and internationally. The multiplex to combine the five 12-channel groups was called a group bank. Ten supergroups were combined in a supergroup bank to form a 600-channel mastergroup. The multiplexing technique for the group and supergroup banks (and continuing for larger groupings developed later) was basically the same as that used in the channel bank. Each multichannel signal was translated in a modulator to a higher-frequency band, passed through a filter to select one sideband, and combined with several other signals translated to adjacent frequency slots to form the larger aggregation. A symmetrical receiving demodulator disassembled the multichannel blocks in the reverse direction [Fig. 15-12].

In the original L1 coaxial system multiplex, eight supergroups were combined for a line allocation from 68 to 2044 kHz (one supergroup being translated down from 312 to 522 kHz to 68 to 308 kHz). Later, two supergroups were added at higher frequencies, and the lowest frequency was shifted slightly to

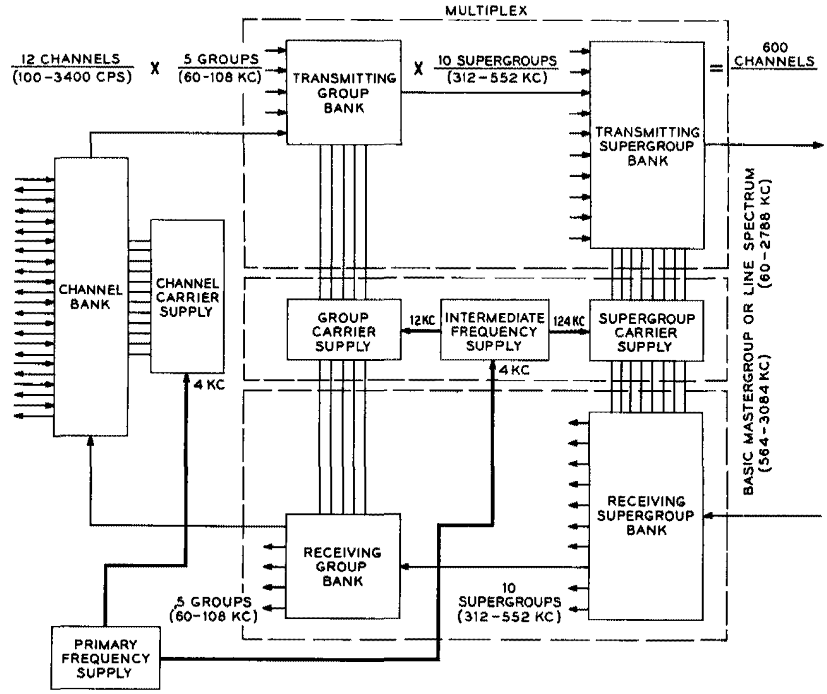


Fig. 15-12. The multiplex terminal to form a 600-channel mastergroup. The L multiplex equipment (LMX) is enclosed in dashed lines.

form a standard 600-channel block, the L-600 mastergroup. As systems grew beyond 600 channels, further stages of multiplexing were required. The L-600 block continued in use for some time, but a new standard assignment, the U-600 basic mastergroup, better suited for additional multiplexing and transmission over the higher-frequency lines, was defined in the frequency range from 564 to 3084 kHz [Fig. 15-13]. As coaxial systems continued to grow in capacity, six mastergroups were multiplexed to form a 3600-channel jumbogroup, three or more of which were then stacked to form the L5 coaxial line signal.

The terminal modulating equipment, beyond the channel banks, required to form (and dissect) a mastergroup was referred to as LMX, and the general plan of groups, supergroups, and mastergroups was known as L multiplex. When the high-capacity long-haul radio systems were designed, it was evident that the multiplex developed for the coaxial systems would also be satisfactory to form the multichannel baseband signals for the radio terminals. The use of the same multiplex for radio and coaxial systems permitted efficient and flexible interchange of traffic at offices using both types of facilities.

2.2 LMX-1

While the techniques for the higher-multiplex stages were essentially the same as those used in the channel banks, there were some elaborations because

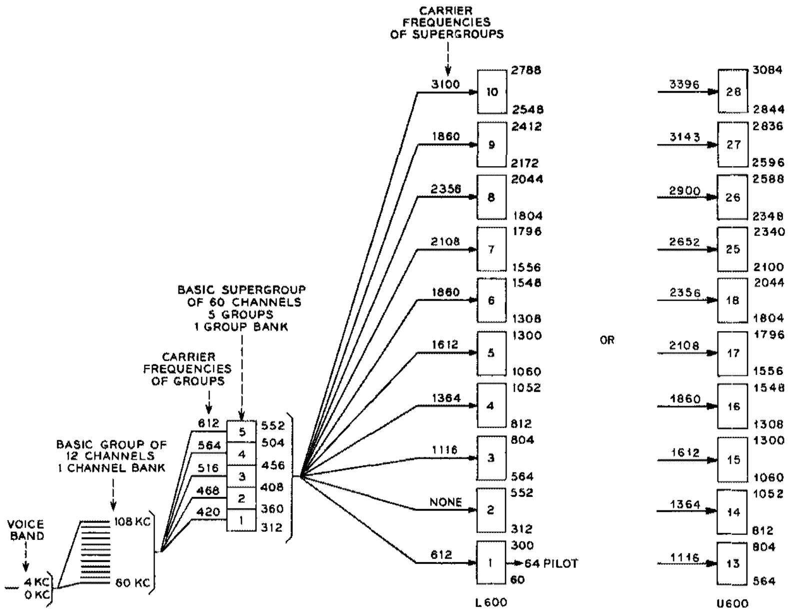


Fig. 15-13. L Multiplex frequency assignments for L600 and U600 mastergroups

of the higher frequencies and broader bands. At first, modulators and demodulators were copper oxide varistors in a balanced-bridge structure. The frequency spacings were such that only the band-pass filters for supergroups 1 and 3 required crystals to obtain the necessary filter selectivity. The rest used capacitors and air-core inductors. In the transmitting-group bank, a 92-kHz pilot was inserted for group-level adjustment.

Combining groups into supergroups and supergroups into mastergroups was somewhat more complex than combining channels into groups. Direct paralleling of a few groups and supergroups was possible, but it was necessary to combine subportions of a supergroup and mastergroup through hybrids to preserve reasonable impedances for filters and to give access for spares. The presence of the L1 2064-kHz, temperature-regulation pilot split the mastergroup into two submastergroups. The six lower-frequency supergroups were first combined, as were the top four supergroups, both combinations using paralleling and hybrids. The two submastergroups were then combined in an additional hybrid to form the complete mastergroup. The groups were directly modulated to and filtered at their final supergroup frequency, as were the supergroups to their mastergroup slots. An amplifier was required at the receiving-bank output to bring the signal to the proper output level. Additional amplifiers were needed to offset the loss of filters, combiner-splitters, and modulators. Gain adjustments were provided for initial lineup and periodic maintenance.[8]

Unlike the channel bank, where the two directions were combined in a single modem, the two directions of transmission were separated in the group and supergroup equipment. Two transmitting-group banks could be mounted in one bay, and two receiving-group banks occupied another. For supergroup banks, several bay arrangements were available. Generally, a transmitting supergroup bank would occupy one bay, and the receiving bank required two.

The LMX-1 carrier supply was a centralized design, requiring a 4-kHz sine-wave source to drive harmonic generators. The carrier supply derived all of the channel-bank, group-bank, and supergroup-bank carrier frequencies, which were then selected with filters and distributed by buses to the multiplex equipment. The carrier supply could support 30 channel banks and up to 3000 channels of group and supergroup equipment.

Since the LMX equipment carried many more channels than any earlier terminal, great precautions were taken to ensure continuity of service. The terminal was designed with liberal provision of spares and ready means for replacing any equipment that might require maintenance [Fig. 15-14]. For easy substitution of spare equipment, the spare and regular circuits were connected together through hybrid transformers on either the input or output side. All patching for substituting the spare for the regular equipment was done at special patching bays, as was all transmission measuring and testing.

The terminal equipment for the first 600-channel coaxial lines was designed around the vacuum-tube and filter technology of the late 1930s and 1940s. It

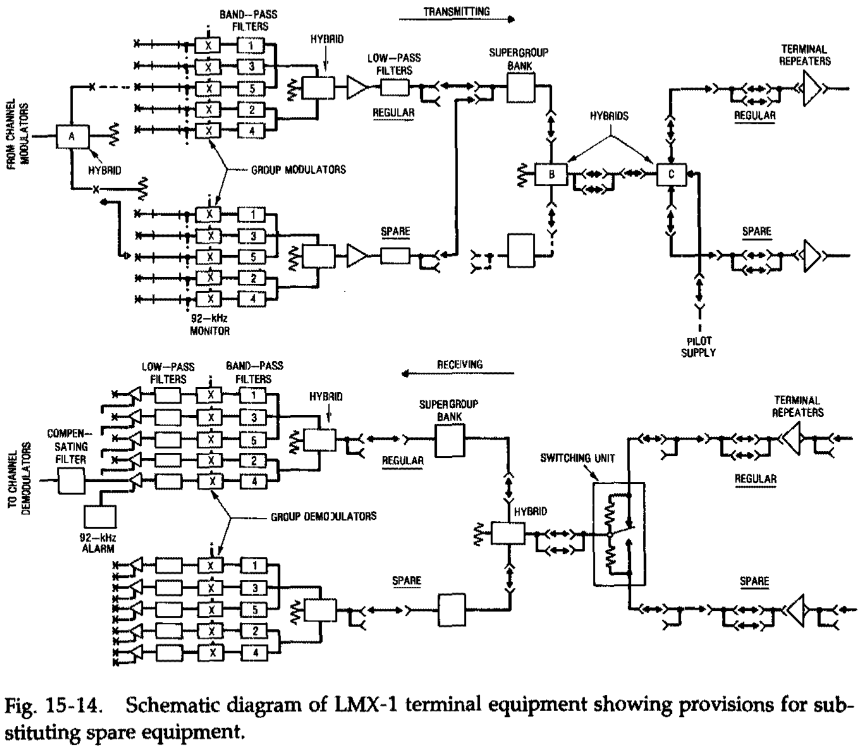


Fig. 15-14. Schematic diagram of LMX-1 terminal equipment showing provisions for substituting spare equipment.

worked quite well and reliably, serving the coaxial and early radio systems for about 20 years, with numerous improvements but no major change in architecture. Nevertheless, even by the standards of that era, it must have appeared quite bulky. A full 600-channel complement required 33 bays of channel banks, 20 of group banks, and 3 of supergroup banks, all 11 feet, 6 inches high. When the carrier supply, patching, and other bays were included, the total came to well over 60 bays.

2.3 LMX-2 and LMX-3

2.3.1 LMX-2

The transmission performance of the original LMX-1 terminal was generally satisfactory, but the spread of direct distance dialing brought more stringent demands on the gain stability of the higher-multiplex stages. Successful machine switching control signaling over a variety of trunks in tandem required tighter control of loss than with manual switching by operators. With LMX-1 equipment, maintaining the correct transmission loss of the channels depended on skilled personnel to carry out a coordinated program of test and adjustment procedures. Also, a particular set of adjustments was appropriate only for the assemblage of equipment in operation at the time the adjustments were made. Patching to an alternate facility or substituting spare equipment would disturb the equalization. Studies showed such mobility in the plant to be a principal cause of channel net-loss deviations. The development of LMX-2 was started in the late 1950s with several objectives in mind. The designers wanted to improve the uniformity and stability of channel net loss, with less dependence on skilled maintenance. They also intended to make the plant more suitable for data transmission, which was beginning to emerge as an important new signal to be handled.[9] Finally, they wished to take advantage of the large advances in technology since the design of LMX-1.[10,11] (The designations LMX-1, LMX-2, etc., were not used at the time but came into general use later when the sequential generations of design were recognized.)

The first objective was realized in part by the use of feedback-stabilized, transistor amplifiers in all parts of the multiplex but also by the use of pilot-controlled regulators in the receiving-group and supergroup stages. The regulators improved the overall transmission stability by automatically and continuously correcting for loss changes anywhere along the circuit between transmitting and receiving terminals. (One of the great bonuses of the solid-state revolution was the ability to use the new devices in an almost profligate way. Feedback amplifiers and automatic regulation were established techniques, and their virtues were well recognized long before the design of LMX-1. It is difficult for designers of a later generation to understand how reluctant designers of the vacuum-tube era were to add even a single tube to their circuits. The rows and rows of power-consuming bays, of course, held the answer.)

For easy initial adjustment and subsequent checking of the circuit loss and pilot level, scanner circuits were built into the receiving-group and supergroup banks. These provided an in-service measurement of each pilot, selected by a switch on the front of the panel. The readout was displayed on an expanded-scale meter, located so that it could be read conveniently while making any adjustment required. Meter readings were in terms of departure from the correct level to eliminate the need to remember or look up the many different combinations of frequencies and test levels.

The pilots used for automatic regulation were not those long familiar in LMX-1. In addition to the need to provide for message service and data services that could be handled over a message channel, there was a growing need for facilities to handle higher-speed data. By the early 1960s, Bell Laboratories was developing a 40.8-kilobit data modem, designed to transmit its signal over an entire group band, and even wider-band data terminals were in prospect. This led to a decision to change the frequency allocations of the terminal pilots of the L-type multiplex. A 92-kHz pilot had been used to monitor the transmission of each group, and the pilot in the third group, translated to 424 kHz, was used to monitor each supergroup. These choices were satisfactory for message service because the pilots were located near the center of the band they monitored and fell between adjacent message channels. But when the entire band was needed for data, a pilot near the center was a serious restriction. The problem was resolved by moving the pilots close to the edges of the group and supergroup bands, using 104.08 kHz for the group pilot and 315.92 kHz to monitor supergroup transmission. The offset of 80 Hz from an exact multiple of 4 kHz eliminated the need for an expensive crystal filter to guard against channel-carrier leak. It was recognized that changing the entire Bell System plant to new pilot-frequency allocations would incur substantial expense, but the integration of wideband data and message service was considered essential to future efficient use of the plant.

When the development of LMX-2 was undertaken, it was quite obvious that, in addition to improved performance and flexibility for wideband data services, large benefits could be realized by exploiting modern components and design techniques. The vacuum-tube circuits were replaced with much more compact and reliable transistor circuits. A large saving was realized in power equipment and power costs. The modulators and demodulators used silicon diodes, which resulted in substantially less loss to the signal. New ferrite materials allowed a dramatic reduction in the size of the filters. All the amplifiers were mounted in plug-in assemblies, introducing a high degree of flexibility and ease of installation and repair. The patching capability provided in the original LMX-1 design was retained, but the access jacks were provided on the multiplex equipment rather than in a separate patching bay.

The LMX-1 carrier supply had been designed for heavy cross-sectional use. Start-up costs were high, and it was not economical where comparatively few channels were needed. In LMX-2, a large part of the carrier supply was decentralized and integrated with the equipment it served. A channel-carrier supply for 30 channel banks continued to be centrally located, but the group and supergroup supplies were mounted in or close to the bays they served. Each group and supergroup supply was common to six bays (three transmitting and three receiving), and each of these bays had its own local carrier-distributing amplifiers and buses. This arrangement reduced the number of cables from the common group-carrier supply from one per modulator to one per frequency, effecting a reduction in numbers from 50 cables to 5. All the equipment for translating the output of 50 channel banks into one mastergroup was arranged in one shop-wired bay 11 feet, 6 inches high, with the corresponding receiving equipment included in a second bay. The total amount of equipment required to assemble, transmit, and receive a 600-channel mastergroup was reduced from over 60 to 7 bays, including 5 bays of the newly designed A5 channel bank.[12,13]

2.3.2 *LMX-3*

The LMX-2 design was in production by 1963 and furnished the analog multiplex terminals for the large transmission plant expansion of the 1960s and early 1970s. By 1973, continued growth and the rapid advance in technology made the design of a new multiplex desirable. As always, the objectives were to reduce costs and improve transmission performance. However, as materials costs were progressively reduced, the costs of field engineering, installation, and operation and maintenance assumed greater relative importance. Special emphasis was directed to these aspects of the LMX-3 design, first shipped in late 1975.[14]

Economy in engineering and installation was achieved by versatile packaging arrangements, suitable for both large and small offices. The total cost for a fully-loaded terminal was lower than that for LMX-2, and the initial cost was kept to a minimum by the almost exclusive use of plug-in units, which could be bought only as needed. Power consumption was one-half of that required by LMX-2. Transmission performance and reliability were both improved; there was no routine maintenance.

For small offices, a combined bay that formed a mastergroup from 50 separate group inputs was provided. In addition, group-bank and supergroup-bank-only bays were provided. The latter, along with group and supergroup distributing frames, offered great flexibility in equipment arrangements. This flexibility improved the efficiency of equipment utilization and reduced the work involved in providing and rearranging circuits.

A block diagram of the LMX-3 terminal would be almost identical to that of LMX-2 or LMX-1; the big difference was in the physical realization of the units. The multiplex functions were provided on plug-in printed-wiring boards that could be used in all bay configurations. The key plug-in unit, in terms of system cost, size, and performance, was the modem, which contained most of the active circuits for both directions of transmission. It included the modulator and demodulator, a common carrier-drive amplifier, the band-pass filters, and a pilot-regulating amplifier. The extensive use of plug-in units, especially the modem with its integral band-pass filters, minimized the initial bay investment. Most of the capital outlay for equipment could be deferred until circuits were needed.

All of the modems for group or supergroup were identical, except for the band-pass filters and the carrier frequency that determined their operating frequencies. The use of transistors in active modulators was instrumental in producing an efficient modem design. The modulator was characterized by low harmonic distortion and low carrier leak. It provided a modest conversion gain that allowed the elimination of an amplifier in the transmit side, and it required very little carrier drive power. The use of an active demodulator in the receiving side provided isolation and permitted the use of filters at both the input (band pass) and output (low pass) with no interaction because of reflected signal energy. Integrated circuits were widely used to realize the circuit functions in a small space and with low power. Extensive use of balanced circuitry reduced crosstalk and even-order distortion.

The carrier supply consisted of two 4-inch-high shelves that, together with a secondary distribution panel in each bay, furnished the carriers and pilots for three U-600 bays (1800 channels). The use of individual, phase-locked oscillators facilitated the generation of extremely pure carriers. There was no need to suppress large amounts of energy at unwanted frequencies.

The combined-, group-, and supergroup-bank bays all contained both the transmitting and receiving directions of transmission in one 7-foot-high, factory-wired frame, to which plug-in modules were added as needed [Fig. 15-15]. The combined bay accommodated up to ten working group banks, plus two spare group banks, one working and one spare supergroup bank, and a carrier supply that served three U-600 bays. As in LMX-2, patching jacks were included in each bay. With A6 channel banks, LMX-3 provided all the equipment needed

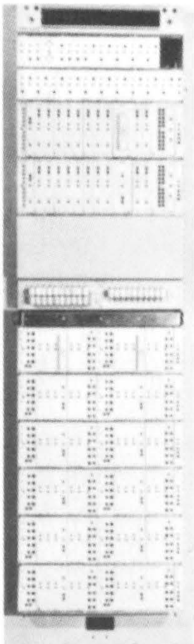


Fig. 15-15. The LMX-3 U600 mastergroup combined bay. All the equipment needed to multiplex and demultiplex 50 twelve-channel groups was contained in a single 7-foot-high bay.

for multiplexing both directions of 600 channels in the equivalent of little more than three 11-foot, 6-inch bays. (By 1975, however, equipment was generally furnished in 7-foot bays). This represented a two-to-one size reduction over LMX-2 and a 20-to-1 reduction over LMX-1, with greatly improved transmission and reliability, as well as large reductions in installation, maintenance, and power costs.

2.4 Mastergroup and Jumbogroup Multiplex

2.4.1 *MMX-2*

When the L4 coaxial system was developed in the mid-1960s, the solid-state A5 channel bank and the LMX-2 group and supergroup multiplex terminals were available to stack up channels into a basic mastergroup of 600 channels. A new mastergroup multiplex terminal, MMX-2, was designed to assemble the 3600 channels for transmission over L4 (the three-mastergroup multiplex (MMX) of L3 was retroactively dubbed MMX-1). MMX-2 multiplexed (and demultiplexed) six basic U-600 mastergroups, each occupying the frequency spectrum from 0.564 to 3.084 MHz, to form a signal extending from 0.564 to 17.548 MHz for transmission on the L4 line [Fig. 15-16]. As radio capacity was expanded in the TH system, a version of MMX-2 that combined three mastergroups was also developed and called MMX-2R (for radio), while the six-mastergroup version was named MMX-2C (for cable).[15]

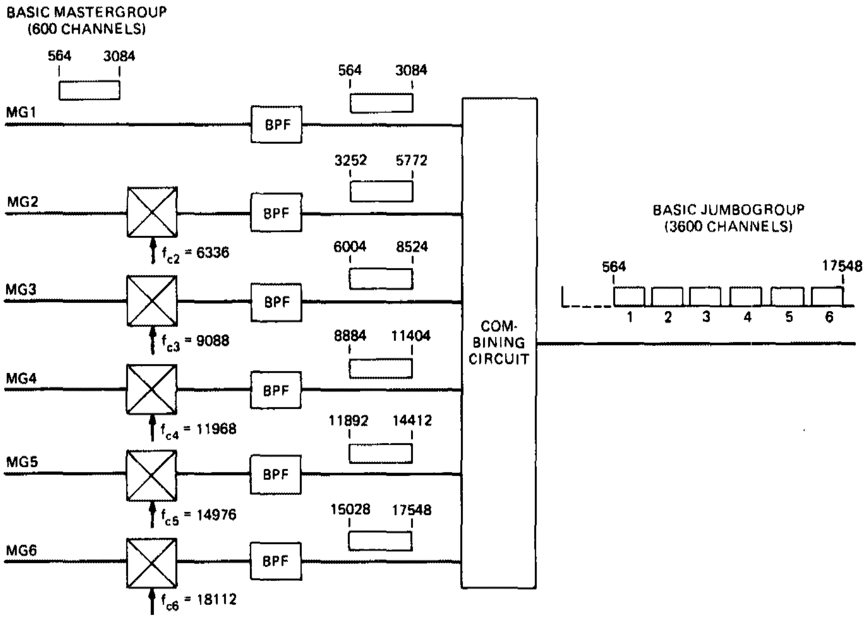


Fig. 15-16. The MMX-2 mastergroup multiplex terminal. Six basic mastergroups were stacked to form a 3600-channel jumbogroup.

The lowest-frequency mastergroup was connected to the line without frequency translation. Stacking of the remaining five mastergroups was by the conventional techniques of mixing with a carrier in a diode rectifier, filtering for the desired sideband, and combining them into the composite signal. The modulating element was a balanced ring consisting of a matched set of four silicon diodes in a single encapsulation. Carrier-drive amplifiers were used to isolate the modulators from each other. Signal amplifiers followed the modulators and the band filters to maintain proper levels. A single design of modulator and signal amplifier served for all mastergroups.

A new method of carrier generation was introduced in MMX-2. Earlier systems had used saturable inductors as nonlinear elements to generate harmonics of the base synchronizing frequency of 64 kHz. In MMX-2, all the binary multiples of 64 kHz were generated by using frequency doublers. Combinations of these frequencies were added or subtracted in modulators for each mastergroup to derive the required carrier frequencies [Fig. 15-17]. The carrier-generating circuitry was duplicated and automatically switched if either generator failed. Meters indicated the relative phase difference between the generators, and, to avoid a phase hit, an adjustment could be made to minimize the difference before forcing a manual switch.

A pilot for automatic regulation was inserted in each mastergroup at 2840 kHz, near the top edge of the band, in keeping with the general plan to leave

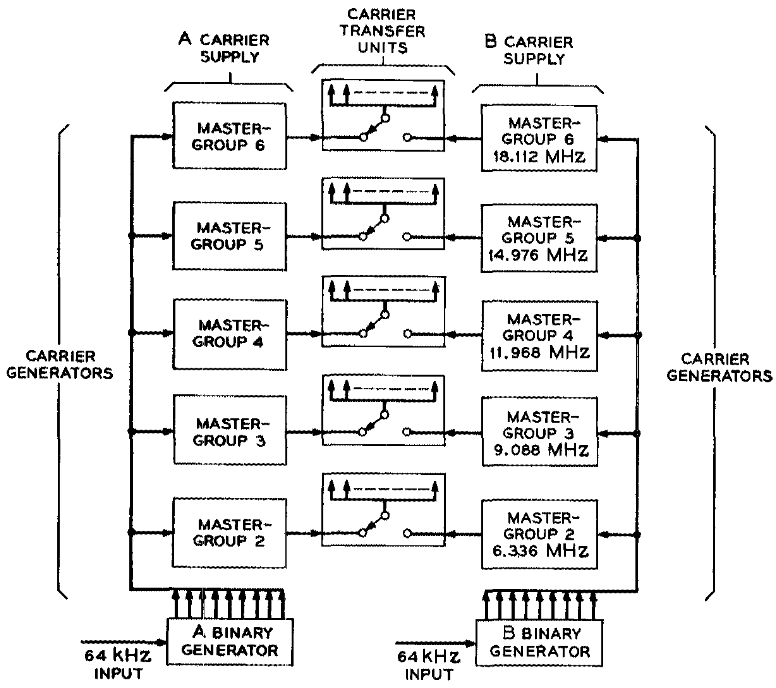


Fig. 15-17. MMX-2 carrier supply arrangement.

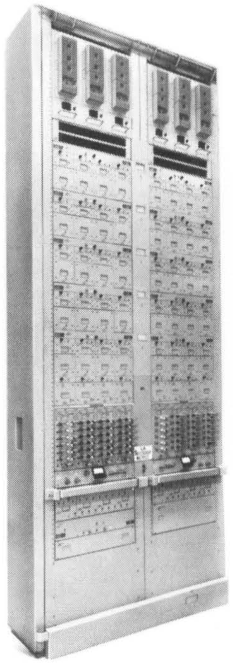


Fig. 15-18. The MMX-2 mastergroup multiplex terminal. The terminal provided all the transmitting and receiving equipment for stacking up to 18 mastergroups into six 3600-channel jumbogroups.

as much of the band as possible free for data. In common with the practice in lower multiplex levels, deviation of the pilot beyond a prescribed range (±3dB) raised a minor alarm. A larger deviation or total loss of the pilot raised a major alarm and, for the first time in a terminal multiplex, initiated an automatic switch to a spare modulating unit. One spare modulating unit was provided for every three working mastergroups at the same line-frequency. The spare units could be locked out of the protection mode and used as a working mastergroup multiplex for alternate routing and restoration of other systems.

The MMX-2 was provided as a shop-assembled, wired, and tested 11-foot, 6-inch double bay that contained the transmission, carrier-supply, switching, regulation, patching, and alarm equipment for a maximum of 18 mastergroups (10,800 channels). It also contained the six spare mastergroup units [Fig. 15-18]. The basic module was a mastergroup shelf assembly containing four (three regular and one spare) plug-in transmission and pilot-detector modules [Fig. 15-19]. A maximum of three MMX-2 bays was required in an office terminating a fully equipped 20-tube L4 cable.

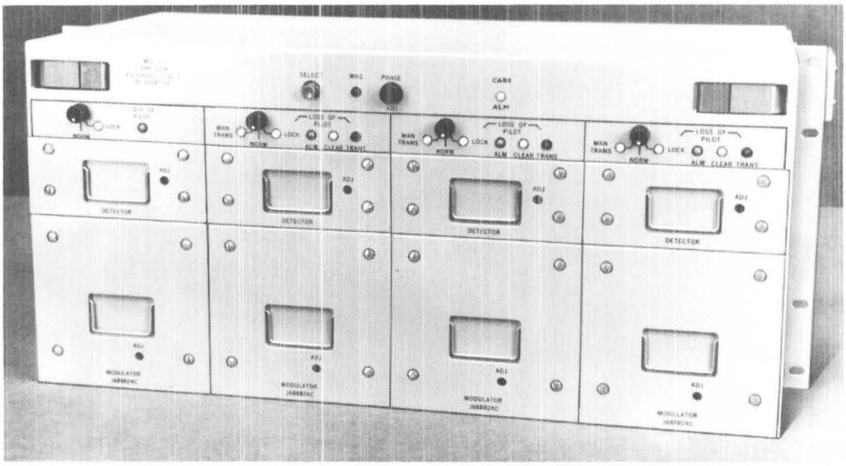


Fig. 15-19. The MMX-2 mastergroup transmitting shelf assembly. The assembly contained three regular units and a switchable spare unit, each of which translated a mastergroup to its jumbogroup frequency slot.

2.4.2 *Jumbogroup Multiplex*

When the L5 coaxial system was planned in the late 1960s with a band of 60 MHz, about three times that of L4, an additional stage of multiplexing was required. Initial plans were to assemble a set of 15 mastergroups (9000 voice channels), starting with the L4 system's six-mastergroup frequency assignment from 0.564 to 17.546 MHz, designated a basic jumbogroup. There were to be 2-1/2 basic jumbogroups translated to form the 15-mastergroup spectrum. During development, the L5 capacity was increased to 18 mastergroups (10,800 voice channels), so the jumbogroup multiplex (JMX) design was changed to multiplex three basic jumbogroups. This arrangement used the MMX-2 multiplex to form the basic jumbogroup.[16]

In the L4 system, the spacing between mastergroups within the basic jumbogroup was wide enough so that filters could be used to select and suppress contiguous sets of mastergroups. This permitted connecting mastergroups between different coaxial and radio systems without requiring mastergroup multiplex equipment. A set of equipment, called the basic jumbogroup trunk bay, was designed for this interconnection. For similar reasons of system interconnection, the L5 line-frequency assignments between 3.124 and 60.556 MHz were spaced to permit single or contiguous jumbogroup interconnection between L5 coaxial cables or routes, without requiring demultiplexing and remultiplexing.

Since the L5 line did not transmit much below 3 MHz, direct connection of a basic jumbogroup, which extended down to 0.564 MHz, was not possible [Fig. 15-20]. It was necessary to shift the first jumbogroup (JG-1) to the range from 3.1 to 20.1 MHz. Furthermore, as the output spectrum substantially overlapped the input signal range, a two-step modulation sequence was necessary

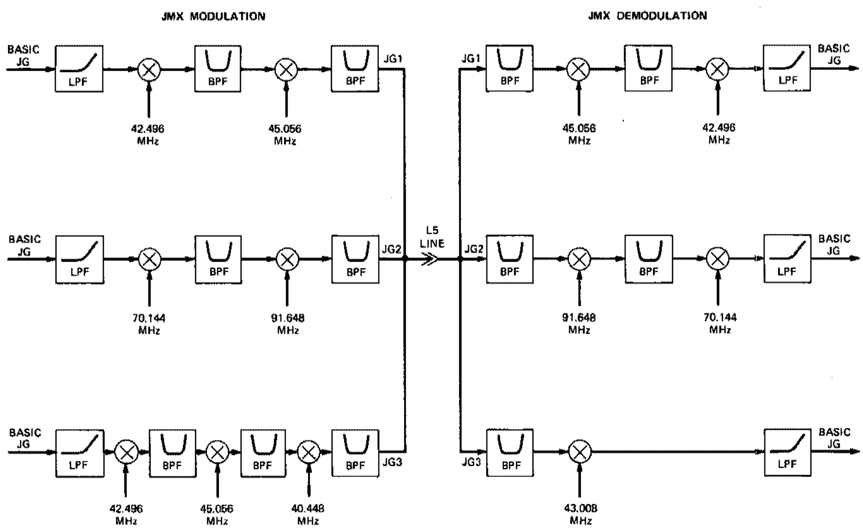


Fig. 15-20. Simplified block diagram of the JMX terminal emphasizing transmission networks.

in which the 17-MHz band was first translated to the band from 24.9 to 41.9 MHz, filtered, and shifted back to 3.1 to 20.1 MHz. For JG-2, even though the overlap was not present, a two-step translation sequence was used to ease filter problems. In JG-3, the modulations used for JG-1 were used, in effect, to form a new basic jumbogroup in the range from 3.1 to 20.1 MHz. This was more amenable to translation by a third modulation step to the top line frequencies than the lower band of frequencies from MMX-2. The three steps of modulation for JG-3 in the transmitting terminal were offset in the receiver, where a single-stage translation to the basic jumbogroup was possible.

The JMX design included automatic-protection switching on a one-for-one basis. Carrier supplies were dedicated to each transmitting and receiving jumbogroup and were automatically switched in and out of service as the signal transmission path was switched by the automatic protection. A synchronization-distribution panel provided accurate reference frequencies from a jumbogroup frequency supply (JFS) for use by the carrier supplies. This panel was also automatically protected. The JFS, with its frequency held to within a few parts in 10^{10} of a network reference signal, supplied the basic frequency-determining signal to as many as 20 JMX bays. (The JFS was made extremely precise and stable as, in addition to its L5 role, it was intended to serve as a regional reference-frequency supply.) A 5.888-MHz pilot in each jumbogroup was used for regulation in the receiving terminal.

Spare equipment and test access were provided to make maintenance and service restoration as easy as possible. Test and monitoring equipment in the office could be connected and programmed from remote automatic-test centers to minimize manual testing at the JMX bay. Other features of the design included equalization for the cables connecting to the basic jumbogroup trunk bay and to the L5 line bays. Matched-length cables were used in conjunction with carrier-phase adjustment to achieve nearly hitless (no transient) switching between in-service and spare equipment.

The large bandwidth of the signals and stringent requirements meant achieving much better performance than had been needed with previous multiplex equipment. Low-noise amplifiers and amplifiers with good modulation performance were developed in HIC technology. Stringent matching requirements were imposed on Schottky diodes in the modulators. Ferrite advances made possible improved filter performance and remarkably good transformers. Broadband hybrids were made small enough to mount immediately behind the jacks on patching bays. The L5 development was not nearly as limited at 70 MHz by transformer performance as L4 had been at 20 MHz. In final back-to-back tests, JMX channel noise was 6 dB better than the original allocation of 18 dBrnC0. Spurious tones, phase jitter, and crosstalk performance were similarly gratifyingly low. This performance was achieved with signal frequencies as high as 91.6 MHz and signal-level differences as great as 128 dB.

2.5 Mastergroup and Multimastergroup Translators

In the late 1960s and early 1970s, a great deal of attention was focused on achieving the maximum capacity in the single-sideband radio system, AR6A. It was finally concluded that as many as 10 mastergroups (6000 channels) could be transmitted in the 30-MHz radio channels of the 6-GHz microwave band, if the spacing between mastergroups could be reduced to 0.168 MHz—much less than in the standard jumbogroup. A little later, favorable experience with the initial design led to a similar interest in increasing the capacity of L5 coaxial by squeezing in more mastergroups. At the same time, advances in device technology and the availability of new circuits, such as precise and inexpensive frequency synthesizers, were reducing the cost of complex circuits and greatly improving reliability. The terminal designed to assemble the more closely packed mastergroups was called the mastergroup translator (MGT).

2.5.1 Mastergroup Translator

In the mastergroup translator, common power and carrier-frequency supplies were abandoned in favor of providing these functions with each mastergroup. The MGT was designed without redundancy or protection, resulting in a two-to-one space savings and considerable savings in costs as well. Alarm systems that localized problems to a single plug-in module enabled a quick repair when failures did occur. Start-up costs were especially reduced, as there was a minimum of common equipment, and mastergroups could be added as needed.

MGTs were designed to assemble basic multimastergroups of seven or eight mastergroups for transmission over L5 coaxial, and of five mastergroups for transmission over single-sideband radio. In all cases, further modulation of the multimastergroup block by a multimastergroup translator (MMGT) was needed for transmission over the line. The increase in packing density was achieved at the expense of the ability to block and branch contiguous sets of mastergroups in the basic multimastergroups. Interconnection between systems, if the entire spectrum could not be accepted, had to be at the basic mastergroup of 0.5 to 3 MHz. The basic mastergroup band was not included in the new multimastergroup spectrums to ease filtering and enable simpler modulation plans in the higher-level multiplex.

2.5.2 *Multimastergroup Translators*

In the MMGT for cable (MMGTC), the first multimastergroup (MMG-1) was connected directly to the L5E line. The seven-mastergroup MMG-2 and eight-mastergroup MMG-3 required only single steps of modulation to their line-frequency assignments. Spacing between multimastergroups at line frequency was adequate to provide blocking and branching of multimastergroups [Fig. 15-21]. At baseband, MMG-1 and MMG-2 could be interconnected or exchanged between systems, but MMG-3, with eight mastergroups, was incompatible with the other two.

MMGTC had automatic switching to a spare and alarms controlled by a microprocessor. Only one spare set of equipment was used to protect 20 regular units. Solid-state switches were used to route the signal to the regular or spare equipment as required. The switches had a loss of less than 1 dB from 1 to 70 MHz. Remote monitoring points were not provided within MMGTC, but remote test access was provided at the MGT bay and at transmitting- or receiving-line terminating bays at flat transmission points. A regular and spare automatically switched carrier supply provided separate feeds for the 21 regular and spare modulation units in MMG-2 and MMG-3 that were needed to load a 22-coaxial-unit L5E line to its full 132,000-circuit capacity. Six 11-foot, 6-inch bays held all the regular and protection MMGTC equipment for the 60 transmitting and receiving MMGs required to fill an L5E system on 22-tube cable.

For single-sideband radio, mastergroup translators were used to form a 5-mastergroup multimastergroup in the range from 8.6 to 21.9 MHz. The mul-

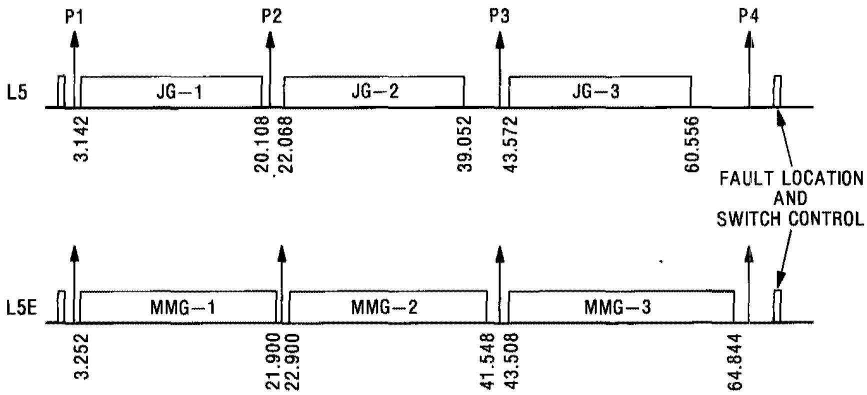


Fig. 15-21. L5 and L5 Expanded (L5E) frequency allocations. In L5E, MMG-1 and MMG-2 had 7 mastergroups and MMG-3 8 mastergroups for a total of 22 mastergroups (13,200) channels per coaxial pair. Nine pilots were unchanged in the expanded system.

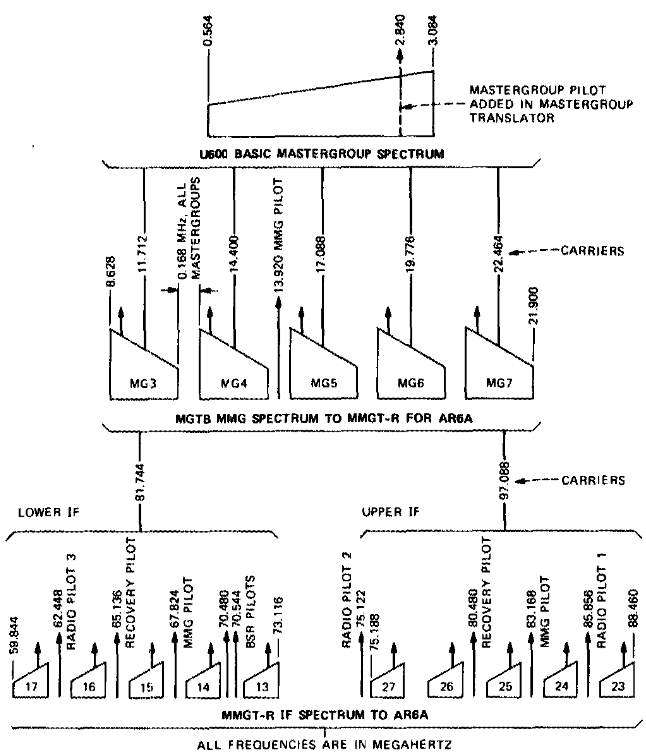


Fig. 15-22. AR6A single-sideband radio terminal frequency allocations using mastergroup translators (MGTS) and multimastergroup translators (MMGTS).

timastergroup translator for radio (MMGTR) then shifted two of these to the radio intermediate frequency band from 59.8 to 88.5 MHz [Fig. 15-22]. The formation of the radio MMGs and line signal used circuits similar to those for cable, with the additional feature of two baseband frequency-recovery pilots. These were compared to a precise, locally generated pilot in the receiver and used to remove the frequency offset and phase jitter acquired during transit through the radio system. MMGTR also supplied additional pilots and test features appropriate to the radio line.

III. INTERCONNECTION AND SYNCHRONIZATION

3.1 Connectors

In 1945, the three principal carrier systems (Types J, K, and L) had the common denominator of 4-kHz voice channels and the basic 12-channel group at 60 to 108 kHz. Channel levels were standardized from the start to permit convenient interconnection at voice frequency. With the rapid increase in circuits, the need soon arose for arrangements to transfer an entire group between systems without the necessity of bringing every channel down to voice frequency. This was the function of a group connector.

Several types of connectors were developed to permit complete flexibility of group connection between the three systems. To provide a flat response across the 48-kHz band while suppressing the pilots and achieving the necessary high out-of-band loss required filters at the forefront of technology. Fortunately, the standard group-receiving level was −5 dB and the transmitting level was −42 dB. With 37 dB of insertion loss available, the required filtering was achieved in a completely passive connector. By 1950, when the need for direct connection of supergroups arose, the designers were not so fortunate. The receiving supergroup level was close to the required transmitting level, and the supergroup connector had to be active to provide the necessary gain to offset the filter-insertion losses.

From 1955 to 1970, the growth of multimastergroup transmission systems generated a need for the cross-system connection of complete mastergroups. Mastergroup connectors were in use by the early 1960s, but the need to provide a flat transmission response across a wide (3-MHz) band and the elimination of in-band pilots again challenged the technology of filter design and component capability.

In the late 1960s, with the advent of wideband data, the group and supergroup connectors were revised to provide delay equalization and to modernize the design. (By the mid-to-late 1970s, connectors had also been developed for group-frequency interconnection between Type L multiplex and the otherwise incompatible N carrier and digital T carrier systems.) The mastergroup connector design was also modernized and the equipment capacity expanded in the early 1970s to accommodate the rapidly growing multimastergroup transmission facilities.

3.2 Distributing Frames

Through-connectors reduced the equipment required and eased the task of network administration but more was needed. The association of terminals and line facilities did not stay put. As the systems grew and traffic patterns changed, it was frequently necessary to reassign blocks of circuits to different facilities. At first, this was done by recabling between the various systems as needed. Needless to say, this was both time- and space-consuming. The cable racks began to fill, making it difficult to run new cables. A group-distributing frame (GDF) to handle the 12-channel circuit bundle was designed to overcome this problem in the 1940s. It allowed reassigning the channel-bank outputs by running jumpers at a centralized cross-connect frame. To facilitate reassigning circuits on an in-service basis, later versions of the GDF provided plug-ended, shielded patch cords and multiple appearances. The GDF became the interface for other circuits besides channel banks and group connectors. It was also used as the access point for program circuits, wideband data, and later as the interface for the connector between digital and analog systems at the group level. A supergroup distributing frame, similar in concept to the GDF, was placed in service shortly after 1950.

The basic 600-channel mastergroup interface in the L-type multiplex hierarchy was the focus of much rearrangement and modification. This was due to the appearance of several multimastergroup radio-line arrangements, mastergroup branching situations, service restoration schemes, the advent of the carrier-transmission maintenance systems, several varieties of mastergroup connectors, and the need for longer intraoffice cables. By 1970, a mastergroup distributing frame (MGDF) was designed to simplify and organize all of these applications and to provide a versatile cross-connect point at the basic mastergroup interface. With the MGDF, only passive cable was required between the cross-connect frame and the LMX-3 bay, the mastergroup translator, or the passive master-group connector. Trunks with gain and/or equalization were required for links to LMX-1 and LMX-2 equipment but were of rather simple design. (It would be interesting to list the variety of items to which the term *trunk* has been applied in telecommunications transmission and switching, but it probably would not reduce the confusion.) Components in the MGDF accommodated all the existing mastergroup formats. These interfaces were contained in plug-in housings that could be added as required for circuit growth. The cross-connect cables were plug-ended miniature coaxial cables.

All of the distributing frames were of the same basic structure, provided in various heights from 7 feet to 11 feet, 6 inches.

3.3 Synchronization

In suppressed-carrier transmission, it was necessary to supply accurate and stable carriers to the receiver demodulators in order to recover the baseband signal without excessive frequency offset or phase instability. Subjective tests in the late 1930s and 1940s established that the frequency offset for voice and program signals should not exceed ±2 Hz. In the Types K and J systems, with maximum modulation frequencies of less than 0.5 MHz, the requirement was met with independent, tuning-fork-controlled oscillators having a frequency accuracy of about 1 part in 10^6. When coaxial systems brought much higher frequencies and the need for correspondingly greater accuracy, the relative accuracy of the carriers was maintained by a synchronizing pilot transmitted over the line from the originating carrier terminal [Fig. 15-23]. Each terminal office had its own primary frequency supply (PFS) (i.e., primary to that office) that generated reference signals for the local multiplex carrier supply and regenerated the synchronization pilot for transmission to terminals further along. The PFS was readjusted as required by comparing a locally generated 64-kHz signal with the incoming 64-kHz pilot. As the radio and coaxial network served by L multiplex terminals expanded, the arrangements for generating, disseminating, and stabilizing the synchronization pilots developed into a nationwide synchronization network.

An early PFS (PFS-1, 1941) consisted of a stable electron-tube oscillator in a bridge-type configuration.[17] Oscillator fine tuning was provided by a variable capacitor under servomotor control. Very small differences in frequency would unbalance the bridge and the servo would act to restore balance. The PFS-1 servo-control had a high-ratio gear reduction resulting in a loop-time constant of about 15 minutes. Early versions of this PFS-1 operated as part of the L1 coaxial system. After 1950, PFS-1 was widely deployed in the network as 64 kHz was also adopted as the terminal synchronization pilot for TD-2 microwave

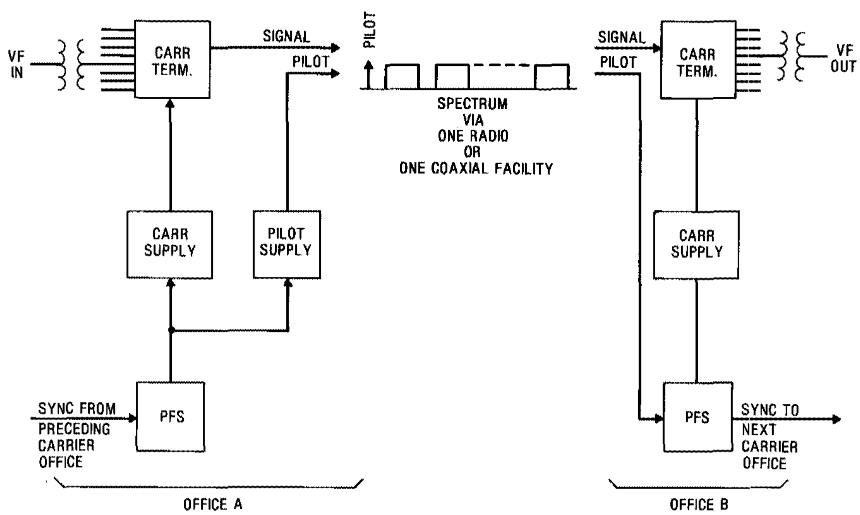


Fig. 15-23. Basic plan for transmitting synchronization, 1950.

radio as well as for coaxial systems. In 1953, PFS-1 was adapted to work with the 308-kHz synchronization pilot of the L3 coaxial system. This version was also widely deployed, as 308-kHz synchronization was used in TH microwave radio, as well as in L3.

Accuracy requirements for the early primary frequency supplies were relatively modest. For frequency offset to be limited to less than ± 2 Hz at the highest-frequency message channel of a single mastergroup spectrum (3.1 MHz), the relative frequency difference between transmitting and receiving carrier terminals could not exceed 7 parts in 10^7. For the three-mastergroup L3 and TH systems, the difference had to be held to one-third of this. With high-quality temperature-controlled crystal oscillators, PFS-1 systems were capable of holding frequencies to within a few parts in 10^8.

The error limits described above applied to the relative frequency offset of one PFS with respect to another and did not address the question of absolute accuracy. For many years, it was sufficient that absolute frequency accuracy would maintain pilots and signaling tones within the appropriate region of their filter passbands. This was usually satisfied by an absolute frequency accuracy in the order of one part in 10^6. For absolute, as well as relative, accuracy a Bell System frequency standard was established at Bell Laboratories in Murray Hill, New Jersey. The standard consisted of three specially designed crystal oscillators that were checked and adjusted to the national standard maintained by the United States Bureau of Standards and the Navy.[18,19] From Murray Hill, a 4-kHz reference frequency was transmitted to the Long Lines building at 32 Sixth Avenue in New York (New York-4) for control of the synchronization pilots, which were transmitted from there to the operating systems.

The master frequency supply in the Long Lines New York-4 office became the frequency source for the synchronization network. The master supply was

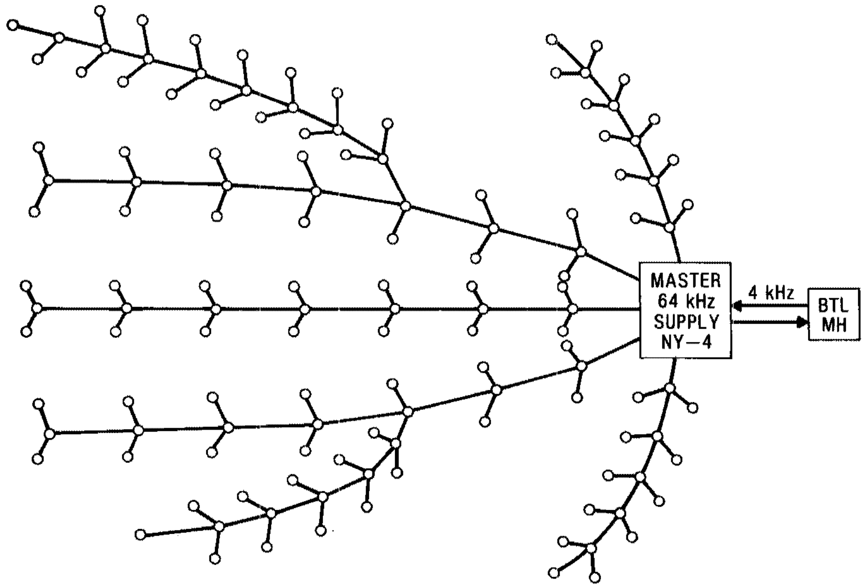


Fig. 15-24. Conceptual view of the 1950s synchronization network.

a highly stable, bridge-controlled crystal oscillator. As a comparison check a 4-kHz signal derived from the master supply was transmitted back to Murray Hill and compared in accuracy to the original Bell Laboratories reference. In this way, the master supply was maintained within a few parts in 10^9 with respect to the national frequency standard monitored at Murray Hill. The synchronization network homed on the master frequency supply in a serial, tree-like topology [Fig. 15-24].

3.3.1 A Phase-Locked Primary Frequency Supply

This method of synchronization worked well for many years, but as the synchronizing network spread and branched westward from New York, situations eventually developed where the pilot could be regenerated as often as 10 or 20 times. With the limited accuracy and slow response time of PFS-1, it became difficult or impossible to maintain the required offset limit in all cases. As a remedy to this situation, when the L multiplex terminals were redesigned in the early 1960s, a new primary frequency supply, designated PFS-2, was developed as a successor to PFS-1. In PFS-2 a phase-locked control loop was used to assure zero frequency offset between the regenerated pilot and the incoming synchronization pilot, and equivalent accuracy in the primary frequencies for the local carrier supply [Fig. 15-25].[20,21,22] In the absence of a synchronization pilot, phase-lock would be lost, and the PFS-2 would free-run at the accuracy of the local crystal oscillator. PFS-2 was widely installed in the network during the large facility expansion of the 1960s and early 1970s.

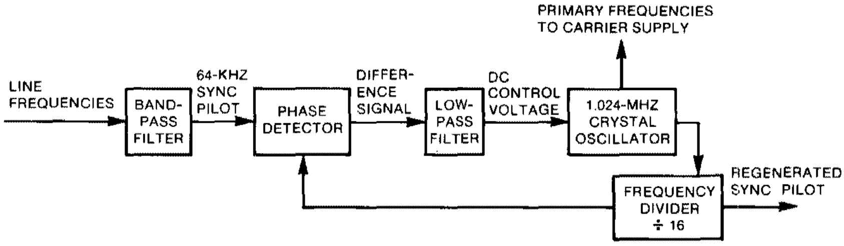


Fig. 15-25. The primary frequency supply, PFS-2, with phase-locked frequency control, 1962.

Unfortunately, the reduction of the frequency offset was obtained at a price. The time constant of the phase-locked loop was about one second, in contrast with the 15-minute response of PFS-1. Facility switching or other maintenance activity could interrupt the pilots and introduce phase hits (transients in the pilot continuity). The older system, with its slow response, largely ignored these hits, but the much stiffer PFS-2 network propagated shock waves through the system. The short time constant of PFS-2 allowed some hits to be propagated through the PFS into carrier supplies, where the modulation process would transfer the hits into the signal paths. An action that may have affected only the pilot or a few circuits could, by this process, disturb all the circuits served by the carrier supply. A phase hit was not a serious impairment to speech transmission, but it could cause errors in data transmission. As more and more of the network became involved in the transmission of data, the consequences of synchronization phase hits became more serious.

Frequency-offset problems also reappeared. PFS-2 was redesigned in 1966 to operate with the 512-kHz pilot of the L4 system. Because L4 had a top frequency of 17.5 MHz, the relative error between two L4 PFSs had to be less than about one part in 10^7 to satisfy the channel-frequency offset limit. It became increasingly apparent that, when it was free running, PFS-2 could not maintain offset limits in the upper mastergroups of L4 under all conditions.

There were problems in addition to the phase hits and frequency accuracy. By the end of the 1960s, the synchronization network was becoming increasingly complex to administer. The number of PFSs in tandem had become unwieldy as the network experienced continued growth. (One transcontinental synchronization route grew to contain 19 tandem PFSs.) The existence of three different synchronization pilots was a complicating factor. The PFS-1 systems still in the network were growing old, some having been in service for 25 years. The performance of these older motor-driven units was often less than optimal. Finally, the development of the L5 and L5E systems in the early 1970s presented new and tighter requirements for synchronization. With a top message frequency of about 65 MHz, frequency-accuracy requirements of a few parts in 10^8 were essential. The old synchronization network could not be depended upon to satisfy such requirements, and a new approach to synchronization was developed.

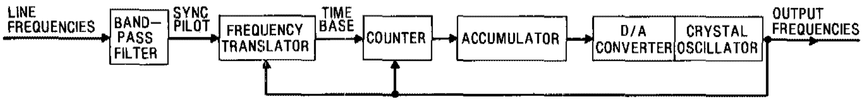


Fig. 15-26. Quasi-frequency-lock control for a regional reference frequency standard, 1975.

3.3.2 *Quasi-Frequency Lock*

The new approach to network synchronization was designed around the jumbogroup frequency supply (JFS) developed in 1974 for the L5 coaxial system; it was used in a modified form as a regional reference supply. The JFS employed three crystal oscillators of a new, improved design.[23] When free running, it had a drift rate of less than 1 part in 10^{10} per day, low enough to operate without correction for several weeks and still meet L5 frequency-accuracy requirements. The oscillator was buffered from transients in the incoming synchronization signal by a method called quasi-frequency lock (also referred to as plesiochronous, where Greek terms are preferred). In this method, cycles of difference between the local signal and the incoming reference were counted and accumulated; no correction was made until the count reached 256. Each time the accumulator filled, a correction of only 2 parts in 10^{10} was made in the appropriate direction [Fig. 15-26]. Under conditions of normal operation, corrections occurred only at intervals of an hour or more. When gross differences between the local and incoming reference occurred, providing the local oscillators were in agreement and appeared healthy, the regional standard disconnected from the network and ran free. The output of the regional supply was normally maintained to within ± 3 parts in 10^{10} of the input signal frequency.

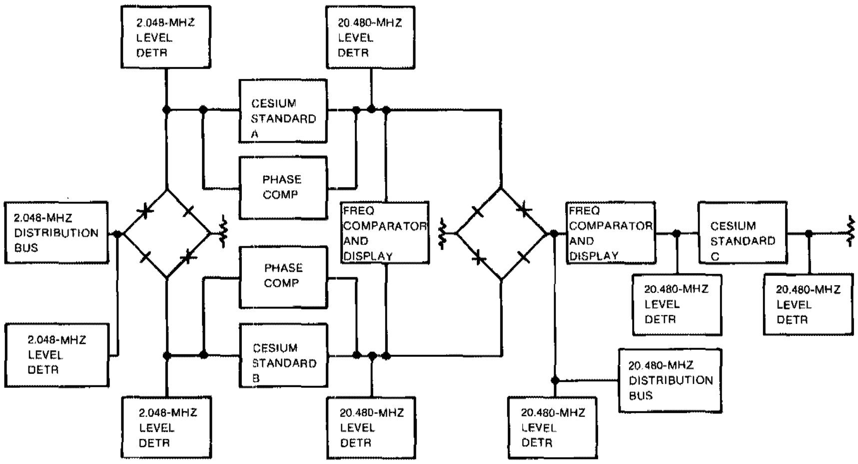


Fig. 15-27. The Bell System reference frequency standard using three cesium atomic standard oscillators, 1974.

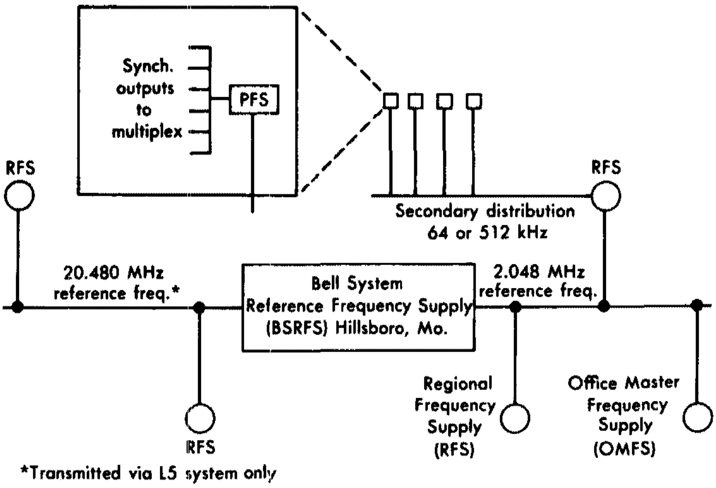
To realize the improved performance, the JFS required a more accurate and stable reference signal than could be provided by the existing synchronization network. Frequency standards had evolved considerably since the design of the crystal-controlled, vacuum-tube oscillators of the late 1940s. By the 1970s, cesium atomic frequency standards with an absolute frequency accuracy of ± 1 part in 10^{11} were commercially available. A new Bell System reference-frequency standard (BSRFS), using three cesium atomic standards, was developed and installed at an underground L5 station at Hillsboro, Missouri [Fig. 15-27]. From there, coaxial and microwave-radio facilities fanned out to carry reference frequencies from the BSRFS to regional JFSs conveniently located throughout the country. The JFSs in turn distributed synchronization to conventional PFSs in their region [Fig. 15-28].[24] A loran receiver at Hillsboro permitted comparison between the Hillsboro standard and the transmissions of the loran C station in Dana, Indiana. Loran C transmissions were referenced to the Naval Observatory's master cesium clock. The Bell System standard was typically within a few parts in 10^{12} of the national standard.

Two reference frequencies, 2.048 and 20.480 MHz, were used to transmit the accuracy of the system reference to the regional JFSs. These frequencies fell in guard bands in the commonly used frequency spectra of broadband analog carrier systems and were used only for reference-frequency transmission. The 2.048-MHz reference was transmitted over all coaxial and radio systems, except for L5, where the 20.480-MHz reference was used. Both references were transmitted without modulation or regeneration and thus always conveyed the accuracy of the network standard.

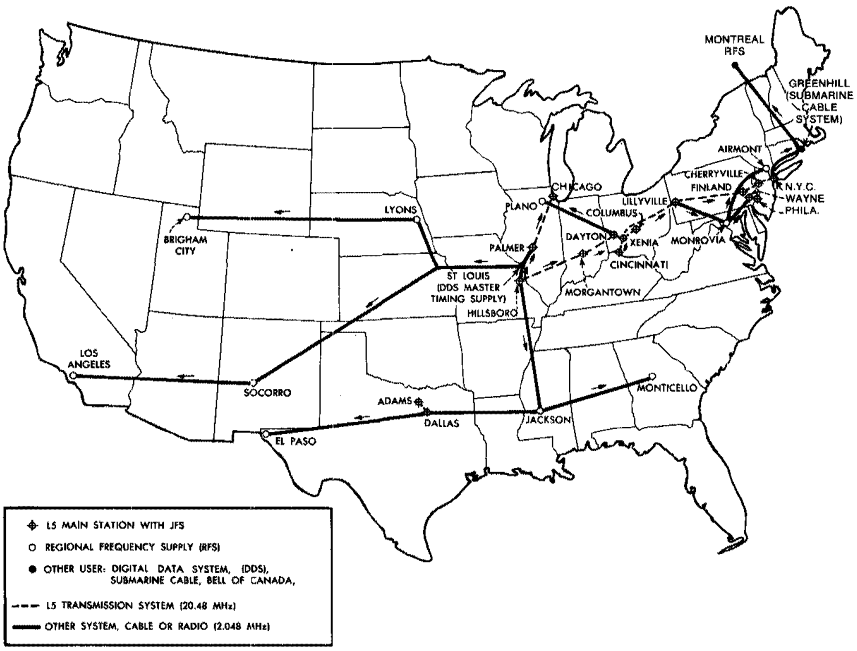
3.3.3 *Synchronization in AR6A Radio Systems*

FM radio systems such as TD and TH transmitted a radio-frequency carrier that in effect provided synchronization from terminal to terminal. The AR6A radio system, with single-sideband AM transmission, had no radio-frequency carrier. Oscillators at AR6A repeater sites were free running; radio-frequency synchronization from repeater to repeater was not required, but frequency synchronization between multiplex terminals was still needed. A two-tone frequency reference was provided on AR6A for this purpose. The two-tone reference frequencies were at 11.200 and 11.264 MHz, between mastergroups in the AR6A baseband spectrum. The 64-kHz difference between the frequencies was derived from the 2.048-MHz BSRFS. As the two tones traversed the AR6A system, their absolute frequencies changed, reflecting the normal drift of repeater oscillators, but the 64-kHz frequency difference remained constant. The frequency difference thus conveyed the accuracy of the reference standard to the receiving terminal and provided the means for frequency control at multiplex terminals.

Quasi-frequency lock proved very successful in the JFS regional frequency-supply application. It was evident that the same features of phase-hit buffering and accurate free running could benefit individual central offices. An office master-frequency supply (OMFS), initially deployed with the AR6A radio system, was developed to provide these features in a more modest package than the JFS. For AR6A, the OMFS was provided with two precision crystal oscillators in quasi-frequency lock to a 2.048-MHz input reference derived from the AR6A two-tone reference frequencies. Output frequencies of the OMFS were 64 and



(a)



(b)

Fig. 15-28. Carrier system synchronization, 1975. (a) Conceptual view of the synchronizing network. (b) The 1975 Bell System synchronization network.

512 kHz, which served to provide synchronization to multiplex terminals and to PFS-2 systems. The quasi-frequency lock of the OMFS maintained the output frequencies to within ±5 parts in 10^{10}.

REFERENCES

1. R. W. Chesnut, L. M. Ilgenfritz,and A. Kener, "Cable Carrier Telephone Terminals," *Bell Syst. Tech. J.* **17** (January 1938), pp. 106-124.
2. J. L. Snoek, "Non-Metallic Magnetic Materials for High Frequencies," *Philips Tech. Rev.* **8** (December 1946), pp. 353-360.
3. H. A. Stone, Jr., "Ferrite Core Inductors," *Bell Syst. Tech. J.* **32** (March 1953), pp. 265-291.
4. F. H. Blecher and F. J. Hallenbeck, "The Transistorized A5 Channel Bank for Broadband Systems," *Bell Syst. Tech. J.* **41** (January 1962), pp. 321-359.
5. W. P. Mason, "Quartz Crystal Applications," *Bell Syst. Tech. J.* **22** (July 1943), pp. 178-223.
6. R. A. Sykes and W. L. Smith, "A Monolithic Crystal Filter," *Bell Lab. Rec.* **46** (February 1968), pp. 52-54.
7. G. W. Bleisch and W. P. Michaud, Jr., "The A-6 Channel Bank: Putting New Technologies To Work," *Bell Lab. Rec.* **49** (September 1971), pp. 251-254.
8. C. H. Elmendorf, R. D. Ehrbar, R. H. Klie, and A. J. Grossman, "The L3 Coaxial System—System Design," *Bell Syst. Tech. J.* **32** (July 1953), pp. 820-825.
9. A. A. Alexander, R. M. Gryb, and D. W. Nast, "Capabilities of the Telephone Network for Data Transmission," *Bell Syst. Tech. J.* **39** (May 1960), pp. 431-476.
10. F. J. Hallenbeck and J. J. Mahoney, Jr., "The New L Multiplex—System Description and Design Objectives," *Bell Syst. Tech. J.* **42** (March 1963), pp. 207-221.
11. J. J. Mahoney, Jr., "New Multiplex For Long Distance Service," *Bell Lab. Rec.* **42** (January 1964), pp. 32-38.
12. R. S. Graham, W. E. Adams, R. E. Powers, and F. R. Bies, "New Group and Supergroup Terminals for L Multiplex," *Bell Syst. Tech. J.* **42** (March 1963), pp. 223-278.
13. W. G. Albert, J. B. Evans, Jr., J. J. Ginty, and J. B. Harley, "Carrier Supplies for L-Type Multiplex," *Bell Syst. Tech. J.* **42** (March 1963), pp. 279-317.
14. R. W. Couture, C. V. Fanuele, and T. R. Helmes, "The New LMX-3 Multiplex Terminals," *IEEE International Conference on Communication* **1**, San Francisco (June 16-18, 1975), pp. 8-1–8-5.
15. W. G. Albert, J. B. Evans, Jr., T. J. Haley, T. B. Merrick, and T. H. Simmonds, Jr., "L-4 System: Terminal Arrangements," *Bell Syst. Tech. J.* **48** (April 1969), pp. 993-1040.
16. R. E. Maurer, "L5 System: Jumbogroup Multiplex Terminal," *Bell Syst. Tech. J.* **53** (December 1974), pp. 2065-2096.
17. L. A. Meacham, "The Bridge Stabilized Oscillator," *Bell Syst. Tech. J.* **17** (October 1938), pp. 574-591.
18. R. R. Stone, W. Markowitz, Jr., and R. G. Hall, "Time and Frequency Synchronization of Navy VLF Transmission," *IRE Trans. Instrum.* **I-9** (September 1960), pp. 155-161.
19. R. R. Stone, Jr., "Synchronization of Local Frequency Standards with VLF Transmissions," *Proceedings of the 16th Annual Symposium on Frequency Control*, Atlantic City, NJ (April 25-27, 1962), pp. 227-240.
20. W. A. Kester, "A New Reference Frequency Standard for the L Multiplex System," *Bell Syst. Tech. J.* **47** (December 1968), pp. 2183-2205.
21. O. P. Clark, E. J. Drazy, and D. C. Weller, "A Phase-Locked Primary Frequency Supply for the L Multiplex," *Bell Syst. Tech. J.* **42** (March 1963), pp. 319-340.
22. J. G. Shields, D. C. Weller, and W. R. Wysocki, "A Phase-Locked Primary Frequency Supply," *Bell Lab. Rec.* **44** (July/August 1966), pp. 236-238.
23. A. F. Flint and H. S. Pustarfi, "L5 System: 39A Precision Oscillator," *Bell Syst. Tech. J.* **53** (December 1974), pp. 2097-2107.
24. J. F. Oberst, "Keeping Bell System Frequencies On the Beam," *Bell Lab. Rec.* **52** (March 1974), pp. 84-89.

Chapter 16

The Subscriber-Loop Network

I. INTRODUCTION

The subscriber-loop network consists of the plant equipment and systems that connect the customer with the central office. Of the huge investment in outside plant, which includes interoffice trunk lines, toll cables, and radio links it is by far the largest part, representing 75 percent of the total. It accounts for almost half the total investment required to provide exchange-area service. The large amount of equipment required has made this part of the plant the subject of extensive research in materials and increasing design and development effort to improve methods and systems and to minimize the costs involved.

A typical subscriber loop consists of inside wiring on a customer's premises between a telephone set and a point of entry, normally equipped with electrical protection. From there, one pair or a few pairs of wires connect with a telephone cable (a distribution cable) at a point called the distribution terminal. In 1925, this terminal was located on a utility pole and the connecting link was called a drop wire. By 1975, the connecting link was frequently buried in the ground and referred to as a service wire. The distribution cable would eventually connect with a larger cable through a splice or, in later years, an access point. The larger cable (a feeder cable) led to the central office, entering through a cable vault and ending on a distribution frame designed to permit interconnection between any pair and terminal of the switching equipment [Fig. 16-1]. In 1925, open-wire lines dominated the field for long loops; short loops in big cities often used multipair underground cable. (See an earlier volume in this series, *The Early Years (1875–1925)*, Chapter 4.)

Every subscriber loop had to transmit two types of signals: audio-frequency voice messages and supervisory signals, including changes in DC state to convey information such as "customer is off-hook," as well as dial tone and ringing current. For reliable supervision, the maximum DC resistance of a subscriber loop was standardized at 1300 Ω for many years, limiting loop length to 25,000 feet (about 5 miles) for No. 24 gauge and 75,000 feet for No. 19 gauge wires. Loops in cable longer than 18,000 feet were loaded to keep the loss at message frequencies to less than 10 dB.

From about 10 million in 1930, the number of loops in the Bell System grew to over 40 million by 1960 [Fig. 16-2]. In the 1960s and 1970s, the average length was about 11,500 feet, with 2,000 feet in distribution cable and the remainder in feeders but with an extremely wide range of variation. From 1925

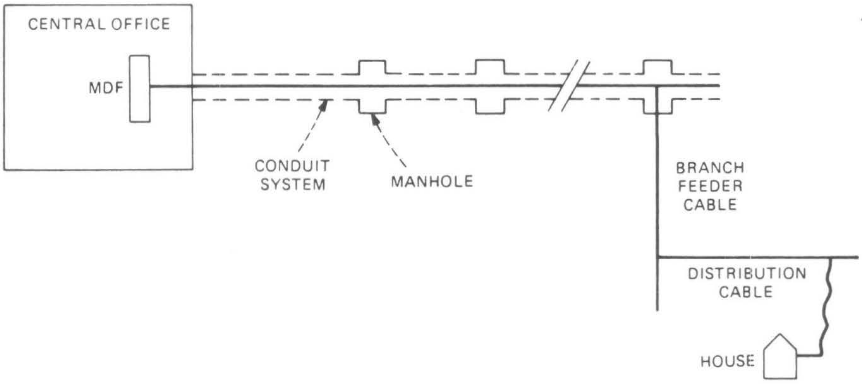


Fig. 16-1. Typical layout of a subscriber loop, 1975.

to 1975, the challenge to improve the subscriber-loop network was threefold: (1) provide universal service of high quality and reliability at a reasonable cost and make single-party service available to all who want it, (2) extend service

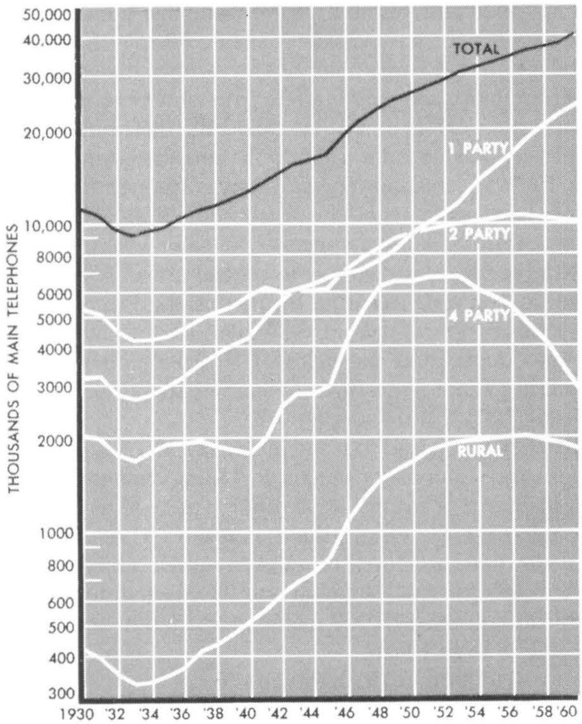


Fig. 16-2. Growth of telephone service from 1930 to 1960.

to rural areas, and (3) provide flexibility to permit the introduction of new services and new technology.

In contrast to other areas of transmission, which occasionally saw such giant technological steps as satellite systems, the development of the loop network took a more gradual approach, characterized by incremental improvements in materials, components, and methods as well as by the incorporation of high-technology systems. This was due largely to the necessity to deal with the circuits one at a time, in contrast to the bulk handling possible in the trunk plant.

1.1 Characteristics of the Subscriber-Loop Network

The subscriber-loop network exhibits a number of unique characteristics. (1) It covers large areas in frequently hostile environments. Ice, heat, dust, chemicals, vibration, and vandalism threaten the equipment. (2) It is custom engineered and installed to be left in place until it wears out. Improvements tend to be added on. It is built mile by mile in a labor-intensive process. (3) It must be designed to be reconfigured quickly and must be flexible to accommodate the daily addition of new customers and cancellations of service. Planning and design must anticipate future demand. When a customer requests service, it is too late to start installing new cable.

II. EVOLUTION OF THE SUBSCRIBER-LOOP NETWORK

Although new telephones were being installed in ever-increasing numbers in the early years, universal service was still a distant goal. In 1925, there were only about 15 telephones per 100 population. Therefore, the subscriber-loop

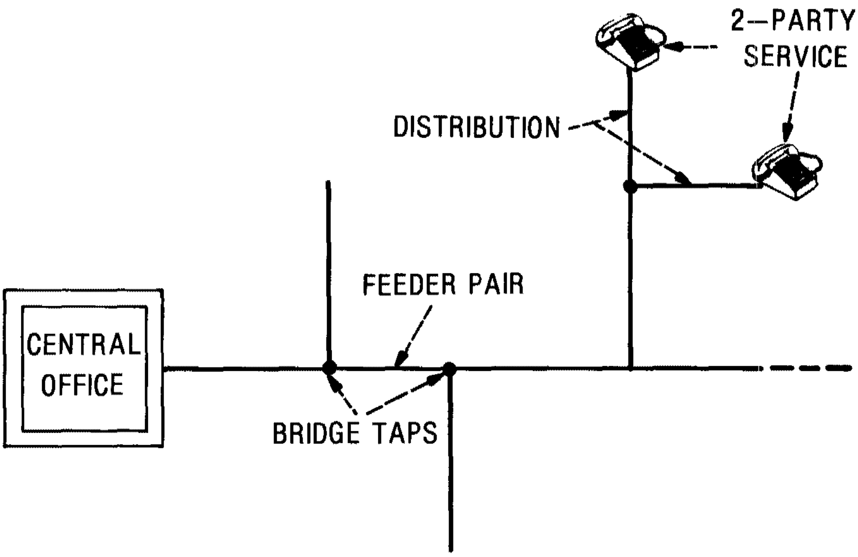


Fig. 16-3. The multiple plant concept.

network was designed to permit rapid addition of new customers through the use of the multiple-plant concept [Fig. 16-3]. Telephone cables and open-wire lines were bridged repeatedly to permit access wherever a new customer demanded service. When a private line was not available, two-, four-, and even eight-party service was offered. The system worked well as long as telephone penetration was well below 100 percent, and people were satisfied with multiparty lines. However, after World War II, the greater mobility of the population and a growing demand for telephones and private-line service proved to be incompatible with the multiple-plant approach, which relied on installing less than one wire pair for each potential customer.

The dedicated outside-plant concept was introduced in 1962.[1,2] It provided one permanently connected pair for each residential housing unit or single-party business location [Fig. 16-4]. Spare pairs, control, and access points were provided for future demands. The plan worked well when enough wire pairs were provided to serve existing and anticipated demand. However, it proved to be somewhat inflexible and not adaptable to unexpected growth. This led to the introduction of so-called permanent plant and the serving-area concept (SAC) in 1971 [Fig. 16-5]. Originally developed for suburban communities, SAC was later modified to meet most requirements of the subscriber-loop network, including special adaptations to rural areas.

SAC embodied two new concepts. One was the separation of the subscriber-loop network into a feeder plant sized for three- to five-years growth and a distribution plant installed only once and sized for ultimate demand. These joined at the feeder-distribution interface (FDI, also called serving area interface, SAI). The other was to assign a specific geographic area to each FDI to simplify record keeping and administration. SAC gave each potential customer an individual distribution pair or pairs, whereas feeder plant could be added economically as needed at appropriate intervals. Permanent plant was uniquely suitable for introducing loop carrier systems and placing plant out of sight.

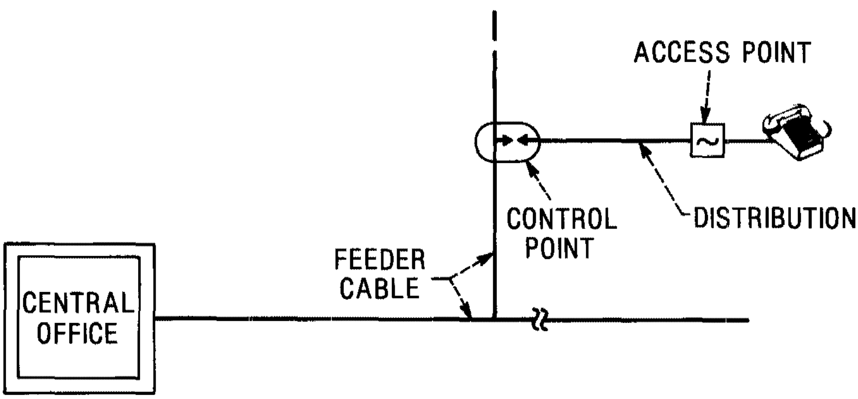


Fig. 16-4. Dedicated outside plant.

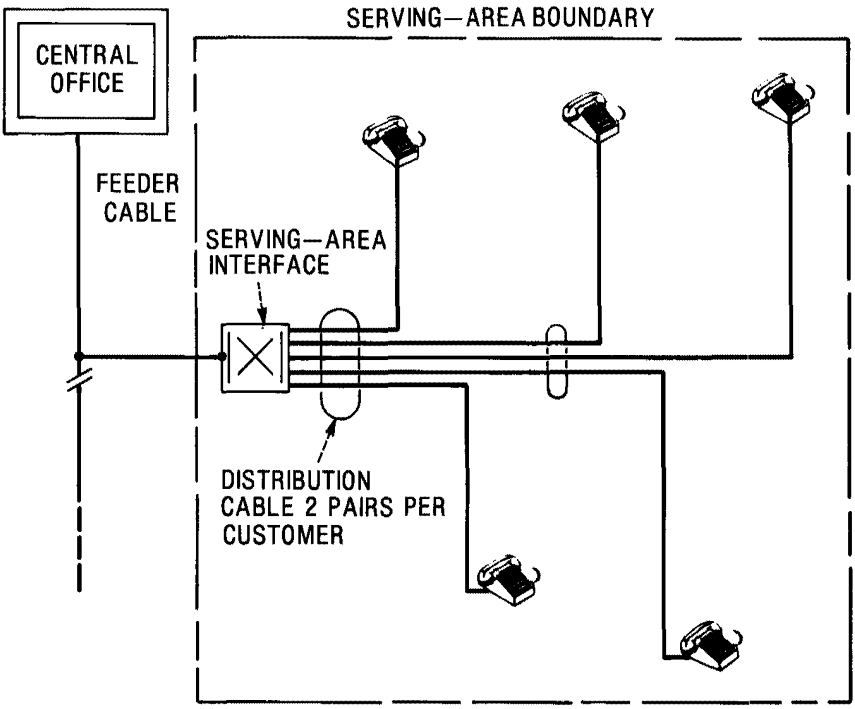


Fig. 16-5. Permanent plant and the serving-area concept.

III. COMPONENT DEVELOPMENT

An important feature of the evolution of the loop network was the development of better cables with new stranding methods and greatly improved insulation and sheaths. In addition, better splicing methods, connectors, splice closures, terminals, and interfaces were developed.

3.1 Cable and Wire Development

The main goals of cable development from 1925 to 1975 were cost reduction and improved durability. By 1925, the capacitance of a wire pair in cable had been standardized at 0.083 μf per mile, and transmission characteristics were predictable and stable.

In the early 1920s, the standard cable design in the Bell System consisted of copper conductors with helically wrapped paper insulation, twisted into pairs. The cable cores were built up in concentric layers with the layers helically stranded to improve mechanical flexibility. The core was then covered with an extruded lead sheath for mechanical and electrical protection. The cables were relatively short (the average subscriber loop was about two miles long) and had to connect with large numbers of subscribers. This led to large-sized cables with many pairs of fine-gauge wires in the feeder plant. In time, it became apparent that the concentric-layer design placed restrictions on cable size (i.e., the number of pairs in the cable) because the cable became quite rigid and difficult to manufacture and to place. Unit construction, introduced in 1928, improved this situation substantially. In this design, the first step was to strand the pairs into a unit (binder group) of up to 101 pairs. Several units were then stranded into a completed core. This structure greatly improved the ease of manufacture and flexibility and, from that time on, was used for all large paired-wire cables [Fig. 16-6].

Other improvements in cable manufacture included the introduction of continuously applied rubber for the insulation of drop wire in 1930 and replacement of paper wrapping with pulp insulation in 1933. The main advantages of pulp insulation were ease of fabrication and cost reduction while retaining a high insulation quality.

In the late 1930s, the British developed polyethylene, a new plastic material with a low dielectric constant and insulating properties far superior to paper. Polyethylene was strong, impervious to water, and easily colored, but in the early postwar years it was expensive and hard to get. Nevertheless, its superior qualities invited the use of polyethylene as an insulation for exchange cable. Early applications included No. 19 gauge, small-size exchange cables (up to 101 pairs). In 1956, fully-color-coded polyethylene-insulated conductor (PIC) cables were made generally available in all four common gauges (Nos. 19, 22, 24, 26).

Another important application for polyethylene was as a substitute for lead, which, until the availability of the plastic, had been the only satisfactory material for cable sheaths since the earliest days of telephony. By the end of World War II, lead was scarce and its price was rising rapidly. In 1947, a composite sheath having a corrugated aluminum shield and an outer black-polyethylene jacket (alpeth) was introduced. Some mechanical limitations, however, soon became apparent, and a more rugged sheath, stalpeth (steel-aluminum-polyethylene), was introduced in 1949. Variations on these composite metal-plastic sheath

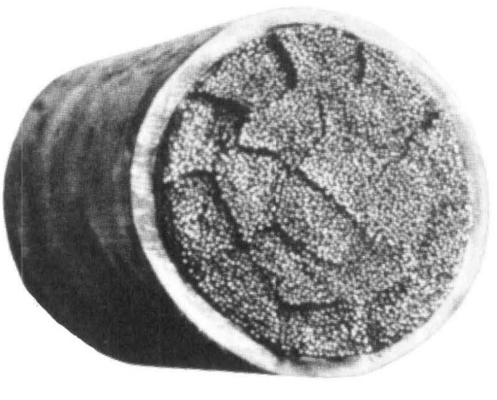


Fig. 16-6. Unit construction applied to an 1800-pair cable manufactured around 1929.

types were developed for specific situations in the years that followed [Fig. 16-7]. The conversion from lead jackets to metal-plastic composite sheath for new cable was virtually total within a few years. It saved billions of dollars in raw materials alone.

In the early 1960s, considerable work was done on the use of expanded plastic insulation in exchange cable. Since the dielectric constant of all plastic insulating materials is higher than that of air, either lower pair capacitance or lower insulation thickness for the same capacitance can be obtained by introducing tiny air bubbles into the molten plastic during the fabrication process to obtain a foam-like or expanded coating. To improve the mechanical strength of this insulation, a thin, solid outer skin covered the foamed inner core. For the same mutual capacitance in a finished cable, the dual-expanded PIC (DEPIC) cable cost less to produce and allowed more pairs in the same-size sheath compared to conventionally insulated wires [Fig. 16-8].

PIC cable revolutionized loop plant design and construction.[3] Its weather-resistance permitted simpler splices and terminals. Direct burial in the ground without pressurization started in 1956. With time, however, it was found that, on occasion, water could enter a cable sheath through an opening or crack, migrate within the cable core for a considerable distance, and finally gain access to the copper wires at splices or through tiny pinholes in the insulation to produce a short circuit. (One of the few advantages of pulp insulation was that, if the cable core became wet, it swelled and blocked water migration.) If

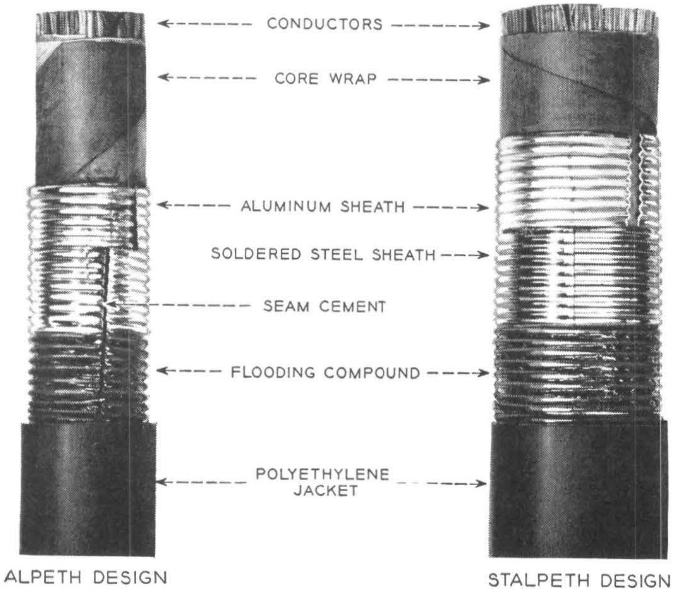


Fig. 16-7. Composite metal-plastic cable sheaths, 1947–1949. Alpeth eliminated the need for lead in sheaths. Stalpeth added a steel sheath for improved strength.

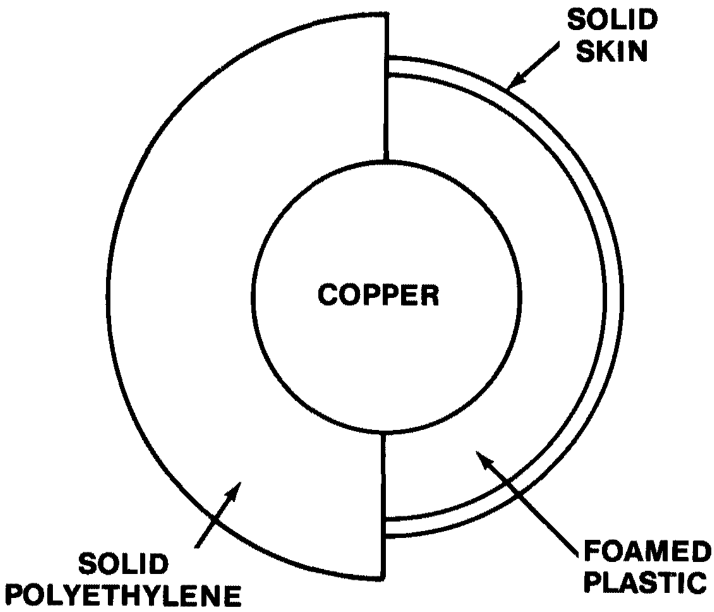


Fig. 16-8. *Right,* dual expanded polyethylene-insulated conductor (DEPIC) and, *left,* a conventionally insulated wire. For the same pair capacitance, the expanded foam plastic insulated wire has a 16-percent smaller diameter.

burying of exchange cable was to succeed, all components below ground—cable, splice closures, wire connectors, and terminals—had to be protected from water. The development of the several components of a completely waterproof cable system began in 1966. The first objective was to fill the air spaces between the plastic-insulated conductors with a water-resistant compound. After considerable experimentation, a mixture of 85-percent petroleum jelly and 15-percent polyethylene was selected as the filling compound (and immediately dubbed *Icky-PIC* in its early messy versions). Later, polyethylene was replaced by polypropylene for the conductor insulation, when the latter was found to be more compatible with the waterproofing compound. The new cable was introduced in selected small sizes in 1970 and was available in all sizes and gauges by 1974.

3.2 Wire Connectors and Cable Splicing

The simplest and time-honored way to connect telephone wires was to strip the ends of insulation, twist the bare wires together, and insulate the joint with a cotton sleeve. By the 1930s, connections for toll cables were twisted and soldered; but, except for this refinement, the simple method was in use until after World War II. While material costs for this connection were negligible, labor costs were high, and poor transmission could develop across twisted joints.

The quest for a better connector accelerated when aluminum cable, which required a tighter and more reliable wire joint, began to appear in the 1950s. Connectors had to be reliable and inexpensive, requiring no wire stripping and no additional insulating sleeve. To meet these needs, the B wire connector was introduced in 1961 [Fig. 16-9].[4,5] It consisted of a thin, springy, phosphor-bronze liner with sharp tangs designed to penetrate the insulation and a soft brass outer shell encased in a plastic jacket to maintain contact between the springy liner and the wire. After insertion of two wires, the connector was pressed together with a pneumatic or hand tool to produce a stable, low-resistance joint. Extensive testing under accelerated aging conditions projected a service life of 40 years. This connector served the Bell System well for many years; several hundred million connections were made each year with a reliability that set the standard for all future connectors.

In underground connections, however, the B wire connector was less satisfactory. It would occasionally produce unreliable joints with the thicker insulation of buried PIC cable, and it was not waterproof. To meet these needs, the 700-series connector family was introduced in 1971.[6] With 700-series connectors, it was possible to splice aluminum as well as copper cables reliably

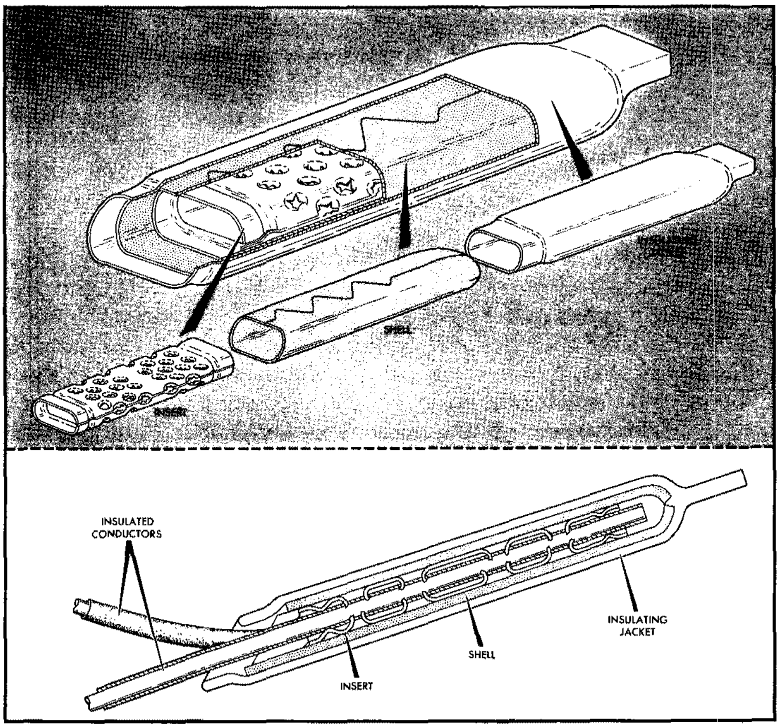


Fig. 16-9. The B wire connector.

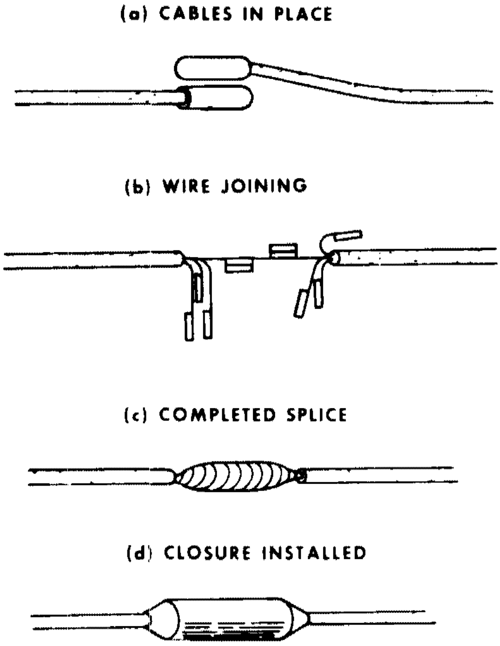


Fig. 16-10. The connectorized exchange cable splicing system.

and economically. The connectors were filled with a waterproofing compound to exclude moisture and contaminants. Designs for butt, half-tap, and bridge-type joints were developed.

In the late 1960s, the quest for greater efficiency, reliability, and cost reduction led to a new approach in cable joining, where 25-pair binder groups were treated as splicing units rather than as so many individual pairs. This approach became particularly attractive after color-coded PIC cables permitted ready identification of each pair in a large cable, a convenience that had not been available in pulp-insulated cables. In 1972, a 25-pair modular-splicing method, the 710 connector system, was introduced. This system met all the requirements of the 700A connector and produced savings in addition. It continued as the predominant system in use into the 1980s.[7,8,9]

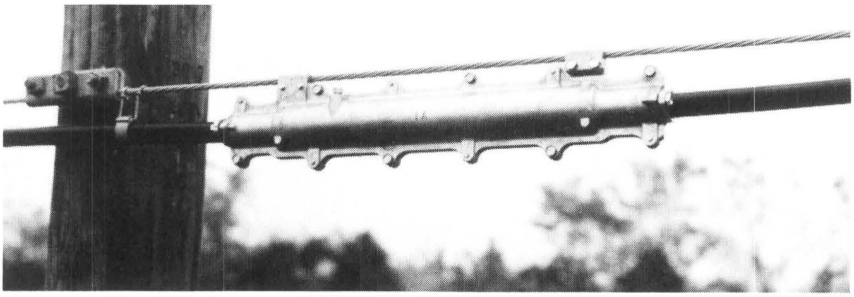
The 710 connector consisted of three parts: an index strip used to order and clamp 25 wire pairs, a connector module that accommodated an additional 25 pairs, and a cap. A specially designed hydraulic tool pressed the three parts together and completed the connection. The heart of the connector was a phosphor-bronze, slotted-beam contact element somewhat like two tuning forks joined at the stem. The contact element cut through the insulation of Nos. 22, 24, and 26 gauge copper wires and made a reliable contact. The connector could be filled with waterproofing compound and, when equipped with indium-plated elements, was suitable for splicing aluminum wires.

The complete system included a test unit to check all 25 connections for faults or shorts. A bridging connector designed for repeated use, which plugged into the connector module in a T configuration, was also available. In a special application of the 710 connector, called the connectorized exchange cable-splicing system (CONECS), telephone cables were equipped with modules at the factory and joined in the field with a simple hand tool, at a substantial saving in time [Fig. 16-10].[10,11,12]

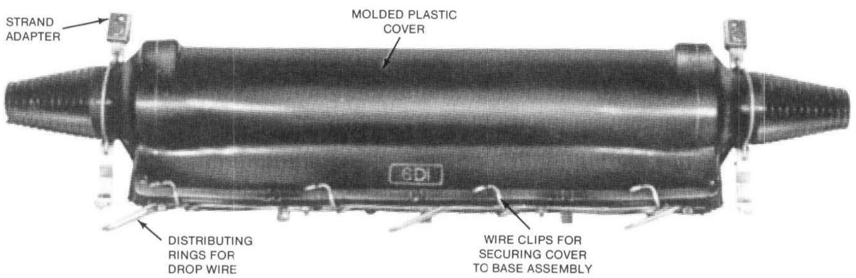
3.3 Splice Closures

After the wires were connected, protecting cable splices was a considerable challenge to developers. Designs had to be suitable for pressurized and unpressurized cables, supported by utility poles (aerial plant) placed underground in ducts and manholes or directly buried in the ground. A common difficult environment was a water-filled manhole in a big city, subject to corrosive stray DC currents from subway electrical systems.

In the pre–World War II years, a lead sleeve was placed over the splice bundle and connected to the cable by solder-wiping to the lead cable sheath to provide a hermetic seal. Later, a variety of splice closures were designed to provide airtight and corrosion-resistant covers. A pressurized mechanical closure



(a)



(b)

Fig. 16-11. Splice cases for aerial cable. (a) Gasketed aluminum closure for pressurized cable. (b) The 6 type plastic closure for unpressurized cable.

for aerial applications was introduced in 1951, consisting of two die-cast aluminum half-shells.[13] The airtight seal was produced by butyl-rubber gaskets that were designed to withstand the punishing temperature variations typical of aerial plant. Later plastic closures for unpressurized aerial PIC cable were relatively simple in design and inexpensive [Fig. 16-11].[14]

Cast-iron closures in place of lead were used for underground plant but were not completely satisfactory because of the heavy weight and corrosion damage under certain conditions. A plastic Type 2 closure system, light in weight and suitable for a variety of pressurized underground cables, was introduced in 1974. Other waterproof, plastic splice closures were developed for underground and buried cable and service wires. The approach was similar to that used for waterproof cables and connectors: all voids were filled with a waterproofing compound after assembly.

3.4 Plant-Placing Methods

Until World War II, loop plant was either suspended from utility poles or placed in conduit and manholes below ground. Development effort centered on finding suitable materials for conduit and simplifying the highly labor-

Fig. 16-12. Plastic conduit, 1968. The 20-foot sections of 4-inch polyvinylchloride ducts are cemented together in waterproof bell and spigot joints.

intensive job of excavating trenches, placing conduit, and pulling cables through the ducts between manholes. Traditionally, each manhole was custom designed and built on site, mostly from reinforced concrete. After World War II, four developments led to substantial savings in installation costs: (1) plastic conduit, (2) standardized prefabricated manholes, (3) self-supporting cable, and (4) the widespread use of direct-buried plant.

Plastic conduit combined low weight, low cost, and the possibility of watertight construction [Fig. 16-12]. Introduced in 1968, it became highly popular and was widely used. Prefabricated manholes were developed in 1969 and gained gradual acceptance because they combined cost savings, high quality, and fast installation, a feature of great importance in busy cities [Fig. 16-13].

Aerial-plant placing was also a labor-intensive process requiring, among other things, the lashing of aerial cable to a steel strand. A self-supporting cable was created in 1963 when a PIC cable was combined with a 0.25-inch high-strength steel strand in a plastic jacket. This type of cable could be placed in a single operation at significant savings.

Directly buried cable plant was introduced in the 1950s. It posed a number of new development problems in cable and connector design to avoid the entry of water. As mentioned earlier, a completely waterproof cable, connector, and

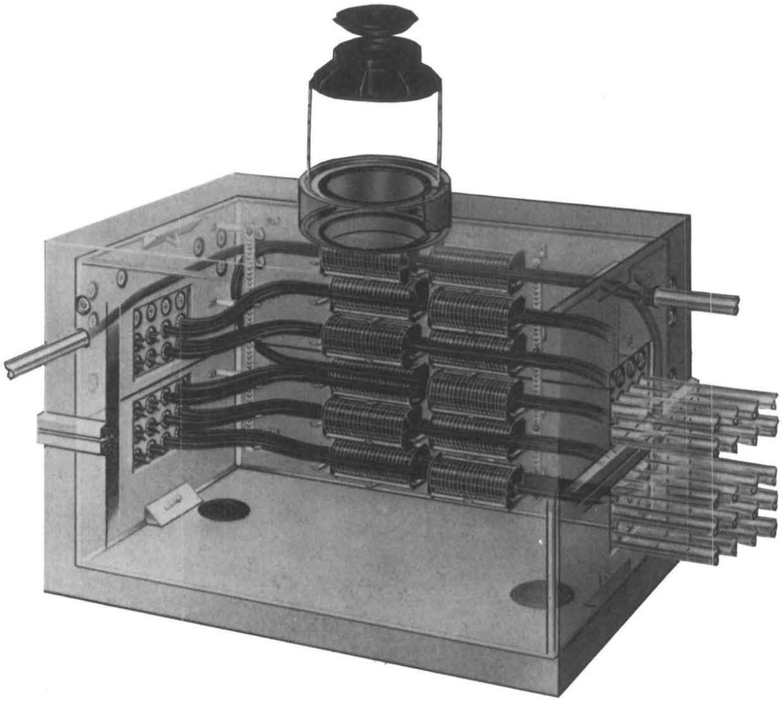


Fig. 16-13. A prefabricated manhole, 1969.

splice-closure system had been perfected by the early 1970s, and a great deal of specialized construction machinery became available. In rural areas, plowing the cable directly into the ground turned out to be fast and cost effective.

IV. INTERFACES AND TERMINALS

Of necessity, loop plant was installed ahead of time, when customer demand was anticipated but not yet firm. Plant engineers relied on forecasts and judgment for sizing cables and locating access points. Over the years, a number of approaches were used to achieve flexibility in providing service when and where it was needed. Convenience in rearrangements and additions was important to compensate for deviations from the forecasts. So long as pressurized pulp cable in lead sheath was used exclusively, access to the plant was not easy, and the provision of extensive extra feeder and distribution cable capacity was necessary to ensure flexibility. PIC changed this situation dramatically. With PIC, it became possible to design a simple, unpressurized cable terminal for aerial cable that gave ready access to every pair in a distribution cable [Fig. 16-14].[15] Introduced in 1956 as the 49 type ready access terminal, this device gave much-needed flexibility to the subscriber network at a time when the system was growing rapidly.

However, the very flexibility of this terminal and the nature of multiple plant proved to be somewhat of a problem when party lines were widely replaced with private lines and universal service came closer to reality. Frequent

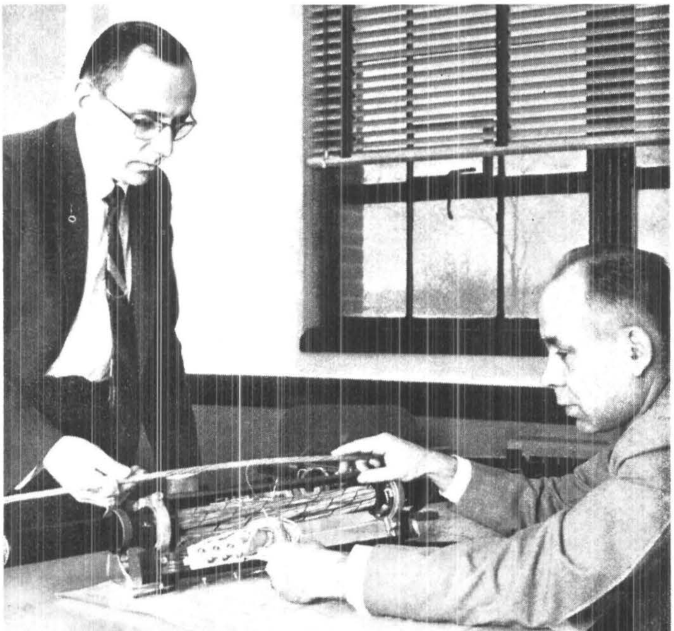


Fig. 16-14. The 49 type ready access terminal, 1956.

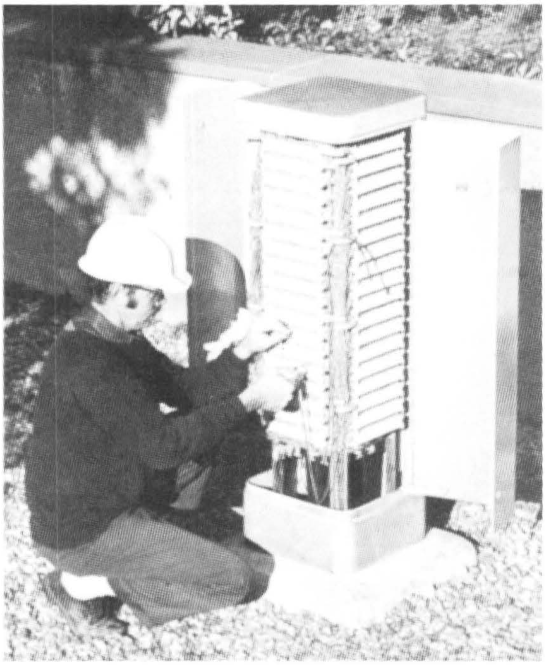


Fig. 16-15. A feeder-distribution interface, 1971. This was widely used as part of the serving-area concept.

entry into and rearrangements within the terminal produced messy tangles of wire. The conviction grew that, in the long run, every residence and business would need its own individual line. Dedicated plant was introduced in 1964 to accomplish this. This concept required new and more substantial interfaces and terminals, called control and access points. Subsequently, dedicated plant was replaced by the more flexible so-called permanent plant, which used an FDI, first introduced in 1971 [Fig. 16-15]. Designed for buried plant, it permitted rapid connection with short jumpers from any feeder pair to any distribution pair. The jumpers were equipped with push-on quick-clip connectors, similar in concept to those used in the 700- and 710-series wire connectors. The first FDI design was later refined and modified to permit housing of loop electronics equipment, and a heavy-duty version used binding posts rather than quick-clip connectors. New distribution terminals for aerial and buried plant abandoned the ready-access concept and offered preconnected binding posts or quick-clip connectors for attachment to drop wires and buried service wires [Fig. 16-16].[16,17]

V. EXTENDING RANGE WITHOUT ADDING COPPER

In the 1950s, growth in rural service, which until then had lagged behind urban growth, began to increase rapidly. This focused attention on the technical and economic problems of the rural plant. A 1964 loop survey showed that only 4 percent of all subscriber loops were longer than 40,000 feet, but these

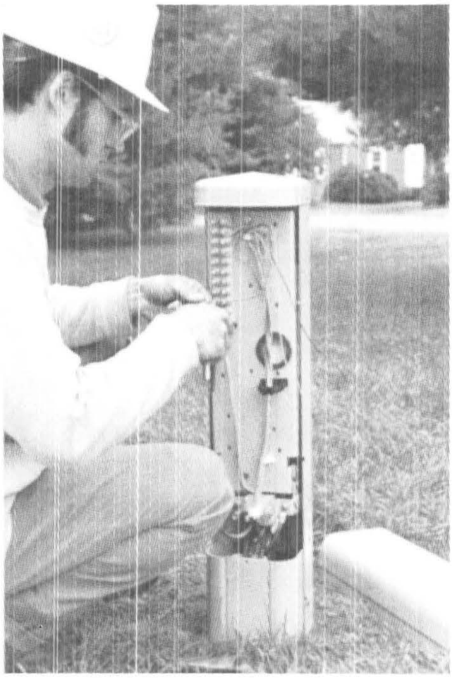


Fig. 16-16. A small distribution terminal for buried plant.

were responsible for 15 percent of total loop outside-plant investment. Loop costs increased, not just in proportion to length but even more rapidly, as heavier gauges and additional equipment were needed. For the longest loops, only expensive open wire met the need. For several years before the 1964 survey and with increased emphasis following it, development effort was directed to solving the problem of high-quality service at lower costs on long loops.

Two major problems were addressed. In most offices, the supervisory range was still limited to loops of 1300 Ω or less. Means were needed so that call origination, dialing, ringing, and termination could be extended to loops of much higher resistance. Transmission quality had to be maintained or improved. Loading was used beyond 18,000 feet, but, for loops above 1600 Ω, an active booster or repeater was needed. A third requirement was implicit. Providing long-loop services should be simple, requiring little if any engineering time.

5.1 Dial-Long-Lines Circuits and Repeaters

The first step was the introduction of dial-long-lines (DLL) plug-in circuitry at the central office [Fig. 16-17]. In DLL, special sensitive relays detected call origination (off-hook) and dial pulse signals and transmitted this information to the switching equipment. At the same time, improved switching equipment

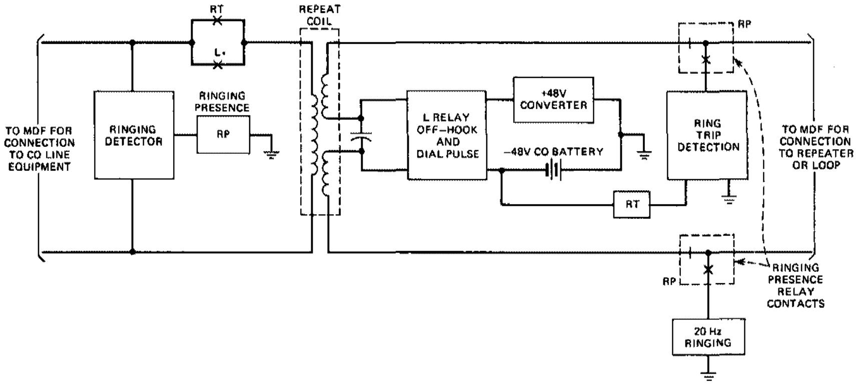


Fig. 16-17. Dial-long lines (DLL) circuit. The DLL provided central office equipment to operate a subscriber loop with a resistance up to 3600 Ω.

permitted operation with increased line resistance, e.g., 1540 Ω for second-generation No. 5 crossbar offices and 1600 Ω for 1ESS*, 2ESS, and 3ESS switching-equipment offices.

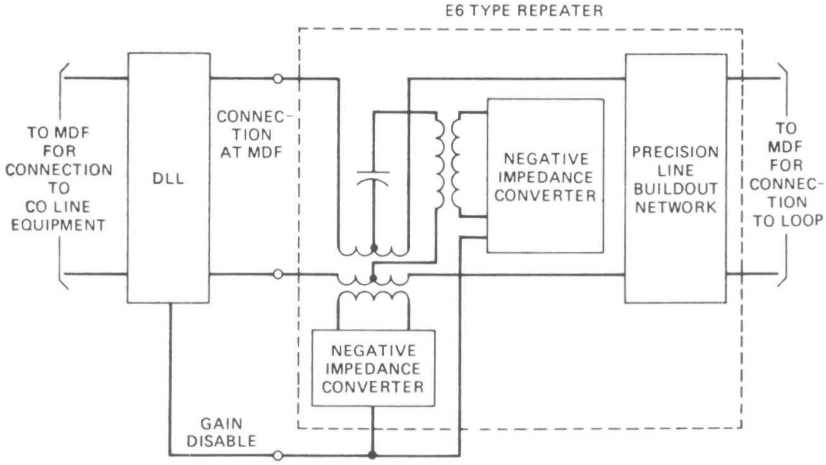
Negative impedance repeaters, designed with vacuum tubes, had been used to a moderate extent for voice amplification on two-wire trunks as early as 1948.[18] However, it was not until an all-solid-state version, the E6 repeater, was introduced in 1959 that their use on long subscriber loops become widespread.[19,20] The E6 repeater was often used in conjunction with the DLL circuit [Fig. 16-18]. The E6 repeater was particularly convenient to use because its negative-impedance elements were inductively coupled to the ring and tip leads, allowing DC supervision and 20-Hz ringing currents to pass through. The E6 repeaters represented substantial progress over earlier vacuum-tube designs but required very careful adjustment of impedance-matching circuits to avoid unwanted oscillations (singing).

With the DLL circuits and E6 repeaters, significant savings were realized for long rural loops, but the administrative difficulties in long-loop design were many. A DLL unit was needed for loops greater than 1300, 1540, or 1600 Ω, depending on the switching machine. Inductive loading was needed for loops longer than 18,000 feet, and voice amplification was needed beyond 1600 Ω (46,000 feet of No. 22 gauge wire). Essentially, each long loop had to be custom designed before service could be provided. Changes in the feeder or distribution plant could require a revamping and redesign of all the long loops involved.

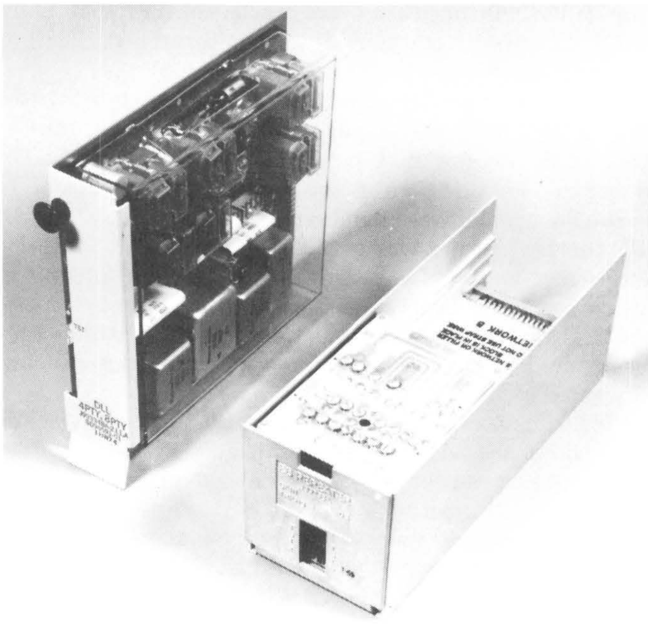
5.2 The Unigauge System

Suburban subscriber-loop networks presented a different problem. Loops in the suburbs were generally shorter than 50,000 feet, but subscriber density was higher than in rural areas. Individual-line treatment with DLL circuits and

* Trademark of AT&T Technologies, Inc.



(a)



(b)

Fig. 16-18. The dial-long lines (DLL) circuit used in conjunction with an E6 type negative impedance repeater for long subscriber loops. (a) The circuit schematic. (b) *Left,* the DLL and, *right,* E6 repeater plug-in units.

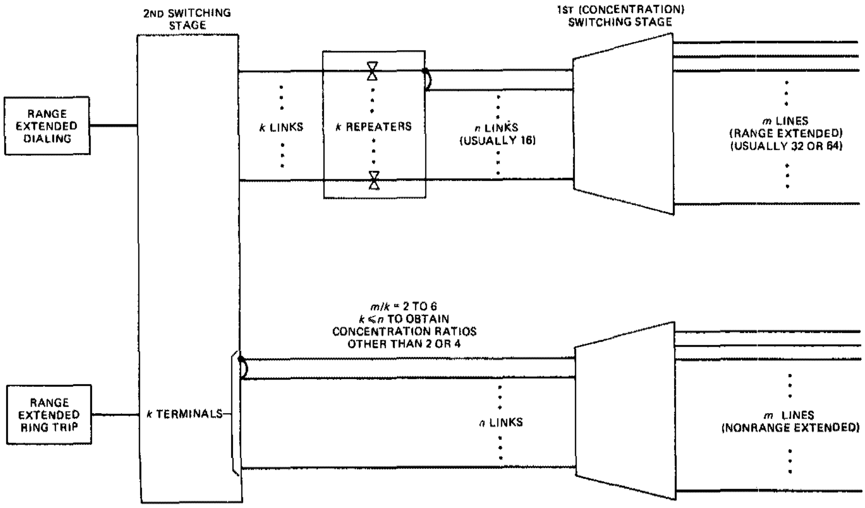


Fig. 16-19. Switched range extension. Voice repeaters were located between the first and second switching stage to take advantage of line concentration.

E6 repeaters tended to be very expensive, so a new approach, called switched range extension, was applied in No. 5 crossbar switching equipment. Long loop capability was provided in the switching system by wiring the equipment behind a stage of switching concentration. In this way, the cost was shared among a number of loops. The use of concentration ratios, typically about three to one, reduced the cost of long-loop service significantly [Fig. 16-19]. An added advantage of this scheme was that it was designed for cables with relatively inexpensive No. 26 gauge wire. The resulting design approach to the subscriber-loop network, introduced in 1966, was called the Unigauge system [Fig. 16-20]. The Unigauge system was used to special advantage in new wire centers where large additions to the cable plant were needed. In existing wire centers with sizable investments in Nos. 24 and 22 gauge cable, nonstandard design approaches were needed to supplement Unigauge.

5.3 The Long-Route Design

Unigauge offered an attractive option for the treatment of loops up to 52,000 feet but did not apply to still longer rural loops. In that range, additional improvements were needed, together with substantial simplifications in route engineering. This was accomplished in 1971 with the introduction of the long-route design system. In this concept, all long loops were divided into five resistance zones and a fixed loop design was prescribed for each zone. Administration was greatly simplified and the custom design of each loop was avoided [Fig. 16-21]. At the heart of the new system was a plug-in unit, called range extender with gain (REG), that combined the features of the DLL circuit and the E6 repeater but greatly simplified the impedance-matching procedure. Over

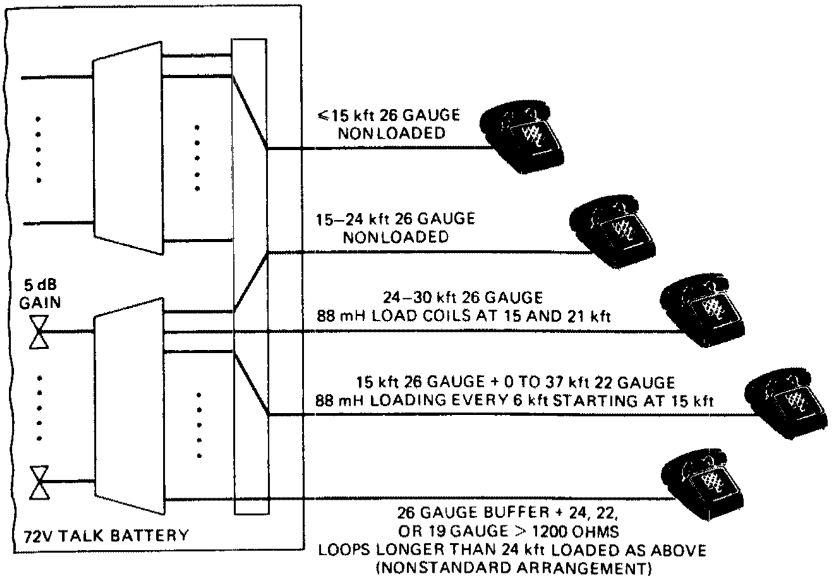


Fig. 16-20. The Unigauge concept, 1966. The design concept was for No. 26 gauge wire in the subscriber loops. Other gauges required a nonstandard No. 26 gauge buffer to avoid singing of the repeaters.

the years, the REG circuitry was improved until the 5A REG was introduced in 1976 [Fig. 16-22]. It featured circuits for automatic gain control and impedance matching that eliminated all field adjustments. The repeater was designed to keep total-loop loss below 8 dB, a value selected after extensive tests of customer reaction to various loop designs. Long-route design and REGs were being widely used by the Bell System companies into the 1980s.[21]

VI. ACCOMMODATING GROWTH WITHOUT ADDING CABLES

Two additional techniques were explored in the postwar years to reduce costs and make lines available on demand without installing a great deal of plant in reserve: line concentration and carrier transmission. Line concentration relocated part of the switching function closer to customers, with fewer lines from the central office to the remote concentrator. Carrier transmission placed multiple voice channels on a single pair of wires. Concentration was almost universal in the trunk plant, and carrier became common on toll trunks. The problem was to adapt these techniques to the peculiar environment of the subscriber loops. The 1A line concentrator, introduced in 1962, concentrated 100 subscribers on 20 talking and 2 control trunks. Unfortunately, it did not

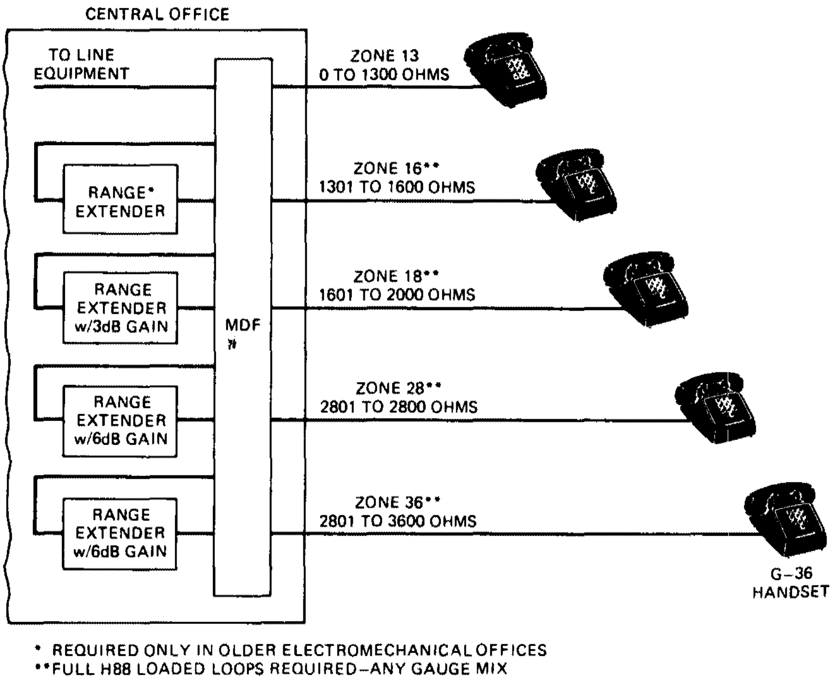


Fig. 16-21. The long route design system, 1971. Long loops were divided into five resistance zones with a fixed loop design for each zone.

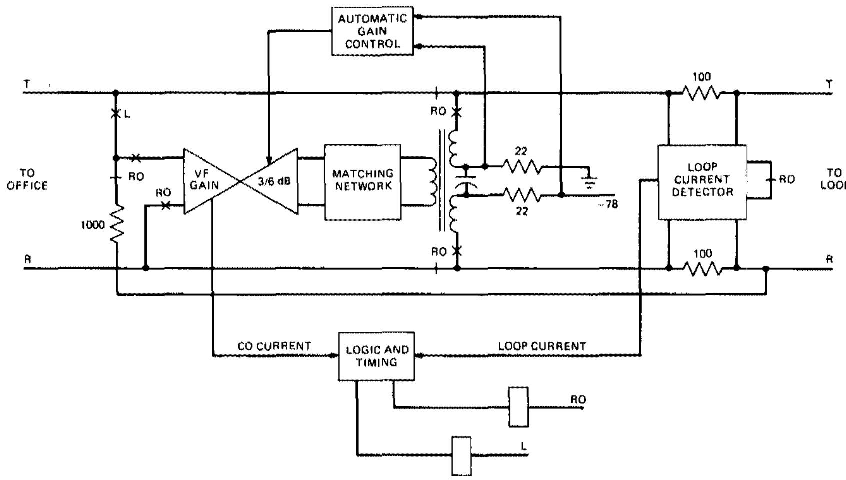


Fig. 16-22. The 5A range extender with gain (REG), 1976. With the 5A REG, loop loss was kept to below 8 dB and no field adjustments were required.

cope well with traffic variations or outdoor locations. The line-concentrator concept proved to be of limited application except as incorporated into loop carrier systems.[22]

6.1 Loop Carrier Systems

Until late in the 1970s, trunk carrier systems rarely proved economical compared to voice-frequency lines for distances of less than 15 to 25 miles. In addition, the early systems used vacuum tubes with limited service life, which required occasional maintenance and replacement. Such systems were obviously ill suited to the loop plant. When, in the late 1950s, improved junction transistors brought increased reliability, low power consumption, and low costs, the situation changed dramatically. Even so, the early solid-state Type P1 carrier system was not successful, even on trunks. It was not until the advent of digital transmission technology that the potential of solid-state electronics could be realized in the special environment of the loop plant.

In 1962, digital technology was introduced into the Bell System on a large scale with the T1 carrier system. The first T1 systems were built with discrete devices, but, in the later 1960s, integrated circuits with more and more devices on a single silicon chip appeared. By the late 1960s, integrated circuits and T1 digital techniques were used to develop the subscriber-loop multiplex (SLM) and the subscriber-loop carrier (SLC) systems.

6.1.1 Digital Subscriber-Loop Carrier Systems

Work on the SLM system started four years before its practical introduction in 1971. The SLM system used the 1.544-megabit T1 carrier repeatered line with 24 digital channels, but, with the help of a built-in line concentrator, it provided access to 80 subscriber lines. It was designed for installation of voice channels as required on an individual basis at up to six remote terminals. The old dream of adding subscriber loops only when and where needed had finally come true—almost. The moderate concentration ratio and circuitry at the remote terminals that gave all 80 customers access to any of the 24 digital lines minimized call blocking. With most of the supervisory and signaling functions in plug-in units serving individual subscriber lines (line packs), the remote-terminal equipment common to all voice channels was greatly simplified. This lowered the first cost of the carrier-system installation.

The complete system was divided into four parts: (1) the line concentrator, (2) the loop-carrier terminal at the central office, (3) the digital line with repeaters (a T1 line), and (4) a remote terminal [Fig. 16-23]. For flexibility, the remote terminal could be subdivided into as many as six units to be installed in separate locations. Each unit could serve as many lines as needed, provided the total did not exceed 80 lines. Each line was served by a line pack that acted as a modulator/demodulator (modem), connected the line to the assigned digital channel, and performed all line-signaling functions. Extensive monitoring, alarm, and trouble-shooting circuitry, and provision of a spare digital line to the central office produced a highly reliable system that could be repaired quickly with plug-in units. The SLM system performed well wherever it was installed and proved to be a sound, reliable addition to the loop plant.

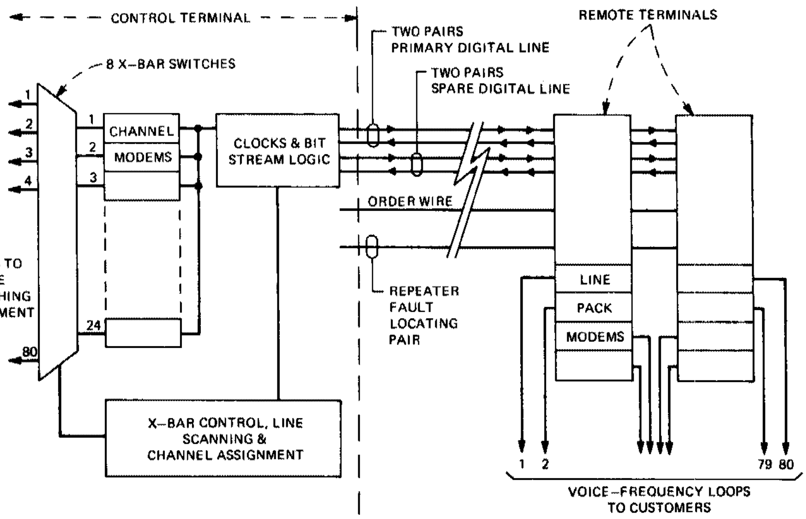


Fig. 16-23. The subscriber loop multiplexer system, 1971.

Rapid advances in digital and integrated-circuit technology subsequently led to the development of a less expensive and simpler SLC system, the SLC-40, which replaced the SLM system in 1974. Designed for 40 full-time voice-frequency channels transmitted over a single digital line with no concentration, SLC-40 had a single, compact remote terminal that could be as far as 50 miles from the central office [Fig. 16-24]. SLC-40 used a delta modulation scheme that permitted transmission of 40 voice channels over a modified T1 digital line originally designed for 24 voice channels. Standard repeaters from the T1 carrier system were used.

The SLC-40 system was a step on the road to widespread loop carrier because it demonstrated reliable digital transmission over long subscriber loops. However, better testing and maintenance features were needed, and there was a growing requirement for a variety of special service features not available in SLC-40. Most importantly, a concept was emerging for a digital subscriber network employing the standard pulse-code-modulation format, compatible with the digital trunk network and with the new digital switches then under development. These considerations led to the development of a new generation of highly successful digital subscriber-loop carriers, the SLC-96 systems, but their realization falls beyond the scope of this history.[23]

6.1.2 Analog Loop Carrier Systems

Despite their enormous development in the trunk plant, analog carrier systems played a relatively minor role in the loop area. However, one analog system, the single-channel SLC-1 carrier system, proved attractive because of its simplicity and low cost. With one carrier channel multiplexed on top of an

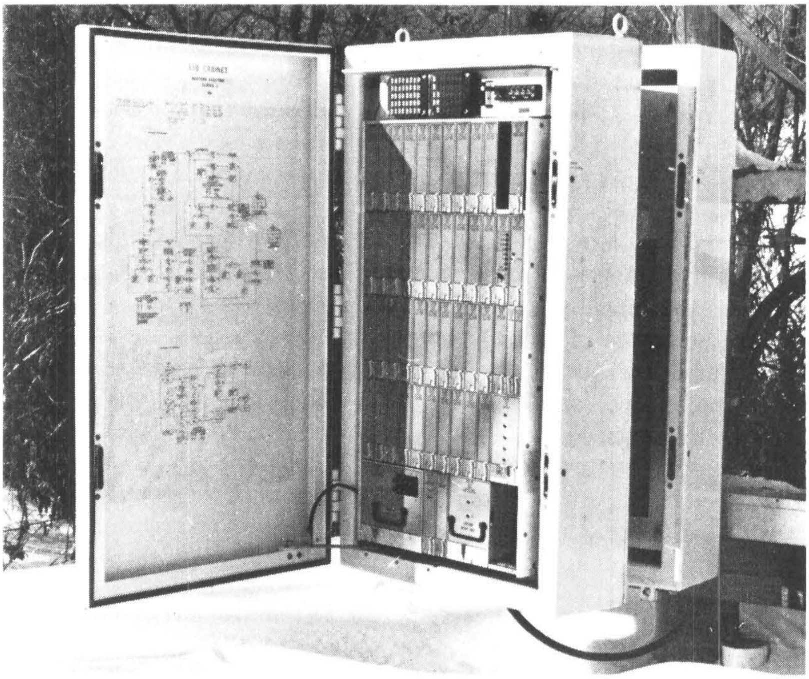


Fig. 16-24. The SLC-40 system remote terminal. The terminal was designed in an aluminum cabinet for pole mounting.

existing voice channel, it was possible to add a subscriber line with a minimum of equipment without adding wires and without disturbing the existing baseband voice channel. Such a system came close to the concept of installing a subscriber line on demand, but it had its limitations. Its range was limited to distances less than 18,000 feet, where subscriber lines were free from load coils, since carrier signals could not propagate on loaded transmission lines. But within its range of application, the single-channel system proved to be economical and useful.

The SLC-1 system was developed and introduced in 1975. Using integrated circuits, SLC-1 achieved a combination of high reliability and low cost that was beyond reach in earlier years. The system consisted of a modem at the central office and one on the customer premises for the add-on line. Two widely separated carrier frequencies were used for voice communications to and from the carrier customer. The customer with the baseband channel was protected from carrier signals with a low-pass filter. Compandor circuitry limited crosstalk on the subscriber line [Fig. 16-25]. Ease of installation and low cost made the SLC-1 system ideal for rapid provision of service on demand. Should digital service be required at a later date, the system could be easily removed and used elsewhere.[24]

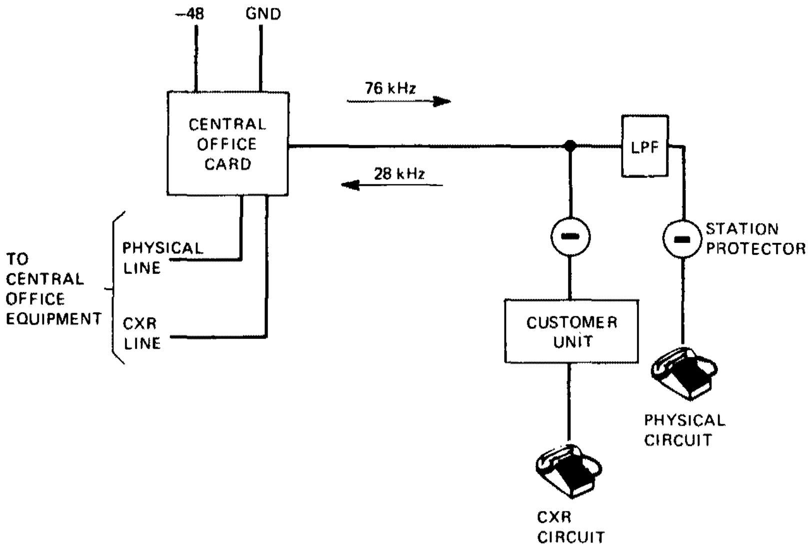


Fig. 16-25. The SLC-1 system, 1975. SLC-1 added one 2-way channel to an existing voice-frequency pair by analog carrier.

VII. THE FUTURE OF THE SUBSCRIBER-LOOP NETWORK

It took more than 50 years to advance from multiparty to mostly single-party service and over 20 years to introduce successful subscriber-loop carrier systems. From the perspective of the mid-1980s, it seems likely to take less than 10 years to reach the next stage of technological development: the general introduction of fiber-optic systems into the loop.

A truly broadband, low-loss, inexpensive, and rugged link between the central office and the customer premises has been a stumbling block in the way of many new services. (It was one of the major obstacles to the achievement of switched person-to-person PICTUREPHONE* visual telephone service in the 1970s). If, as seems likely, optical fibers prove to be an economical solution to basic telephone service in many areas, then the door will be opened to a host of new services, limited not by the costs of special outside plant dedicated to each service but only by the costs of the service provider and terminals. Many new and exciting things, such as high-quality face-to-face video and video conferencing, electronic mail and newspapers, and instant access to vast stores of information, would then be possible. Moreover, these would be available as required at the choice of and at a time convenient to the customer, not with the time and content selected by a broadcaster. Many of these services

* Service mark of AT&T.

are already technically possible with existing systems operating at 1.544 megabits per second. Fiber-optic systems will vastly extend the possibilities.

It is relatively simple to add broadband transmission media such as optical fibers and suitable electronics to the arterial feeder plant. The problem lies in the lesser paths and the capillaries of house-to-house distribution. If this link is short, as in large commercial buildings or high-rise apartments, economical solutions appear to be available. The broadband link between a feeder distribution interface and a suburban residential customer poses an economic problem that is under intense study. If the present trend of decreasing cost for improved fibers and complex electronics continues, the solution should not be far in the future.

REFERENCES

1. P. A. Gresh, "Dedicated Outside Plant," *Bell Lab. Rec.* **43** (January 1965), pp. 26-29.
2. W. J. Lally, "The Changing Pattern of Exchange Outside Plant," *Bell Lab. Rec.* **39** (December 1961), pp. 422-430.
3. See reference 2.
4. H. J. Graff, J. M. Peacock, and J. J. Zalmans, "Development of Solderless Wire Connector for Splicing Multipair Cable," *Bell Syst. Tech. J.* **42** (January 1963), pp. 131-153.
5. S. C. Antas, "The B Wire Connector For Cable Splicing," *Bell Lab. Rec.* **40** (September 1962), pp. 293-296.
6. D. R. Frey, D. T. Smith, and J. P. Pasternak, "700 Series Connectors for Aluminum and Copper Telephone Cable Conductors," *Proceedings of the 21st International Wire and Cable Symposium*, Atlantic City, NJ (December 6, 1972), pp. 130-137.
7. R. W. Henn, "Systematic Encapsulated Cable Splicing Connector for Aluminum and Copper Wires," *Proceedings of the 22nd International Wire and Cable Symposium*, Atlantic City, NJ (December 5, 1973), pp. 195-200.
8. AT&T Product Application Bulletin/Outside Plant, *710 System*, 4th Ed. (August 1980).
9. D. R. Frey and D. C. Borden, "A System of Modular Wire Joining with Connection Verification," *Proceedings of the 2nd International Symposium on Subscriber Loops and Services*, London (May 1976), pp. 68-72.
10. D. R. Frey and G. S. Cobb, "CONECS: A Complete Splicing System for Aerial, Buried and Underground Telephone Plant," *Proceedings of the 27th International Wire and Cable Symposium*, Cherry Hill, NJ (November 14-16, 1978), pp. 287-292.
11. A. W. Carlisle and D. R. Frey, "A New Multifunctional Mass Termination Connector," *IEEE Trans. Compon. Hybrids Manufacturing Technol.* **CHMT-3** (December 1980), pp. 610-617.
12. D. R. Frey and A. G. Hardee, "CONnectorized Exchange Cable Splicing (CONECS)," *Proceedings of the 25th International Wire and Cable Symposium*, Cherry Hill, NJ (November 16-18, 1976), pp. 125-128.
13. W. C. Kleinfelder, "A New Cable Splice Closure," *Bell Lab. Rec.* **32** (November 1954), pp. 405-409.
14. P. P. Koliss, "A New 'Ready-Access' Distribution Terminal," *Bell Lab. Rec.* **36** (November 1958), pp. 411-413.
15. See reference 13.
16. N. Osifchin, "Connective Systems: Evolving New Strategies," *Bell Lab. Rec.* **50** (February 1972), pp. 45-51.
17. D. P. Dobbin, "Distribution Terminals Cut Costs in the Loop Plant," *Bell Lab. Rec.* **54** (October 1976), pp. 249-252.
18. J. L. Merrill, Jr., "A Negative Impedance Repeater," *Trans. Amer. Inst. Elec. Eng., Pt. II—Appl. Ind.* **69** (1950), pp. 1461–1466.

19. H. T. Uthlaut, Jr., "The Application of Negative Impedance Repeaters on Long Rural Telephone Lines," *Trans. Amer. Inst. Elec. Eng., Pt. I— Commun. Electron.* **77** (May 1958), pp. 230-234.
20. G. T. Hawley and K. E. Stiefel, "Voice Frequency Electronics for Loop Applications," *Bell Syst. Tech. J.* **57** (April 1978), pp. 1071-1107.
21. See reference 19.
22. M. E. Krom, "The 1A Line Concentrator," *Bell Lab. Rec.* **40** (September 1962), pp. 297-302.
23. M. T. Manfred, G. A. Nelson, and C. H. Sharpless, "Digital Loop Carrier Systems," *Bell Syst. Tech. J.* **57** (April 1978), pp. 1129-1156.
24. T. N. Rao, "Analog Loop Carrier Systems," *Bell Syst. Tech. J.* **57** (April 1978), pp. 1109-1128.

Chapter 17

Supporting Technology (1950–1975)

I. COMPONENTS

1.1 Negative-Grid Vacuum Tubes

The development of special, low-power vacuum tubes for high-frequency application in the Bell System started in 1934 as part of a research project in

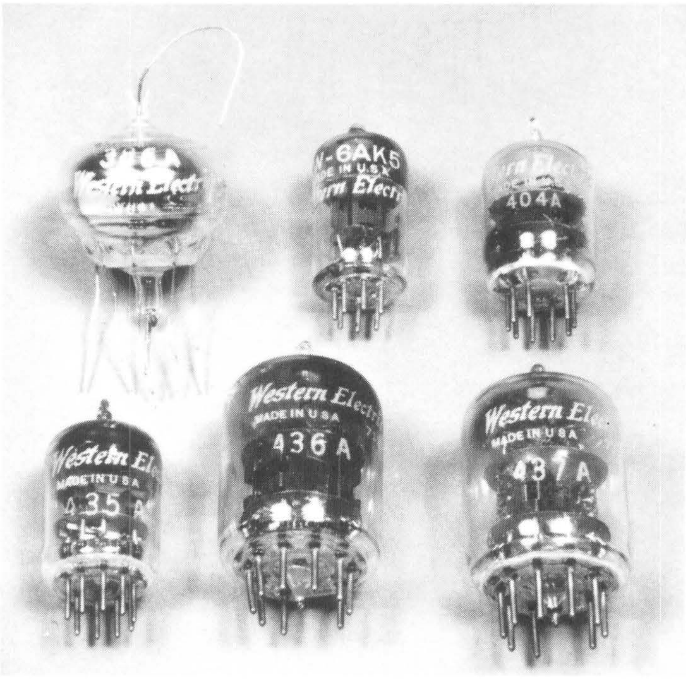
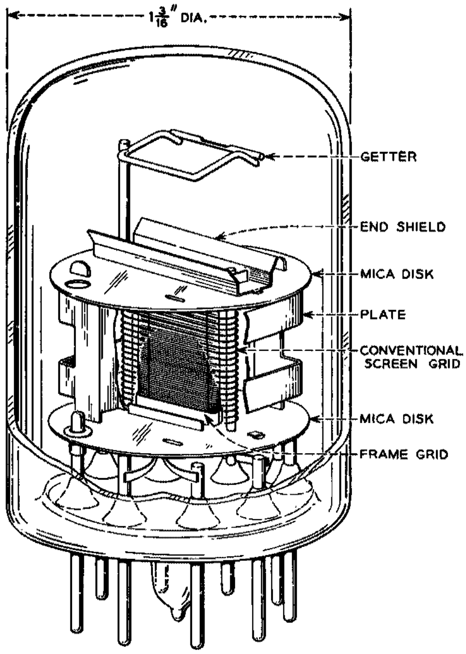
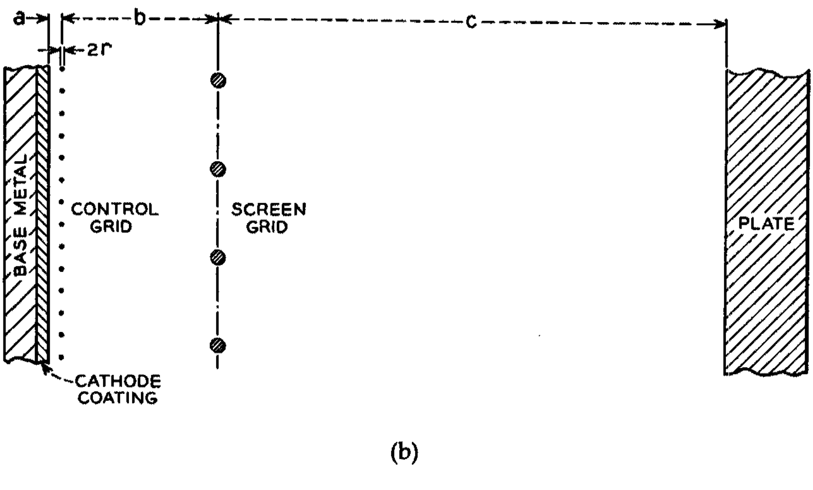


Fig. 17-1. Western Electric high-performance vacuum tubes, 1939 to 1951.

the field of radio communications. When the development of the L1 coaxial system began, it was recognized that similar tubes would be needed but with somewhat different specific requirements, and some of the effort was directed



(a)



(b)

Fig. 17-2. The WE 436A tetrode for the L3 coaxial system. (a) Cutaway view. (b) Electrode spacing shown to scale. The grid-cathode spacing was 0.0025 inch.

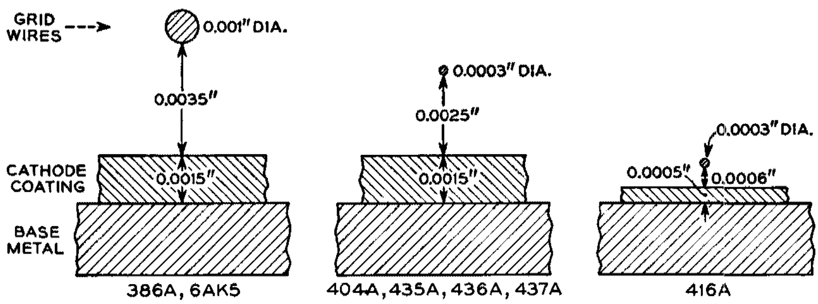


Fig. 17-3. The trend in grid-cathode spacing and grid wire size, 1939 to 1950.

to the special coaxial requirements. By 1939, this work produced the Western Electric (WE) 384A and 386A tubes used in the L1 coaxial amplifiers. In response to wartime needs, and building on experience gained with the L1 tubes, a high-gain tube for use in radar intermediate frequency (IF) amplifiers, the 6AK5, was designed by 1943. (See another volume in this series, *Electronics Technology (1925–1975)*, Chapter 3.) Shortly after World War II, a long-life version of the 6AK5, the WE 408A, was developed in Bell Laboratories for telephone repeater use. By 1949, an improved version, the WE 404A, was used in the IF amplifiers of the New York–Boston TDX microwave system and in the first TD2 systems. Starting at the end of the war, and overlapping the design of the IF tubes, a new family of tubes, the WE 435A, 436A and 437A, was specially designed for the 8-MHz L3 coaxial system [Fig. 17-1].[1]

In all these tubes, leads were made as short as possible and carefully placed to reduce capacitance and inductance. All excess baggage of mounting bases was eliminated for the same reason. In earlier tubes, the typical electrode structure consisted of a self-supporting helical grid between a cylindrical cathode and plate. The postwar tubes featured a planar cathode and parallel-plane grid geometry. Grids were spaced very close to the cathodes and designed to act as a uniform potential plane controlling the current, while offering a minimum of physical obstruction. This objective was approached by winding many turns of very fine wire on a carefully designed grid frame [Fig. 17-2]. Placing the grid closer to the cathode increased the undesirable grid-cathode capacitance, but the transconductance increased more rapidly, resulting in improved performance when designed into a feedback circuit. Designers created both tetrodes (WE 435A and 436A) for high-gain stages and a triode (WE 437A) for the higher-power output stage. The process of close electrode spacing was carried to a pre-solid-state limit in the design of the experimental 1553 triode, in which the spacing between the grid and cathode was only 0.0006 inch. The commercial version of this tube became the WE 416A [Fig. 17-3]. These tubes presented many mechanical as well as electronic problems to the designers. For example, in the WE 416A the tungsten grid wires had to be wound with a residual stress of about one ounce to preserve the geometry when the tube was hot. In the 0.0003-inch-diameter wire, this was equivalent to 200,000 psi, about ten times the normal working stress limit in structural steel.[2]

1.1.1 Microwave Amplifier Tubes

The first microwave amplifiers in the TDX system were designed around klystron-type tubes. These were succeeded in TD2 and, in some later 4-GHz systems, by the highly successful and durable 416A triode. (See Chapter 11, Section I.) The triode amplifiers met the need, but, with the early versions, system developers were constrained by their limited power, narrow band, and the somewhat complex adjustments required in the amplifiers. Looking ahead, the broad band, high gain, and high power promised by traveling-wave tubes (TWTs) looked extremely attractive. During the decade following the war, an intense research effort was devoted to solving the many problems in the way of realizing a commercial TWT.[3,4] Progress was such that a TWT was chosen for the output amplifier when development started in 1955 on the first 6-GHz system, TH1.

The 5-watt (w) 444A TWT used in TH1 was successful, but it required forced-air cooling, was large and heavy because of the need for a magnetic shield, and was relatively expensive. A much simpler structure was developed starting in 1962, for a 5-w TWT amplifier for the second generation 4-GHz system, TD-3. Later TD systems reverted to triode amplifiers when improved triodes yielded increased power—ultimately 5 w—, at much lower cost than a TWT. But when active-satellite work started in 1961, the broad band, freedom from adjustments, and demonstrated long life of traveling-wave tubes left no serious alternative to their choice as the microwave output amplifier in the *Telstar* satellites [Fig. 17-4].[5] TWTs were highly successful in satellites and were used through several generations of design, until they were finally replaced by solid-state devices in the late 1970s.

A TWT, with its high power, broad band, and stability, was also the only serious contender for the output amplifier during the development of the single-

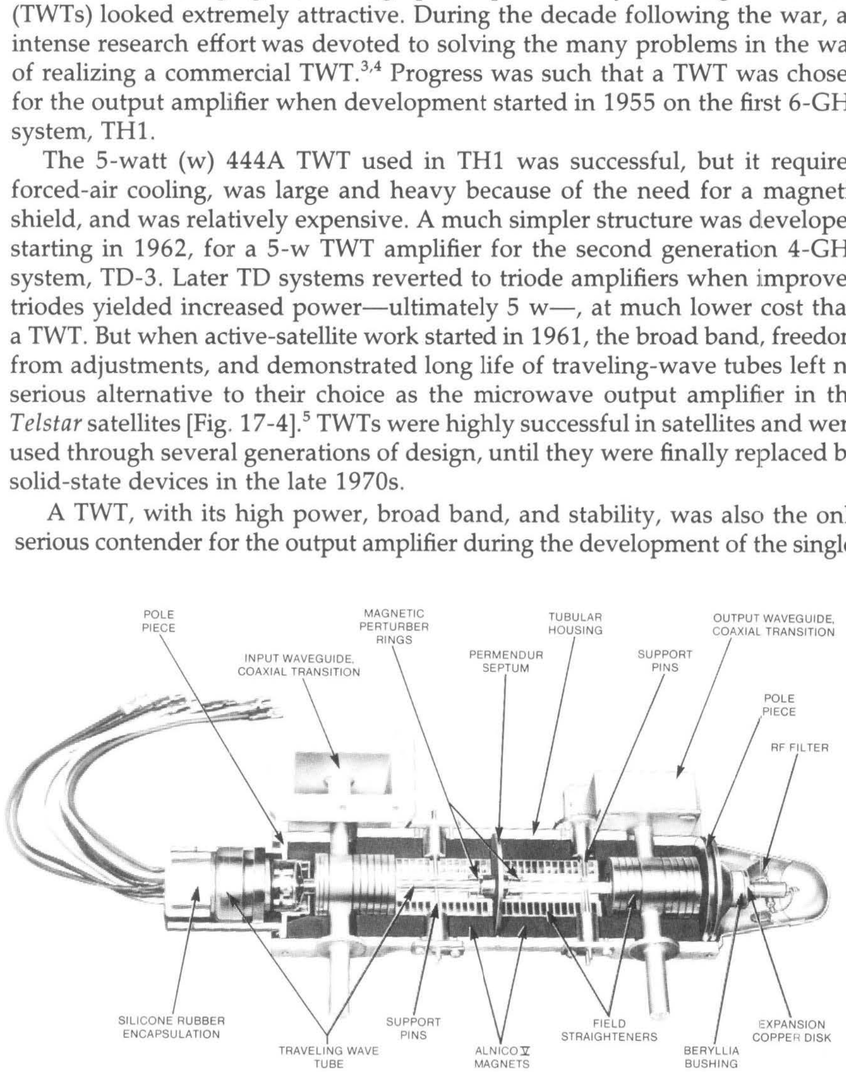


Fig. 17-4. Cross-section view of the M4041 traveling-wave-tube amplifier used in the *Telstar I* satellite.

sideband AM microwave-radio system, AR6A, in the early 1970s. In 1984, it remained one of the last surviving hot-cathode vacuum tubes in production for telephone use.

1.2 Transistors and Integrated Circuits

1.2.1 Discrete Devices

A few solid-state devices were used in transmission long before the invention of the transistor. Galena crystals and other rectifying detectors were used at a very early date. Copper oxide varistors and silicon point-contact diodes were used as modulators and detectors in transmission systems by the late 1930s. With the invention of the transistor in 1947, a wave of excitement went through the transmission development organizations, but it was to be several years before transistors presented a serious alternative to the highly developed vacuum tubes.

Transistors were first used in a transmission system in the rural subscriber-line Type P carrier, developed and tested in 1954 and 1955. Type P carrier was not very successful, as much perhaps because of the system concept as because of the embryonic devices, but, by the late 1950s, highly reliable junction transistors were becoming available at reasonable costs and with valuable properties unattainable in vacuum tubes. The first time assignment speech interpolation (TASI) terminals, developed from 1957 to 1960, were all solid state and would have been completely impractical with vacuum tubes. By 1960, transistorized amplifiers were fitted into the Type N carrier system as the N1A design, and the N2 terminals and line-amplifier that followed, designed around transistors from the outset, provided a substantial improvement in stability, power consumption, and life over the vacuum-tube versions.

The N2 amplifiers used alloy-junction p-n-p germanium transistors in the first stages and diffused-silicon n-p-n transistors in the output stage [Fig. 17-5]. The epitaxial-silicon output transistors permitted higher operating tem-

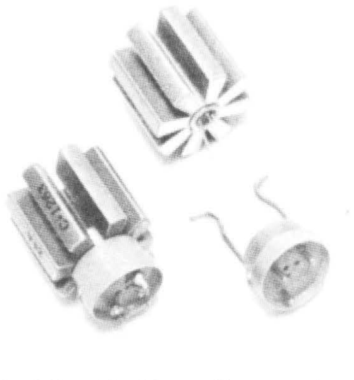


Fig. 17-5. The 24-type silicon transistor used in the N2 line amplifier with its heat dissipating mounting socket.

peratures, allowing highly linear operation over extremely wide ranges of current and voltage swings. Silicon diodes were used in the frequency-frogging modulators at each repeater, providing greater stability with time and temperature than the copper oxide varistors used in the earlier N1 and N1A repeaters.

The compandor compression and expansion functions in N1 (1950 design) were realized by variolosser pads using germanium point-contact diodes. With these devices, the compandor tracking (change in the loss of a circuit as a function of speech volume) could deviate by 1 dB or more. While adequate for manual switching and ordinary voice service, this was unsatisfactory for direct dialing. In N2, special matched silicon-diode pairs were developed. Originally these were diffused silicon mesa diodes, but later, planar silicon devices were developed. With these devices, compandor-tracking deviation was reduced to less than 0.1 dB. The low-current and low-voltage operating biases of the semiconductor devices in N2, coupled with the use of regulated AC/DC converter power supplies, resulted in power consumption approximately one-fourth that of the N1 system.

While the N2 carrier design benefited from the advance in transistors, its needs did not greatly stress the rapidly advancing capabilities of the technology. In contrast, the L4 coaxial system, with twice the bandwidth and twice as many amplifiers per route compared to L3, placed extremely difficult requirements on transistor noise and distortion. Two families of high-frequency, planar silicon

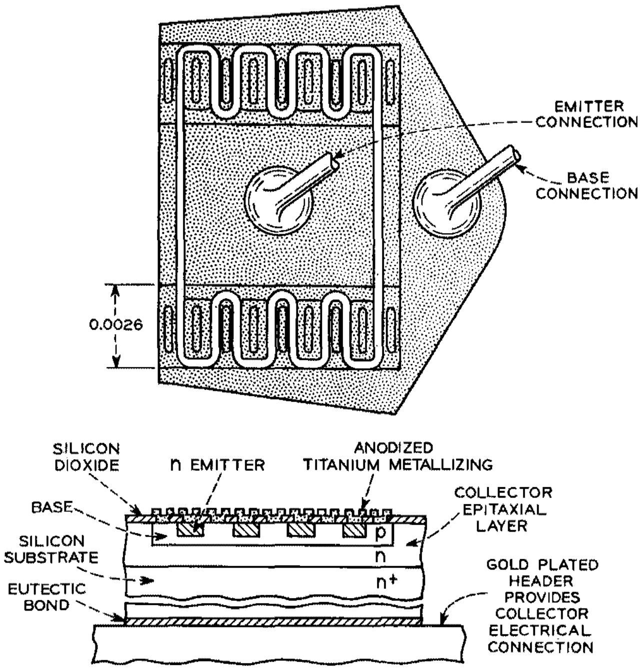


Fig. 17-6. The L4 45D low noise transistor.

transistors were developed to meet these needs. The first family consisted of the 45-type transistors for use in the input stages, featuring low power and low noise, and the second family, the 46-type, was used for higher power and low distortion. The transistor chips were encapsulated in packages designed to minimize parasitic capacitance and lead inductance while providing low thermal resistance to heat sinks. The remarkable ceramic, beryllium oxide, with its unique combination of high electrical resistance and high thermal conductivity, was used in the power transistors.[6]

The L4 input transistor, designated 45D, achieved a noise figure (the ratio of transistor noise to thermal noise) of less than 4 dB by realizing a base resistance under 25 Ω through the use of multiple emitter stripes [Fig. 17-6]. A gain-bandwidth product of 900 to 1300 MHz was achieved by making the width of the base approximately 0.5 μm. (The common-base current gain of a transistor, generally denoted as alpha (α), is a function of frequency. Gain bandwidth is related to the frequency, where α is 0.707 (−3dB) of its low-frequency value. It is used as a figure of merit for comparing high-frequency performance.)

Several thousand L4 transistors were subjected to accelerated aging under conditions of high power and thermal stress. (See *Electronics Technology (1925–1975)*, Chapter 1.) In these tests, the devices were operated at junction temperatures much higher than those expected in service, in some batches until the life was only days or hours. From other batches at lower temperatures the extension in average life was observed, until a projectable relationship of life versus temperature was established. (With sound devices, life usually increased exponentially with lowered temperature.) Extrapolation of the data on the L4 devices to the expected junction temperature of 125 degrees C led to a predicted failure rate of less than 10 FITs (failures in 10^9 device hours). Failures at that rate would correspond to an average device life (mean time between failures) of more than 10,000 years. Later field experience proved to be consistent with the extrapolated results.

By the mid-1960s, continued advances in semiconductor devices made a coaxial system with three times the bandwidth and capacity of the L4 system appear possible. Analysis of requirements and early experiments made it clear, however, that the L5 repeater would place truly extraordinary requirements on the transistors. The gain-bandwidth product had to be extended in proportion to the system bandwidth to 3 GHz. Noise had to be as low as or lower than that in the L4 system. The most challenging requirement was a reduction in distortion by a factor of 100 compared to L4, a consequence of the increased number of channels and halved repeater spacing of the new system.

To meet this need, the device designers produced the 76 and 77-type transistors in six codes, each chosen for a specific purpose, such as low noise or low distortion.[7] The 3-GHz gain-bandwidth cutoff was realized by keeping current densities low and reducing the base width to a nominal 0.27 μm. Noise and distortion objectives were met by using an interdigitated base-emitter structure, with a heavily-doped base grid that minimized resistance to a peripheral metal contact. The noise figure was only 2.4 dB at normal operating current. The structure permitted the fairly large total current necessary for low distortion, while keeping the current densities low enough to meet bandwidth

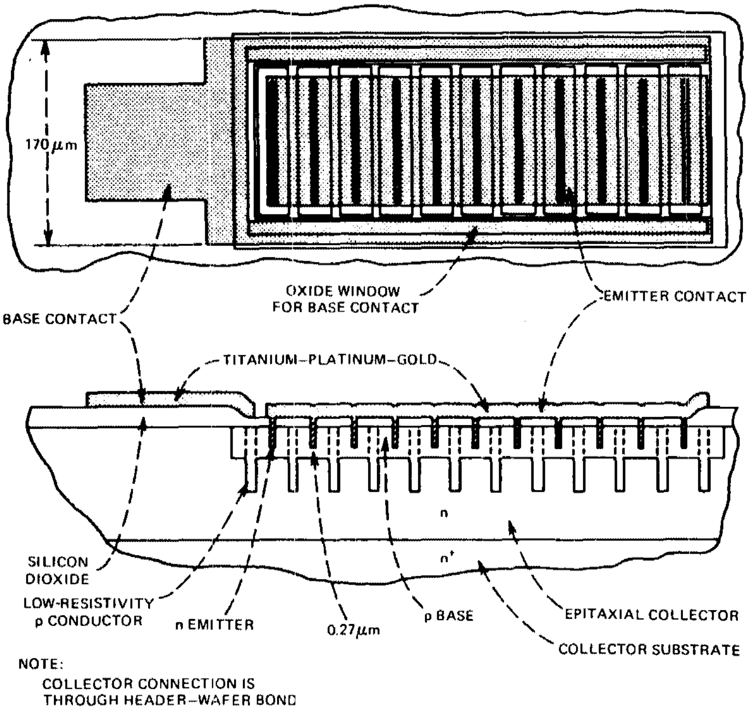


Fig. 17-7. The interdigitated structure L5 transistors.

and noise objectives. A novel method of computer modeling was vital to the low-distortion performance [Fig. 17-7].

Device designers had developed computer models that gave a highly accurate prediction of the characteristics and related them to the way in which the transistor was made, that is, the gradation of dopants that defined the base, collector, and emitter.[8] Circuit designers also had computer models based on external measurements of the devices that predicted the distortion with various source and load impedances and as a function of bias voltage and current. Use of the models made it possible to carry out simulated experiments on a computer, which would have been impossible in the laboratory. Combining the models provided insights not available from either alone and resulted in the design of transistors, starting with the materials and processes, that gave optimum performance in the final circuits.[9]

1.2.2 Hybrid Integrated Circuits

By the late 1960s, the performance potential and small size of the new devices was best realized when they were assembled as hybrid integrated circuits (HICs). HICs were vital in critical circuits, such as the A6 channel bank and L5 repeater. HICs consisted of the solid state devices, tantalum thin-film con-

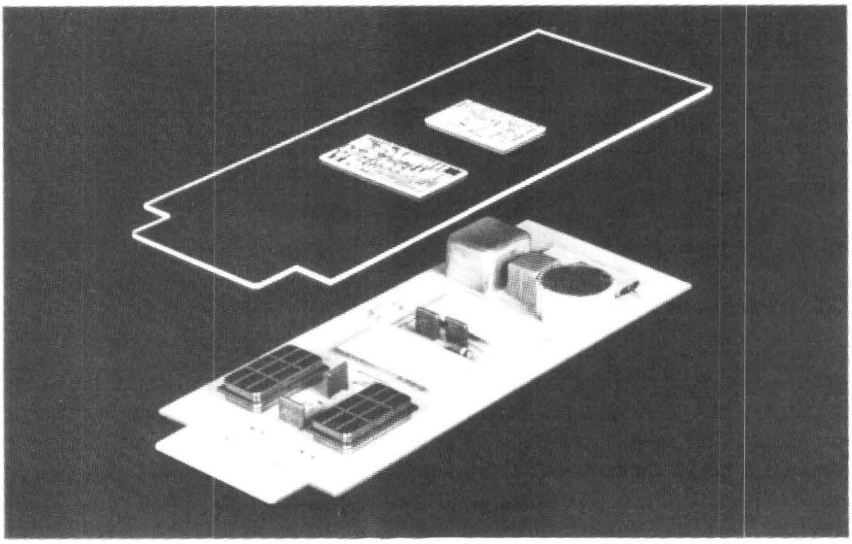


Fig. 17-8. *Bottom,* The A6 modem unit. The raised outline of the circuit board, *top,* shows the two hybrid integrated circuits (turned face up).

ductors, resistors, and some low-value capacitors on an alumina ceramic substrate. The patterns were defined with precision by photolithographic techniques, giving extremely close control of placement and coupling between circuits. Larger capacitors, discrete transistors, and beam-leaded small-scale silicon integrated circuits were appliquéd via precisely located mounting holes or con-

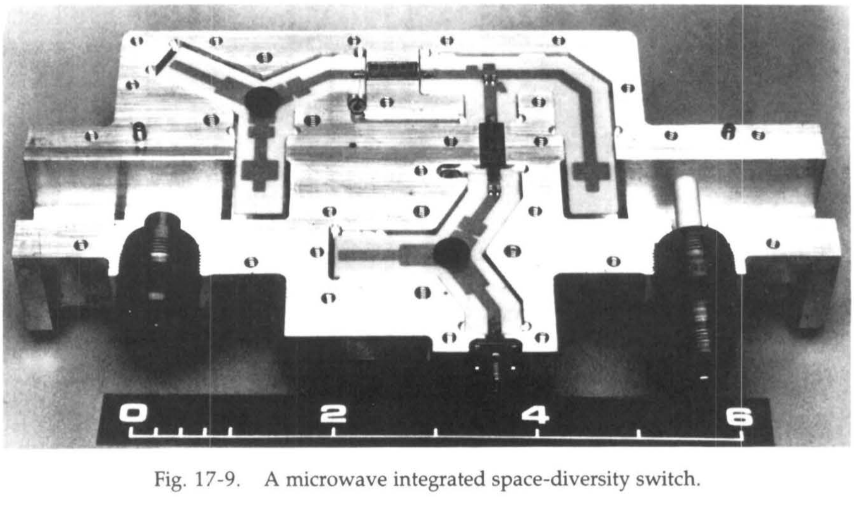


Fig. 17-9. A microwave integrated space-diversity switch.

nector pads. These compact subsystems allowed for simpler and smaller printed-wiring boards as the next level of assembly [Fig. 17-8].[10]

Some of these concepts and techniques were carried into microwave systems, where filters, hybrids, and directional couplers were realized by photolithography in so-called strip-line circuits. These were combined with circulators, modulators, and other waveguide components to form microwave integrated circuits [Fig. 17-9].

II. MEASUREMENTS (1945-1975)

2.1 Transmission-Measuring Sets

In the years from 1945 to 1960, the development of transmission measurements continued along the lines of increased frequency range and improved accuracy. As systems extended to higher frequencies and became more complex, the volume of measurements and amount of time devoted to making them became a major factor in the total development effort. The measuring-set designers responded by providing greatly increased convenience, flexibility, and speed, as well as extended range and better accuracy. By 1949, a direct-reading transmission-measuring set (TMS) was available with an accuracy of 0.05 dB and 0.25 degree for the range up to 3.6 MHz.[11] Within a year or so, the range was extended to 20 MHz and the accuracy improved to 0.02 dB and 0.1 degree to meet the needs of the 8.3-MHz L3 coaxial system. Measurements on these

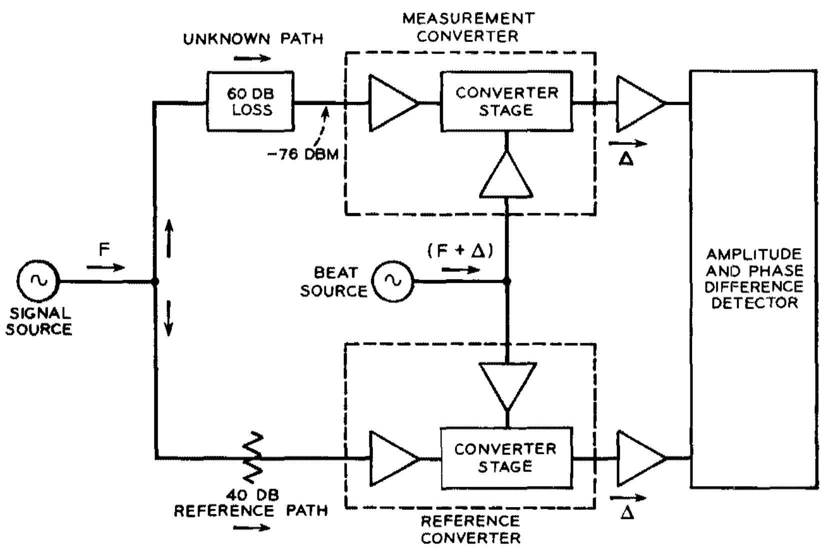


Fig. 17-10. Direct-reading transmission-measuring set at 60-dB loss level. All measurements are at the IF difference frequency Δ.

sets used the so-called IF substitution method, in which the standard for comparison was always at a fixed intermediate frequency [Fig. 17-10].

The early direct-reading TMSs required the measuring circuits to be stable over the period during which measurements were taken, typically one-half hour or more. To avoid drifts associated with warm-up, test sets were often powered 24 hours a day. Even so, changes over the measuring-time interval often set the limit on accuracy. A major step to eliminate this problem was the development of the rapid-comparison technique in 1951.[12] This technique, sometimes referred to as the microbel technique, was one in which the device or circuit under test and the transmission standard were rapidly interchanged, thus eliminating drifts in circuitry common to both positions of the comparison switches. The switches were usually driven by a 60-Hz AC power source so that comparisons between the transmission through the device under test and the standard were made 60 times a second. (For very narrow-band or low-frequency measurements, lower comparison rates were used.) Any difference between the two was readily detected as an AC signal that could be nulled by adjusting the standard until the set was in balance. This highly useful principle was first applied in a transmission test set to measure residual misalignment in sections of the L3 coaxial-cable system [Fig. 17-11]. Accuracy of about 0.01 percent (0.001 dB) was achieved over a frequency range from 0.05 to 8.3 MHz. Further benefit from the rapid-comparison technique was realized in later systems, where the common-path circuitry included frequency translations, or could include a long length of cable with its inherent phase instability.

The basic architecture of new TMSs was essentially unchanged through the 1950s. Development effort was concentrated on extension of the frequency range, reduction of instrument residuals so that measurements would be easier to make, and provision of direct reading without changes of scale or ranges.

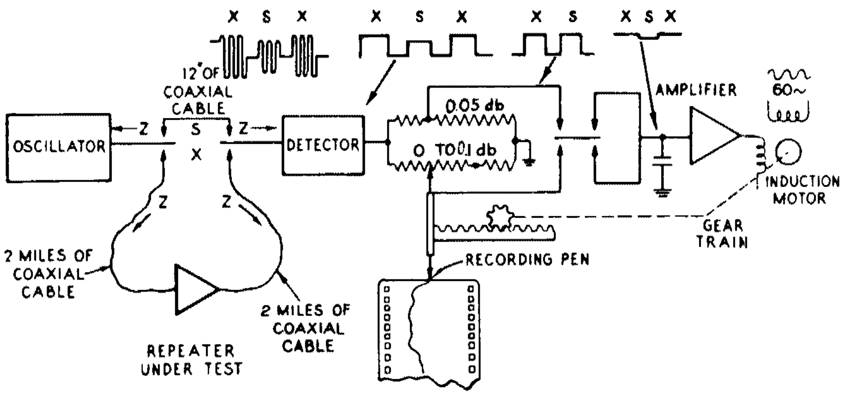


Fig. 17-11. Block diagram of the rapid comparison (microbel) measuring circuit. The switches alternated between the unknown X and the standard (S) 60 times a second. The difference output was adjusted to zero by an automatic balancing mechanism. [Slonczewski, *Elec. Eng.* 73 (1954): 346.]

Transmission measurements were usually made at line impedances of 75 Ω for coaxial cable or a few hundred ohms for pairs. Most two-terminal impedance measurements continued to be made on bridges customized for the impedance and frequency range. However, the improved accuracy and convenience of the transmission sets made it practical to measure impedance over a considerable range of values by measuring the effect of two-terminal devices on through transmission [Fig. 17-12]. By 1960, the features of IF substitution, rapid comparison, convenient measurement of impedance, and even the determination of transistor parameters were combined in a test set with an accuracy of 0.1 dB and 0.5 degree over the frequency range from 5 to 250 MHz [Fig. 17-13].[13]

In another development, the frequency accuracy of the signal source was made to approach that of a crystal standard, and accuracy in phase measurement was improved to 0.01 degree by the use of counting techniques. Phase was determined by the measurement of the time intervals between zero crossings at a fixed intermediate frequency; loss accuracy was improved by a technique for close control of signal levels in the test set.[14]

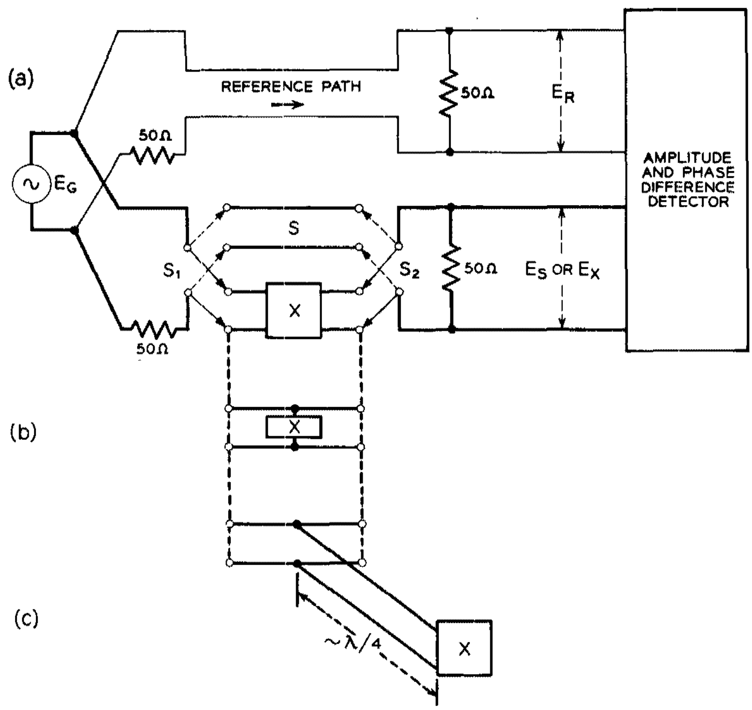
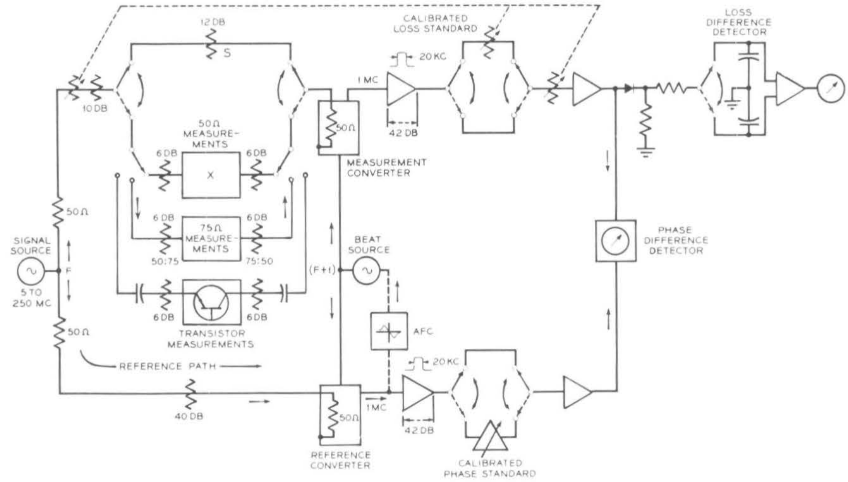
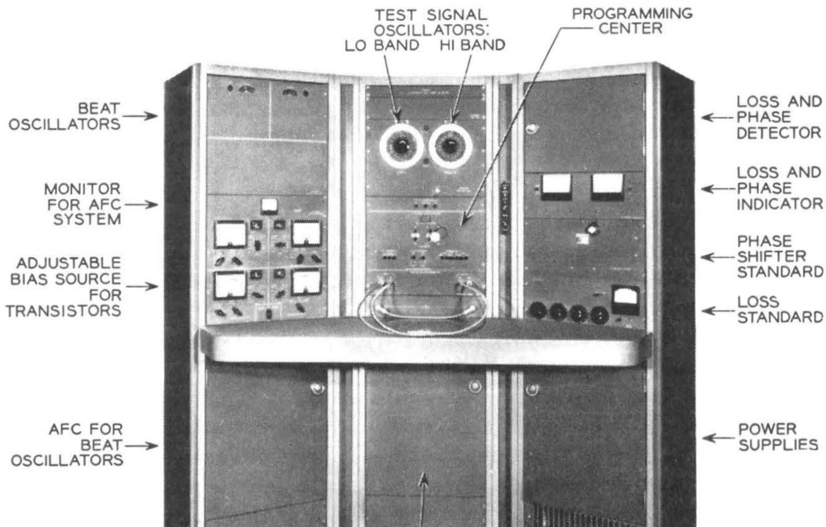


Fig. 17-12. Determining two-terminal impedance by transmission measurements. (a) Rapid comparison TMS set up for transmission measurement. (b) Impedance measurement by bridging. (c) Inversion of high impedance by a quarter-wavelength line transformer.



(a)



(b)

Fig. 17-13. Direct-reading transmission-measuring set (TMS), 1960. (a) The TMS could be set up to measure transmission, two terminal impedances, or transistor parameters. (b) The three-bay 5 to 250 mc phase and transmission measuring set.

2.1.1 *Computer-Operated Transmission-Measuring Set*

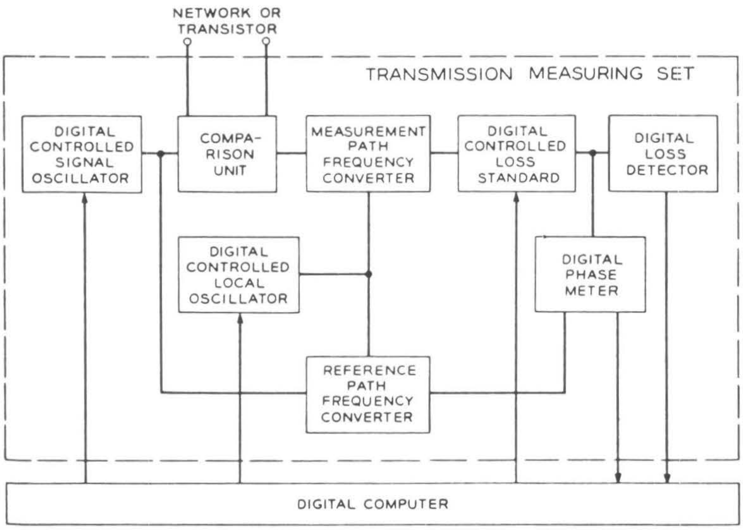
For all this, precise transmission measurements remained a tedious and highly repetitive process that cried for automation. In 1961, a test set using extensive hard-wired logic was developed in which fully automatic transmission measurements from 20 Hz to 20 kHz could be made at a speed 30 times that possible with manual measuring sets.[15] This set demonstrated convincingly the practicality of automated measurements. The next step was to place the operation under the control of a minicomputer of the type that was beginning to appear in the 1960s. This advance was facilitated by the development of components, such as frequency synthesizers for signal sources, that could be controlled by computer-driven circuitry. (The test-set development group at this time was under the direction of S. Doba, who forced the advance of the art with two decrees: only solid-state designs would be pursued, and all new general-purpose measuring sets were to be computer controlled.)

The linking of a computer to a measuring set was undoubtedly the most significant development in transmission measuring capabilities of the 1960s. With a computer-controlled set, it became possible to make large numbers of measurements over extensive ranges of frequency and gain or loss, correct potential instrument errors, speed up operation, and generate data in desired formats. The effects of this approach were profound. (1) Transmission measurements provided both transmission and impedance measurements from a single connection. Two-port characterizations could be converted by the computer from the measured parameters to any other convenient set in a few seconds. (2) The computer could be used to make several measurements under the same conditions and average them. The number of measurements made could be controlled according to the dispersion (derived from the measurements and calculated by the computer) to provide specified confidence limits in the result. The dispersion could be provided along with the mean value. (3) A network or circuit could be evaluated according to a complex figure of merit, such as the root-mean-square misalignment that a coaxial repeater would produce when connected to a nominal cable section.[16]

A computer-operated TMS (COTMS) was developed in the mid-1960s and put into full-time operation in 1967. COTMS measured loss, gain, and phase and calculated delay, impedance, and all two-port parameters over the range from 50 Hz to 250 MHz (later to 1 GHz) with an accuracy of 0.001 dB and 0.01 degree [Fig. 17-14]. It measured at speeds over 300 times faster than the best earlier manual test sets and furnished data in user-desired formats. Even with (or perhaps because of) the increased speed and convenience of the COTMS, use of one set exceeded 700 hours per month on several occasions. Clearly, the new set did not merely expedite measurements that would have been time-consuming with earlier equipment; it made possible measurements and characterization programs that could not have been contemplated before it existed.[17]

Until the appearance of the computer-controlled sets, system designers had accepted, almost as part of the natural order of things, that their ability to measure would lag behind their ability to design or even lag behind what was really needed. With COTMS, perhaps for the first time since the start of electrical communications, the ability to measure was in advance of the need.

(a)



(b)

Fig. 17-14. Computer-operated measurements, 1967. (a) Computer-operated transmission measuring set (TMS). The cabinet at the left contains the measuring circuits. The controlling minicomputer is in the center and plotters are at the right. (b) Computer-operated TMS block diagram.

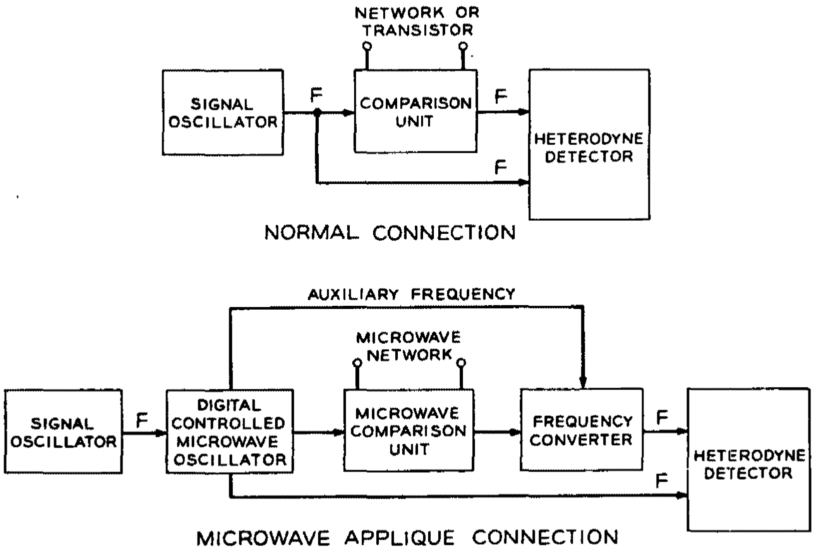


Fig. 17-15. Extension of automated measurements to microwave frequencies.

2.1.2 Microwave Appliqués for COTMS

Appliqués for COTMS were developed to extend the frequency range of measurements into the microwave ranges from 0.9 to 4.2 GHz, 3.7 to 12.5 GHz, and 17.5 to 19.5 GHz [Fig. 17-15].[18] In these ranges, the accuracy was 0.005 dB and 0.04 degree. The rapid-comparison circuit, which made all of the frequency-conversion circuitry common mode, was essential to the successful operation of the microwave adjuncts. Computer control was also applied to transmission measurements in the frequency range from 40 to 110 GHz of the millimeter-waveguide system. (See Chapter 20.) In this frequency range, accuracy requirements were less severe, but offsetting this, the signal sources and detection techniques were less well-adapted to computer control. Despite this, computer operation provided rapid-transmission measurements to accuracies of 0.01 dB.

2.1.3 Swept Measurements—Mini-COTMS

Following the success of the first computer-controlled measuring set, computer control was applied to a number of follow-on designs. The later measuring sets took two main forms: point-by-point sets, which provided the best accuracy but which, at least by the new standards, tended to be slow; and scanning sets, which swept the signal through a frequency range and could measure and display transmission over the complete range rapidly and repetitively but with reduced accuracy. The scanning sets were often used to adjust networks to their best adjustment state, and the point-by-point sets were used to verify that a network met an exacting specification. A computer-operated sweeper, dubbed mini-COTMS, which was introduced in the early 1970s, had the very high speed and display of a scanner, yet could trade speed for accuracy to obtain high precision.[19] This high-speed set operated over the range from 100 Hz to 2000 MHz with an accuracy of 0.01 dB and 0.05 degree. The set could also display transmission and impedance parameters simultaneously in those cases where networks had to be adjusted to meet simultaneous requirements on both parameters.

2.2 Impedance Bridges

The development of impedance measurements at Bell Laboratories proceeded along parallel but separate lines with the development of transmission measurements. Requirements for measurements associated with a new system development often resulted in a joint request for a new transmission measuring set and new impedance measurements, usually by bridges.

Early improvements in impedance measurements were in the areas of calibration methods and in bridge analysis and design. In the years after 1925, over 200 varieties of bridges were developed, each operating over a restricted frequency and impedance range for best accuracy and designed to be direct reading and convenient to use within that range. Many of these bridges were designed to meet special needs. For example, one bridge, important in the vacuum-tube era, used a complex double-Y network of unusual form and was capable of measuring direct capacitance smaller than 0.00001 pf at 465 kHz.[20] Another bridge, which operated up to 10 MHz, used novel conductance standards and a sensitive phase detector to measure the dissipation factor of polyethylene insulation to an accuracy of one part per million. This measurement was of basic importance in submarine-cable structures with their solid-polyethylene dielectric.[21] With the variety and large number of bridges used by Bell Laboratories and Western Electric, it was necessary to have capacitance and inductance standards to check the accuracy of individual bridges. Calibration of standards was done at Bell Laboratories because the National Bureau of Standards did not provide calibrations with sufficient accuracy over most of the range of impedance and frequency.

2.2.1 Computer-Controlled Bridge

As they had in transmission measurements, digital computers provided a major advance in bridge measurements. By coupling a computer to the most advanced equipment structure in the computer-operated Z/Y bridge (COZY), both the operating frequency range (200 Hz to 30 MHz) and the impedance range (10^{-3} to 10^{9} ohms) exceeded those of most previous Bell Laboratories manual bridges taken collectively. The bridge had a single set of terminals for the entire range [Fig. 17-16].[22] In operation COZY would make a quick estimate of the impedance of the device under test by a transmission measurement and then would switch into the bridge configuration most suitable for the measurement. The system had stored corrections for the admittance standards and for any bridge parasitics that affected accuracy. The needed computations were made quickly and accurately by the system computer.

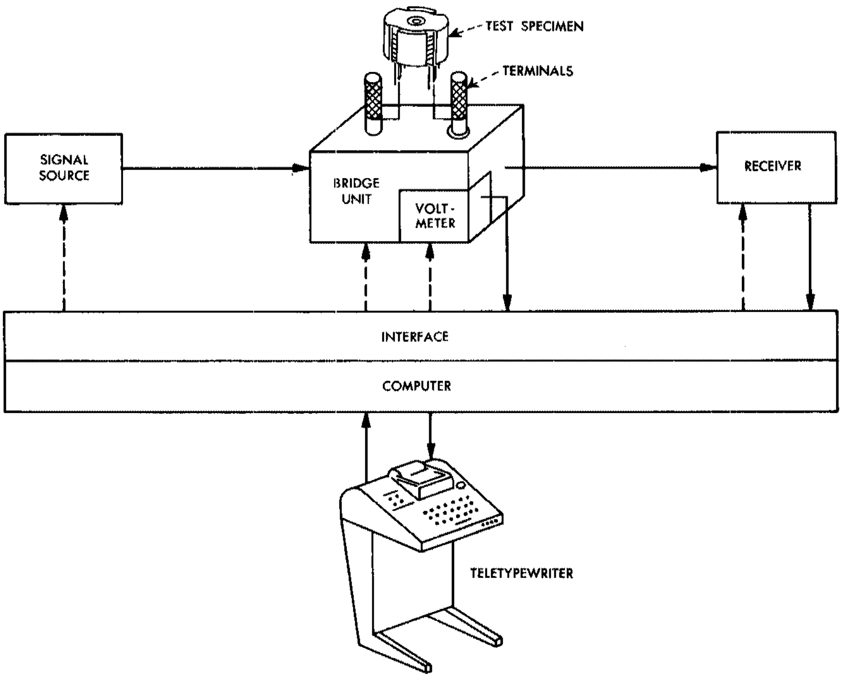


Fig. 17-16. Computer-operated Z/Y bridge (COZY), 1975.

Control of the complex measurement process was practical only in an automatic system. Overall accuracy was about 0.05 percent, equal to or better than the best results available on most manual bridges, and with measurement speeds about 50 times greater.

2.3 General Purpose Measurements—Summary

If transmission measurements are considered from the point of view of accuracy and highest operating frequency, the trend to increased accuracy and increased operating frequencies is apparent. Accuracy penalties were paid when higher frequencies were reached, then improved as techniques were further developed [Fig. 17-17]. The impact of COTMS on wide-frequency coverage along with best achievable accuracy is clear.

The problem of wide-frequency coverage for two-terminal impedance measurements in bridges was more acute, since the bridge standards had to operate at the test frequency. As the need grew for impedance measurements at higher frequencies, special bridge designs were made to provide the needed range with accuracies in the range from 0.1 to 1 percent. The computer-operated impedance/admittance (Z/Y) bridge provided the means for wide-frequency

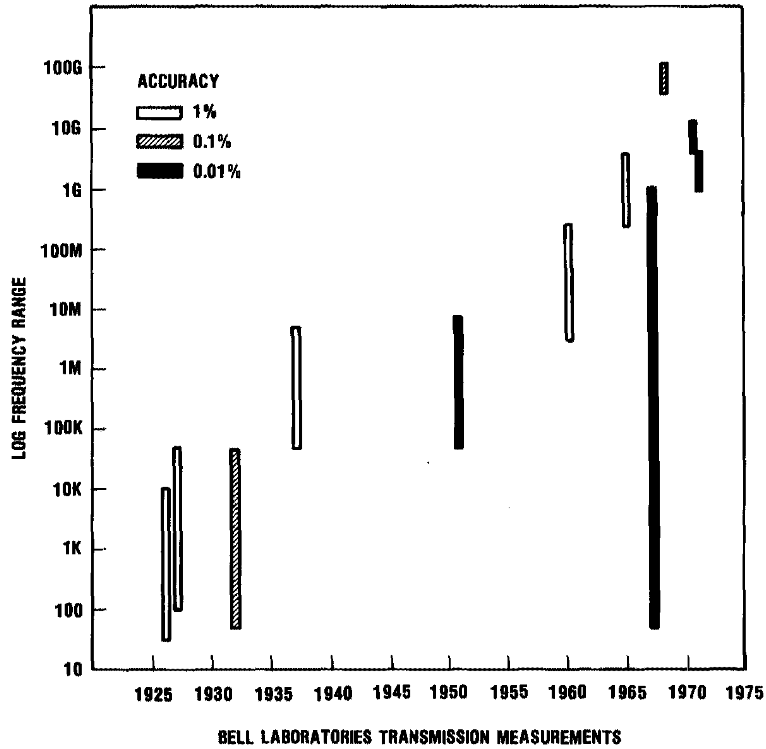


Fig. 17-17. Frequency-accuracy ranges in transmission measurements, 1925–1975.

coverage along with best available accuracy [Fig. 17-18]. Automatic operation was especially needed in the case of bridges to provide the complex control, massive storage of calibration data, automatic checking of calibration standards, and the complex calculations required to transform a group of measurements to the value assigned to the device under test.

When a preproduction version of a new system was tested in the field, measuring equipment was needed to verify that the trial equipment met the design intent. This usually involved shipping to the field measuring equipment with performance as close to laboratory standards as possible. Later, when systems were in regular service, field measuring sets were needed for maintenance. It was necessary, for example, to adjust equipment that drifted with time and to check performance after a repair. The design of suitable field test equipment was an essential part of the development of almost all new transmission systems. A succession of increasingly accurate, convenient, and compact measuring sets were produced. These made use of the latest laboratory measuring techniques in equipment packaged and ruggedized for the field environment.[23,24,25]

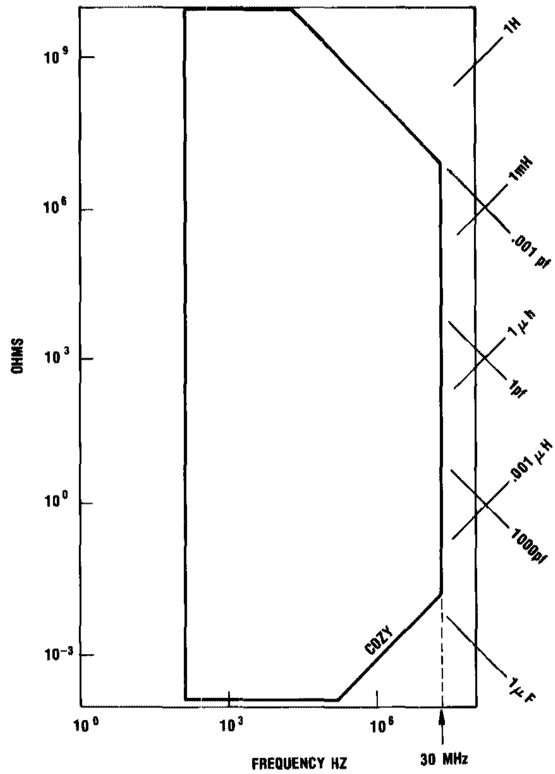


Fig. 17-18. Range of computer-operated Z/Y bridge system.

2.4 Specialized Measurements—Microwave Radio

The measuring equipment described in the preceding sections was of a general nature and was usually applied to a wide range of system developments. When a new system in a new frequency band was introduced, suitable test equipment generally did not exist or, if it did, it was usually inadequate to carry the development very far. Rudimentary microwave measuring equipment had been developed during the prewar research phase, and extended somewhat during the wartime radar work, but when development started on the postwar microwave radio-relay systems, test equipment for laboratory development was inadequate, and equipment suitable for maintaining systems in the field was nonexistent. System developers had to design field-test equipment as well as new laboratory facilities. It was an exceptional, but by no means the only, instance of the need for new systems-oriented test sets.

2.4.1 Power Sources

Any measurement must start with a signal source. For microwave test purposes, klystrons designed during the war became the first signal sources for microwave power in the range of 100 mw. These tubes could be tuned over about a 10-percent band by mechanically altering the cavity dimensions. Fine tuning over a few megahertz was possible by changing the DC repeller voltage. Power could be measured to about ±0.1 dB at 1 mw by heating a thermocouple, thermistor, or bolometer in a bridge circuit.

2.4.2 *Radio-Frequency Sweepers*

Point-by-point measurements on waveguide components using a combination of mechanical and electrical tuning was stupefyingly slow. When a low-frequency voltage was applied to the repeller of a klystron, the frequency could be swept over a range of one percent or so. This, together with a suitable detector and oscilloscope display, made possible the rapid examination of filters, networks, and amplifiers over at least this limited range. Accuracy was not very high, but critical points could be tested more accurately by stopping the sweep and reverting to a single frequency comparison.

By 1953, a sweeper for the 4-GHz band, in which a ceramic vane was rotated by a motor in a tuning cavity, provided a wider-frequency range and more uniform output. This was incorporated in a commercial unit widely deployed throughout the system, a number of which remained in service till 1975 and later. A second radio-frequency (RF) sweeper design used the moving-coil element of a loudspeaker to drive a vane in a microwave cavity. Similar sweepers were developed for the 6-GHz band. The development of ferrite microwave variolossers in the mid-1950s provided an automatic level control that maintained the output constant to ±0.1 dB over the 30-MHz band of a TH system channel at 6 GHz.[26]

In later years, general trade sources designed RF sweepers for extremely wide bands using backward-wave oscillators and, by the mid-1970s, multiband solid-state sweepers were available with electronic tuning. These had excellent voltage-frequency linearity and permitted precise measurements in a short time. Output was held constant within ±0.3 dB over a 10-percent band compared to ±0.5 dB over only a 1-percent band in the first 4-GHz sweepers.

2.4.3 *IF Sweepers*

An early IF sweeper was realized by mixing the output of a fixed-frequency klystron with the swept output of a second klystron. A swept band of about 20 MHz centered at 70 MHz was achieved for the TD2 development by this means. Another IF (70-MHz) sweeper was built using loudspeaker coil action to vary the capacitance controlling a conventional oscillator. For the TH1 system, the IF sweeper was all electronic (no moving parts). The frequency was varied by changing the current, and hence the inductance, in a saturable reactor in an otherwise conventional feedback oscillator. Output was maintained at a constant level by an automatic level-control circuit. This successful mid-1950s design was updated and used in the 92A field test set for TD3 and TH3 in the late 1960s.

2.4.4 *Frequency Measurement*

From the early prewar research work to the mid-1950s, microwave frequencies were measured by adjustable cavities coupled into the waveguide circuits. At its resonance, the cavity produced a dip in the transmission that was visible on a detector or sweeper trace. Originally, the cavity size was controlled by conventional micrometer barrels, requiring a calibration chart to determine the frequency. The early cavities were limited to an accuracy of 1 to 2 MHz in the 4-GHz band. Later cavities were improved in sharpness (higher Q) by better designs and gold plating to reduce dissipation. They were made insensitive to temperature by machining them from invar and were direct reading in frequency. The early 92A field-test sets used the improved cavity meters. By the 1960s, electronic counters, with a very precise clock and time base were adapted for microwave measurements. With these devices, measurements to within a few hertz at multigigahertz frequencies became simple.

2.4.5 Base-band Testing—Noise Loading

Conventional point-by-point or sweeper measurements from 0 to 10 MHz were used to measure the base-band response of radio systems. The system response to a square-wave input provided a quick indication of low-frequency response, important for the video transmission that motivated so much of the early microwave-system work. Multitone intermodulation measurements were made with equipment specifically designed for the purpose, but Bell Laboratories radio-development engineers pioneered in the use of noise loading as a means of simulating a broad band of telephone signals and in measuring the amount of intermodulation resulting.

A broad, flat band of thermal noise was applied to the transmitter input from a high-gain video amplifier with a resistor terminating its input. The band of the applied noise was limited by filters, and one or more band-reject filters eliminated the applied noise from selected portions of the signal spectrum. At the receiving end, band-pass filters eliminated all the received noise except in the cleared frequency slot or slots in which no noise signal had been applied. The received-noise level in the selected slot was therefore due to thermal noise arising in the system at that frequency and to intermodulation from the applied-noise signal in other parts of the band [Fig. 17-19]. By varying the level of the applied noise and thus changing the intermodulation noise, both components could be determined to an accuracy of 1 dB or so. The noise signal was related to an equivalent multichannel telephone load by the analysis developed by

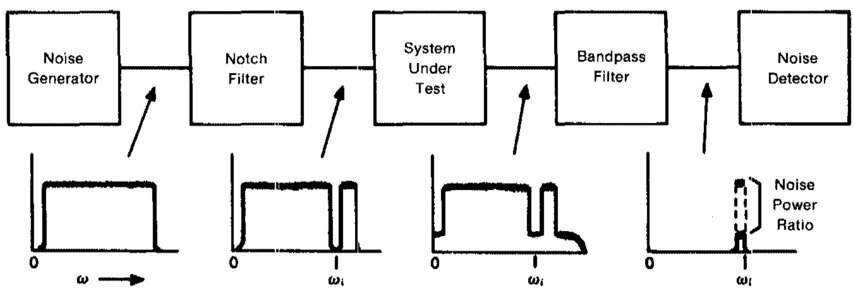


Fig. 17-19. Noise loading test.

W. R. Bennett in the late 1930s. (See Chapter 8.) Noise loading became an industry-wide technique and was applied to cable as well as radio systems. Gas-discharge tubes (fluorescent lamps) mounted in waveguide were also used as microwave-noise sources.

2.4.6 Measurements for Television

In addition to the basic measurements of power, frequency, transmission response, and linearity, a number of measurement techniques were developed to determine how well transmission systems met the special needs of television. As in all picture transmission, the relative delay as well as the amplitude response was important at all signal frequencies. Two-frequency-swept techniques were devised to measure envelope-delay distortion rapidly. The color information in the NTSC color-television signal is conveyed by the amplitude and phase of a 3.58-MHz subcarrier, and these are assumed to be unaffected by the content of the luminance signal. When transmitting color television became important, it was necessary to measure the sensitivity of the subcarrier to total signal amplitude. A number of test sets were developed to measure amplitude-to-phase conversion and differential gain and phase.

Finally, echoes are important in television transmission. In microwave-radio systems, it was possible to discriminate between signals entering the system and those returned from a reflecting point by using waveguide hybrids and directional couplers. By these means, it was possible to construct time-domain reflectometers for echo measurements. These were especially important for the development of waveguide channel-combining networks and antenna feeds in conjunction with long waveguide runs.

III. NETWORK DESIGN (1945–1975)

3.1 Analysis and Synthesis

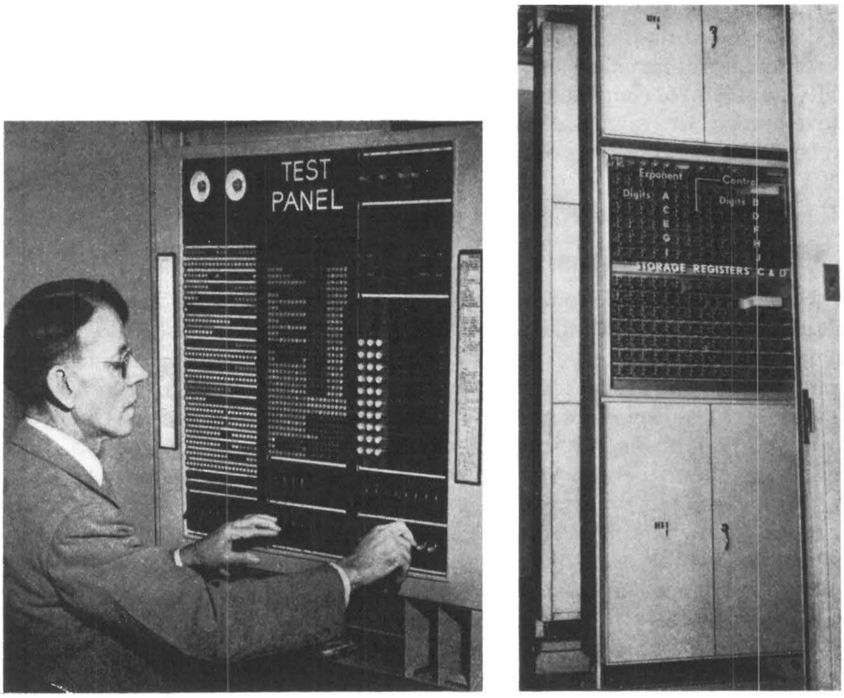
During World War II and in the years immediately following, network design (network in the sense of filters, equalizers, and the like) continued to advance along the lines initiated in the late 1930s. Better measurements and better components contributed to the closer realization of desired characteristics in smaller and less expensive networks. The introduction of ferrite-core inductors in the early 1950s in particular eliminated bulky air-core coils except at the highest frequency. There were refinements in synthesis methods by H. Bode, S. Darlington, and others, especially valuable in the design of networks for feedback amplifiers. Most filters and equalizers continued to be realized by assembling complex networks from simpler building blocks. Designing on an image-impedance basis was still a widely used technique.

3.1.1 Computations

There was probably no field of communications development in which numerical analysis played a larger role than in network design. The preliminary synthesis, based on the cascaded building-block method, was painstakingly analyzed with the use of desk-top mechanical calculators. The analysis of any but the most simple networks often required many hours or even days. More complex methods of synthesis, or attempts to optimize designs by varying components, multiplied the computations required and quickly exhausted the capacity to carry them out. What was finally accepted was often determined more by running out of time for computation than by a conviction that the best possible result had been reached. The more efficient networks and the improved performance promised by Darlington's insertion-loss method required computational capabilities far beyond what was possible with such manual methods.[27]

3.1.2 Application of Computers

Following the early work of G. R. Stibitz, Bell Laboratories engineers designed a number of relay computers for military applications during World War II.[28] After the war, they completed their work in that field by building the relay computer, Mark 6, especially for use in network design [Fig. 17-20].[29] By later standards, the Mark 6, with its relays, punched-paper tape, and only 12 storage registers, was slow, limited, and cumbersome, but it was a great step forward over desk calculators. It had the capacity of dozens of human operators, and



(a) (b)

Fig. 17-20. The relay-based Mark 6 digital computer for network analysis. (a) The control panel for programming operation. (b) The storage register bay.

it was tireless. As it clicked away, day and night, automatically calling up subroutines as needed, it was a harbinger of things to come. It was used from 1949 to 1956 largely on network analysis. It did not change the basic iterative flow of the design process, but it did permit a greatly increased volume of rapid, accurate calculations and thereby widened the horizons of the network designers in searching for more nearly optimum responses to the demands of the system designer.

When the first commercial electronic digital computers became available in the mid-1950s, they were immediately seized upon by the network designers. Of the first two IBM 650 computers acquired by Bell Laboratories, one was dedicated full-time to transmission-network design tasks. The greater speed, data-storage capacity, and versatility permitted programmers to use approaches previously not feasible. The new computers were able to perform some of the iterative procedures automatically, making it possible to apply them to optimization programs. Design approaches like Darlington's insertion-loss synthesis began to be economically feasible. A new day in network synthesis had arrived. The major advances in network design technology now were made in programming and software design, rather than in new mathematical understanding of the properties of reactive networks.

3.1.3 Potential Analogue

Designers still got the ball rolling by drawing on their experience and knowledge of network properties to propose a configuration and the initial values on which the analytical power of the computers could be exercised. Darlington's 1951 paper on the potential analogue was of assistance at the conceptual stage.[30] The potential-analogue method was based on a parallel, first noted by Bode and developed by Darlington, between network performance as a function of the location of the zeros and poles of the analytic function in the complex-frequency plane, and the potentials that would result from similarly located electric charges in a physical plane. This enabled a designer's insight into electric-field properties to be applied to network synthesis and opened a bagful of mathematical tricks useful in network synthesis. Even so, programming of computers moved into the main arena.

3.1.4 Optimization

In 1956, a network-optimization technique was devised in which a computer was used to reduce the difference between a desired and realized (calculated) transmission characteristic by searching for a least-squares minimum of the residues. It was also applicable to the successive steps in realization, from the analysis leading to a paper design, to the realization of a lab model, and to the production and adjustment of a factory product; each step was amenable to computer assistance [Fig. 17-21].[31] About the same time, the new computers were programmed to synthesize a limited class of equiripple so-called Tchebycheff filters following the insertion-loss methods first described by Darlington.[32]

For networks of any complexity, however, even the new computers required long and expensive running time on many of the problems. There was much

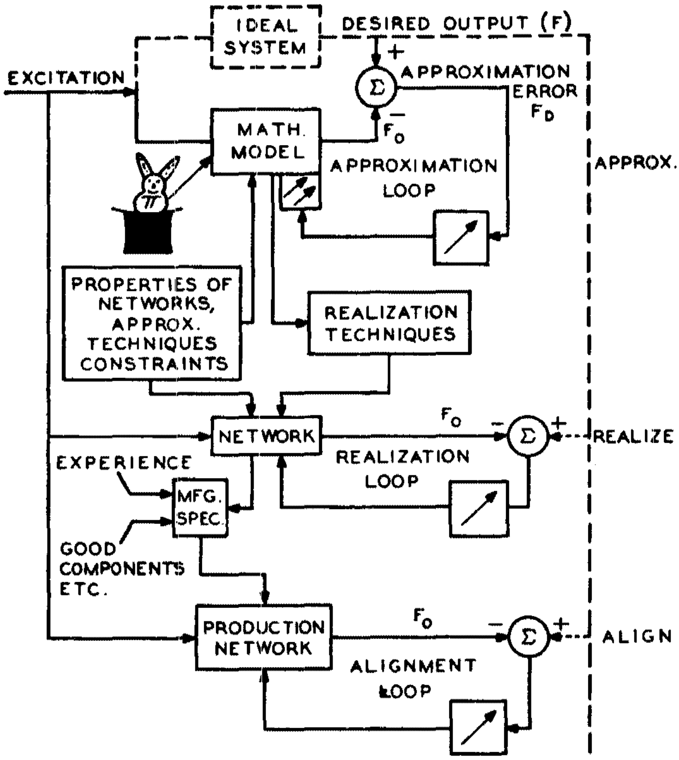


Fig. 17-21. Conceptual flow diagram for least squares optimization in design, lab realization, and production, 1956. [Aaron, *IRE Trans. Circuit Theory* **CT-3** (1956): 225.]

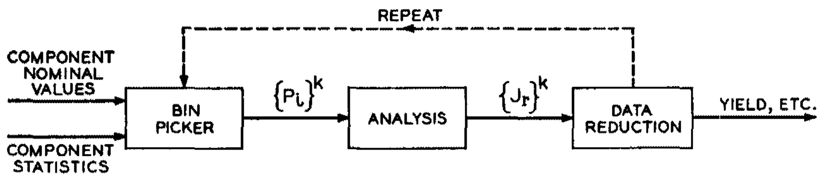
room for ingenious simplifications that allowed the realization of adequate designs with much less computation. In the early 1960s, several synthesis-program packages were devised to arrive at near-optimum results more rapidly, some of which did not even require the designer to start with a proposed topology. In these packages, the computer started with only the basic performance objectives and synthesized a network to meet them by selecting basic network modules whose characteristics were stored within it, essentially following the procedures described by Darlington in 1939.[33,34]

The synthesis programs did not take the designers out of the process. A computer could only arrive at a solution that was best according to criteria laid down by the designer-programmer, and the designer often was not sure what those criteria should be. The question of what was best or adequate was still largely a matter of judgment. By making a number of designs meeting somewhat different criteria readily available, the synthesis programs permitted the network and system designers to collaborate in finding really optimum designs, considering system tradeoffs, and the economics of manufacture, as well as network performance.

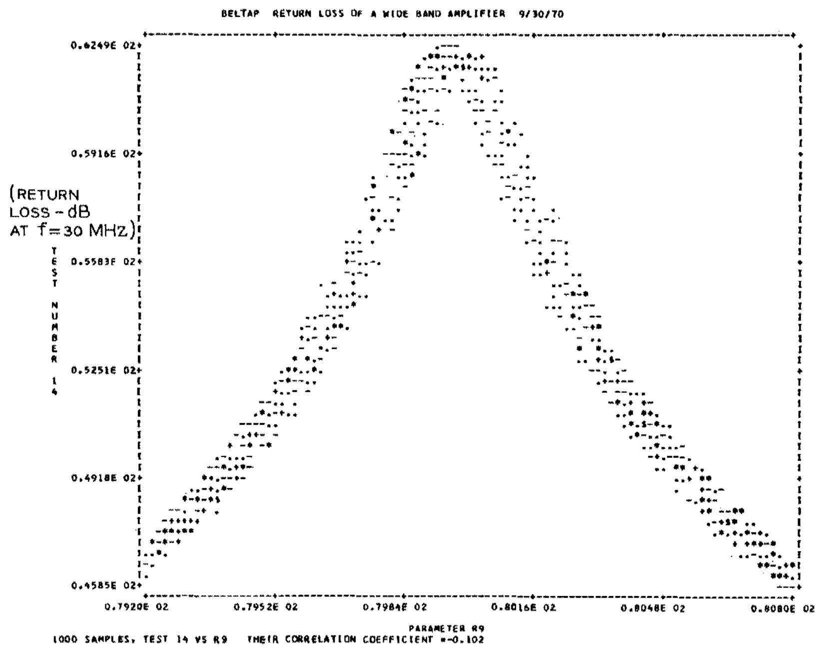
By 1975, computers played a dominant role in every phase of network design. Synthesis, analysis, and optimization were all linked together in the computer. Typically, the designers used interactive graphic displays and control terminals to direct their machines so they could very quickly execute and review many alternative designs.

3.1.5 Tolerance Analysis

Once a design had been determined, computers were used to calculate the expected variation in the manufactured product. Analytic methods could be



(a)



(b)

Fig. 17-22. Tolerance analysis by computer. (a) Monte Carlo tolerance analysis process. (b) Calculated plot of impedance match with a component picked at random from the known range of element values.

used in some cases, but simulation methods permitted the study of more complex designs by using tolerance-analysis programs. In these programs, the computer selected the value of each component at random from a number of "bins" representing the expected range of value for that component, thus simulating the actual process of component selection and assembly in manufacture [Fig. 17-22]. The resulting spread of calculated transmission performance made it possible to specify more soundly based test limits than had previously been possible. The process also could be used in reverse to identify components on which tighter limits might be required to meet a needed overall performance. The use of computer-controlled and -analyzed measurements to characterize the devices as a prelude to tolerance analysis was a closely related activity.[35]

3.2 Crystal Filters and Synthetic Quartz

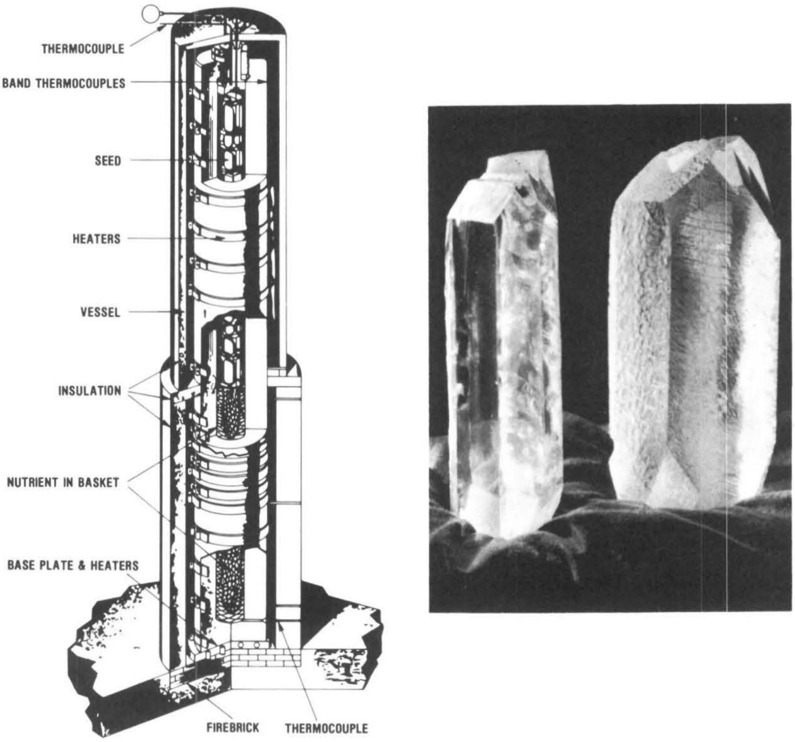
Crystal filters continued to be extremely important components in analog systems. Spurred by the development of the monolithic filter (see Chapter 15) and using the modern synthesis and analysis programs, designers produced a discrete crystal-resonator filter design meeting the group-frequency channel-selection requirements in an extremely small package. The new filter was somewhat larger than the monolithic version, but only 1/100th the size of the original 1938 channel filter.[36]

3.2.1 Synthetic Quartz

The technology of the material for quartz crystals evolved in a dramatic way in the postwar years. Natural quartz, silicon dioxide, is an abundant mineral found practically everywhere on earth, but large natural crystals of electronic grade are relatively rare. The primary known sources were in Brazil. During World War II, when the demand for crystal units grew from a few thousand to millions per year, the dependence on a foreign source became a matter of concern. Following the war, Bell Laboratories devoted a sustained effort to grow quartz of the required quality in the laboratory. From this effort, a high-pressure growing technique, the hydrothermal process, was developed.[37,38]

In the hydrothermal process, small pieces of natural quartz, too small to be of value as electronic quartz, were placed in a pressure vessel in an alkaline solution of sodium hydroxide. Seeds of natural quartz of a size and orientation appropriate to the desired finished crystal were suspended in the upper part of the vessel. The alkaline bath was heated to 750 degrees F at a pressure of 25,000 to 30,000 psi, with the temperature at the seeds about 100 degrees F lower. At these temperatures and pressures, the solution slowly dissolved the small pieces of quartz in the bottom of the vessel and saturated the solution. Through convection, the solution was transported to the top part of the container, where, at the lower temperature, the solution was supersaturated and would deposit the dissolved quartz on the seeds. By this process, large crystals of excellent quality were produced in a growing period of three weeks. Laboratory crystals were being grown by the late 1950s, and in December 1960, mass production of man-made quartz crystals started in the Western Electric works at Merrimack Valley, Massachusetts [Fig. 17-23].

As part of the synthetic-quartz program, an elaborate system of test and



(a) (b)

Fig. 17-23. The production of synthetic quartz crystals. (a) Pilot plant vessel for growing synthetic quartz. (b) Quartz crystals grown at the Western Electric Merrimack Valley works.

analysis was developed. The understanding of the composition and structure of quartz gained from this and earlier research paid off in the mid-1970s. Up until 1975, the Bell System still depended on Brazil for the small pieces of nutrient quartz. In that year, Brazil first placed an embargo and then an export tax on the material that increased its cost by a factor of ten. (This was only a year or so after OPEC's first oil embargo and price escalation. The lessons of pricing a scarce natural commodity were quickly learned.) With the knowledge and test methods gained over 20 years, little difficulty was encountered in locating more than 20 sources of suitable material in the United States and Canada. Since 1975, all nutrient quartz has been from North American sources, with no detectable difference in the finished product.

3.3 Active Filters

Despite their usefulness and advances after the war in the technology of their realization, inductors remained something of a nuisance. This was espe-

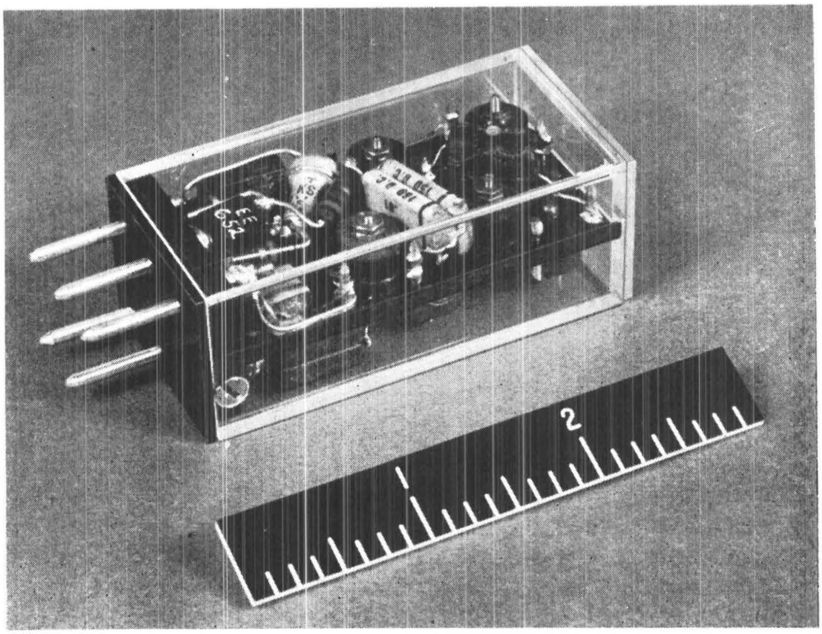


Fig. 17-24. A 98-kHz active channel filter, 1954.

cially true at low frequencies, where they become bulky and lossy. From the work of Bode and others, designers were aware as early as the 1930s that network functions requiring inductors could be synthesized, at least on paper, from resistors, capacitors, and active devices. In 1948, R. L. Dietzold described an active-filter synthesis scheme that placed resistors and capacitors in the feedback path of an amplifier.[39] However, it was not until transistors became plentiful, stable, and inexpensive that active filtering became more than an interesting concept. In 1953, J. T. Bangert published a comprehensive paper showing how transistors could be used in networks to eliminate inductors in filters to produce envelope delay structures with zero loss and to invert the impedance of reactive networks.[40] Models were built to show how transistors could be used to produce a miniature filter even when coupled with small, normally lossy, coils [Fig. 17-24].

3.3.1 *Negative-Impedance Converters*

One of the first synthesis techniques used to realize active networks was the negative-impedance converter (NIC).[41] Network designers were still wary of active devices. NICs were a popular element in early active-filter research, since an impedance could be realized by standard techniques and then its negative realized by interposing a simple one- or two-transistor inverter. Theorists showed how any resistance-inductance-capacitance impedance could be realized with only resistors and capacitors and just one NIC.[42,43] But NIC active filters were inherently sensitive to element variations because they depended on the differences of functions. Small changes in element values could cause large changes in response.

3.3.2 *Operational Amplifiers*

Various techniques were proposed to reduce sensitivity, the most fruitful of which involved the use of an operational amplifier (op amp), an active circuit component that was highly developed in analog computers following World War II. The first op amps were simple active circuits, often a single vacuum tube or transistor, with a great deal of feedback, that could be configured as integrators, summers, and inverters to realize linear transfer functions. In its vacuum-tube form, the op amp was large and expensive, but its descendant, the integrated-circuit op amp, was both small and inexpensive. By the mid-1960s, Bell Laboratories was building active filters in hybrid integrated-circuit form using silicon operational amplifiers and thin-film components.[44] An intensive joint effort in the systems and device areas resulted in a family of devices specially suited to the application. In the late 1960s, designers using these began to fashion elementary active units from which more complex functions could be assembled. The design approach used standard second- or third-order active networks as building blocks to construct higher-order filters. The basic unit could be used in many different active filters, increasing the total production volume and thus reducing unit costs.

The search for a less sensitive, high-production-volume building-block approach to active filters led in 1967 and 1968 to the biquad.[45,46] This unit used up to four operational amplifiers for the realization of a second-order transfer function [Fig. 17-25]. Further developments led to multiple-input and single-amplifier biquads.[47] In one design, the single-amplifier biquad was realized as an array of thin-film capacitors and laser-trimmable resistors, the standard

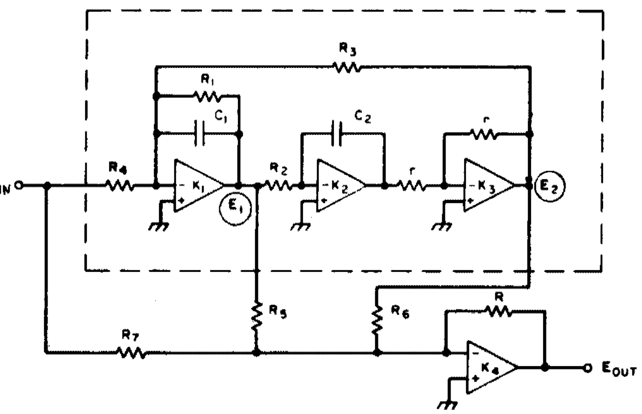


Fig. 17-25. A biquad circuit. [Thomas, *IEEE Trans. Circuit Theory* CY-18 (1971): 358.]

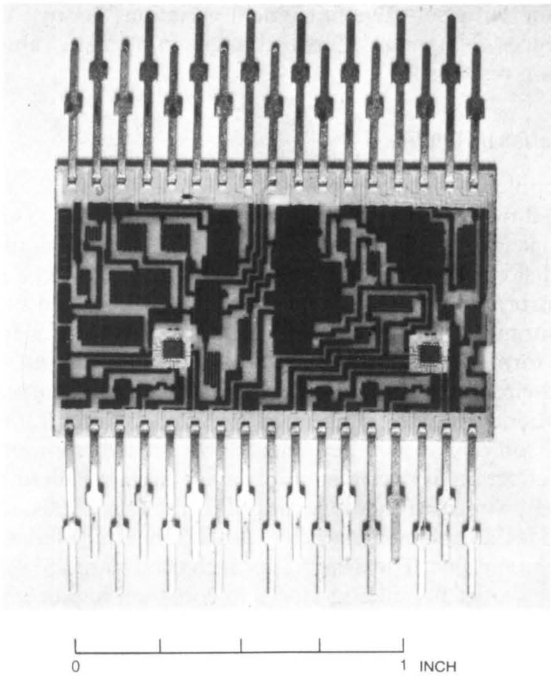


Fig. 17-26. Hybrid integrated circuit realization of the D3 active voice frequency filter, 1972.

tantalum active resonator (STAR). This design was used in multifrequency and Touch-Tone signaling receivers designed in 1973 and 1974.

While the building-block approach was convenient for many applications, for systems in large production a custom design was preferred to optimize the performance and minimize the cost. The voice-frequency transmit and receive filters for the 1972 D3 pulse-code modulation (PCM) channel bank were the first active filters to be manufactured in large quantities (several hundred thousand). They were fifth-order active filters that used just two operational amplifiers. Each filter was realized on one thin-film substrate [Fig. 17-26].[48]

The success of active filters was such that by 1975, most filters for use below 50 kHz were being realized in active form. Besides their widespread use in filters, active-circuit topologies also provided convenient and highly flexible realizations for a variety of other network applications. The ability to adjust an active resistance-capacitance structure by changing only the resistance was especially attractive for applications in equalizers. By the 1970s, active loss and delay equalizers were being used to equalize carrier facilities in private-line data systems.[49]

IV. SYSTEM DESIGN (1945–1975)

By the time the first FM radio systems and second-generation coaxial system (L3) had been designed in the early 1950s, the principles of analog transmission-

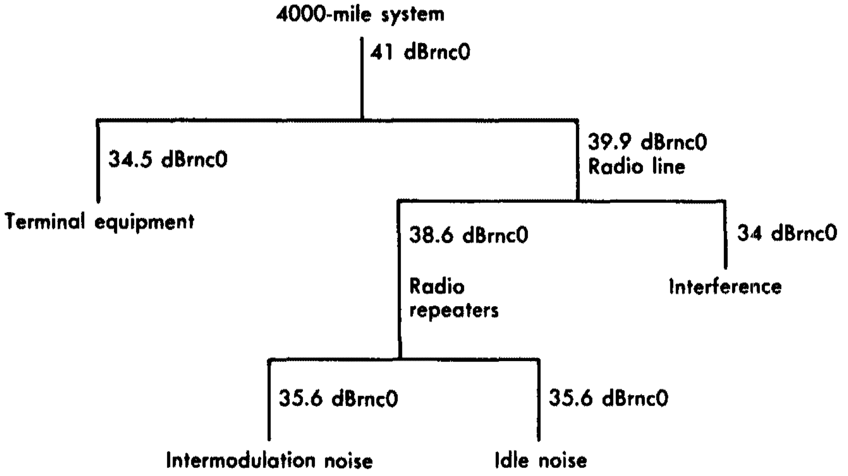


Fig. 17-27. Typical noise budget tree for a microwave radio system, 1970.

system design were pretty well understood and had been codified.[50] Customer-to-customer telephone channel-performance objectives were translated to transmission-system requirements, with appropriate allocation to station sets, local loops, and trunks. For telephony, broadband system designers had little difficulty providing channels of the required bandwidth and transmission uniformity. System design was usually focused on meeting signal-to-noise requirements. The approach was based on the understanding of the sources and accumulation of random noise and the generation and addition of intermodulation products due to nonlinearities. In radio systems, an additional allowance was made for intersystem interference, which often limited the minimum angle between routes, and hence the number of routes that could converge on a city. For each new system, a noise-allocation budget was established in which the contribution that could be accepted from each system component was defined. This budget included all types of impairment. As a development progressed, and components were realized and tested, adjustments and trade-offs were made within the constraints of the overall budget [Fig. 17-27].

4.1 Computation and Optimization

In the early postwar years, the determination of system performance was aided by generalized charts and tables that could be entered according to the specific parameters of the system considered. However, there was much additional calculation by slide rule or by desk calculators. Trial designs had to be verified by laboratory measurement or one or more field tests. As the various parts of a system were necessarily developed in parallel, it was often difficult or even impossible to change performance allocations late in the development cycle. Convergence depended a great deal on the experience and judgment of the designers. Significant margins were built in to cover differences between the anticipated and realized performance of subsystems and to allow for component aging, an important factor in the vacuum-tube era.

As system capacities increased, the system architecture became more complex and requirements tightened. Margins became smaller as confidence in the models improved and as solid-state systems proved to be highly stable. The dependence on calculated performance became more critical at the same time the calculations became more involved. For all these reasons, the earlier manual design process was no longer adequate.

Starting in the late 1950s and early 1960s, the availability of high-speed computers provided system designers with a better approach. Computers readily furnished the solution to complex, previously unmanageable, computational problems and could do so rapidly and repeatedly as parameters were varied. The design process was accelerated by simulating numerous trial designs on a computer without laboratory verification, or by laboratory testing only the most promising options. The accuracy and capacity of the new computers permitted incorporating many refinements and details previously omitted. Finally, computers improved the understanding of complicated systems by the ease and convenience of interaction. A component or subsystem could be varied over the expected range or beyond it, and the effect on the total system could be observed in a short time.

4.1.1 Hybrid Computing

While the first digital computers provided an enormous advance in computational capability, the speed and storage capacity of the available (and affordable) digital computers of the 1950s were not adequate for the direct digital simulation of complete transmission systems or even for major parts of them. At the same time, analog computers, highly developed for some military and aircraft-design applications, were commercially available. To model their problems, transmission-system designers developed a hybrid computation technique—the simultaneous use of both analog and digital computers for system modeling and problem solving. The analog computers were especially well adapted to the solution of nonlinear problems and integrations that would either take too long or were beyond the capacity of the digital computers. An analog computer, operating in a high-speed, repetitive mode, simulated the system or part of it and displayed the response on a cathode-ray tube [Fig. 17-28]. Such displays, however, were often lacking in precision or might not provide useful guidance on favorable directions for change. Data from the analog process could then be fed into a digital computer, with its extensive processing capacity to assist in obtaining more quantitative results and to help in decision making.[51,52] The equations, which defined the dynamic system and the performance-criterion function, were programmed on the analog computer; the parameter adjustments to improve the system performance were determined by the digital computer. Hybrid optimization retained the interaction and real-time display advantages associated with an analog computer and made it possible for the designer to obtain insight into both the simulation and the optimizing process itself.

By the mid-1960s, both the hardware and software of digital computers had advanced to a state that direct digital simulation of transmission systems became both technically and economically feasible. In the early 1970s, simulation on

Fig. 17-28. System simulation by analog-digital hybrid computation, 1964. The analog computer plug-in programming board is at the right.

purely digital computers had largely replaced hybrid computation. Numerous digital computer-aided system-design programs were developed from 1965 to 1975.[53] Most of the programs had optimization and interactive capabilities. Optimization programs permitted the computer to perform some of the iterative procedures, while the interactive capability allowed the designer to evaluate design variations with ease and precision.

4.2 Design Tools for AM Systems

By the mid-1960s computers were used as design aids for every aspect of system equalization and signal-to-noise design. Transmission flatness, while not easily obtained, was usually a tractable problem. Thermal noise and device noise that was similar to thermal noise were well understood and usually well behaved. Much of the effort in enlisting computers focused on the complex, nonlinear, intermodulation-noise problem.

In the early vacuum-tube analog systems, repeater nonlinearity was adequately characterized by a truncated Taylor series, usually the linear, square, and cube terms only. At some point in the system where the band of signals was approximately flat with frequency (usually the grid of the vacuum tube in the repeater output stage), formulas were derived for the number and magnitude of the intermodulation products falling in various parts of the transmission band. The variation of feedback across the band was taken into account, but the vacuum-tube nonlinearities were assumed to be independent of frequency and of the presence or absence of other signals.

The entry of solid-state devices changed the picture significantly. As system bandwidth increased, signals over a wide range of frequencies and levels were impressed on the devices. The input-output relation of a repeater using transistors could no longer be adequately characterized by a "memoryless" nonlinearity, and a better characterization was needed. In 1967, transistor amplifier nonlinearities were characterized by a Volterra series, a generalization of the Taylor series used in the earlier analysis.[54] Volterra series permitted a frequency-dependent characterization of amplifier nonlinearities and yielded a closer approach to the realized performance. The new analysis provided a great deal of insight into the distortion behavior of various cascaded transistor stages and was instrumental in the choice of the L5 coaxial amplifier circuit configuration. Working back, the circuit and amplifier modelers collaborated with the transistor designers, who had developed device models of their own, based on basic material and dimensional considerations. The happy outcome was a device design and circuit performance several decibels better than could have been obtained without the joint computer simulations, and without which L5 performance requirements could not have been met.[55,56]

To support the laboratory and field measurements of intermodulation by noise loading, designers derived analytical formulas for the determination of the system noise-power ratio (NPR) for typical nonlinear characteristics and developed a computer program to evaluate them. This simulation was used extensively in trade-off studies during the development of the AR6A single-sideband AM microwave-radio system, in which a predistorter was used to compensate for the output-stage traveling-wave-tube nonlinearity.

The increasing complexity of analog terminals and the low noise allocated to them in the system-impairment budget (usually about one-tenth of the line allocation) required extensive calculation to evaluate the interactions resulting from the repeated modulation and filtering of the signals. A program called design aid for multiplex equipment (DAME) was developed in the early 1970s to evaluate and display the energy spectra of a multiplex terminal characterized by modulators, demodulators, and filters arranged in any tandem order. With this tool, the calculation time was improved by several orders of magnitude. An interactive capability allowed changes in filter responses, modulator coefficients, or signal levels to be evaluated quickly. Many multiplex terminals, including LMX-3 and the mastergroup and multimastergroup translators, were designed with the extensive use of DAME. Experience showed that the originally assumed filtering requirements could often be relaxed, resulting in lower cost for the terminal equipment.

4.3 Design Tools for Frequency-Modulation Systems

Despite the difficulty in realizing large channel capacity in the first microwave-radio system, the hope persisted that FM radio would be, in some way, inherently a low-noise system. This was fostered, no doubt, by the success of the high-deviation broadcast FM hi-fi systems. As experience was gained, it was realized that thermal noise and noise from nonlinearities were just as important in FM as in AM systems and that they arose from a bewildering array of sources. Signal impairments came from thermal noises, transmission deviations, amplitude-to-phase conversions, echoes, and interference from signals in adjacent channels. As in an AM system design, the designer was faced with the adjustment of system parameters to obtain an optimum balance. The computation of noise was extremely important, not only at the planning stage of a new design, but also for improving existing systems, which was such an important part of the radio development program from 1960 to 1975.

For many years, the computation of intermodulation noise in FM systems eluded exact analysis. Several techniques with various degrees of complexity were advanced during the period from 1950 to 1970.[57,58,59] Most of the methods employed approximations of one kind or another, and the equations used did not apply to all cases of practical interest, but they did provide improvements in understanding and contributed significantly to the substantial system improvements that were realized. During the 1960s, a number of new digital and hybrid computer programs were developed for the computation of intermodulation noise due to FM transmission deviations. By the late 1960s, a digital-computer program was written based on a set of analytically derived formulas using truncated-series approximations. The program was applicable to broad-shape transmission deviations in low-modulation-index systems, and it was extensively used in the system design of TM, TD3 and TH3 radio systems. A hybrid computer simulation of an FM system, devised about 1970, extended the range of application by treating arbitrary pre-emphasis of the system and medium as well as low-index modulation.[60,61] Both digital and hybrid simulations were useful design tools.

REFERENCES

1. G. T. Ford and E. J. Walsh, "The Development of Electron Tubes for a New Coaxial Transmission System," *Bell Syst. Tech. J.* **30** (October 1951), pp. 1103-1125.
2. J. A. Morton and R. M. Ryder, "Design Factors of the Bell Telephone Laboratories 1553 Triode," *Bell Syst. Tech. J.* **29** (October 1950), pp. 496-530.
3. J. R. Pierce, "Traveling-Wave Tubes," *Bell Syst. Tech. J.* **29** (January 1950), pp. 1-59; (April 1950), pp. 189-250; (July 1950), pp. 390-460; and (October 1950), pp. 608-671.
4. J. R. Pierce, *Traveling-Wave Tubes* (Princeton, NJ: D. Van Nostrand Co., 1950).
5. M. G. Bodmer, J. P. Laico, E. G. Olsen, and A. T. Ross, "The Satellite Traveling-Wave Tube," *Bell Syst. Tech. J.* **42** (July 1963), pp. 1703-1748.
6. N. J. Chaplin, G. A. Dodson, and A. M. Jacobs, "L-4 System: Solid State Devices," *Bell Syst. Tech. J.* **48** (April 1969), pp. 983-992.
7. F. A. D'Altroy, R. M. Jacobs, J. M. Nacci, and E. J. Panner, "L5 System: Ultralinear Transistors," *Bell Syst. Tech. J.* **53** (December 1974), pp. 2195-2202.
8. H. C. Poon, "Implications of Transistor Frequency Dependence," *IEEE Trans. Electron Dev.* **ED-21** (January 1974), pp. 110-112.
9. S. Narayanan and D. L. Scharfetter, "Mathematical Models Help Cut Distortion In High-Frequency Amplifiers," *Bell Lab. Rec.* **53** (July/August 1975), pp. 295-299.
10. L. H. Steiff, "Applying Integrated Circuits To Transmission Systems," *Bell Lab. Rec.* **48** (November 1970), pp. 286-291.
11. D. A. Alsberg and D. Leed, "A Precise Direct Reading Transmission Measuring System for Video Frequencies," *Bell Syst. Tech. J.* **28** (April 1949), pp. 221-238.
12. T. Slonczewski, "Precise Measurement of Repeater Transmission," *Elec. Eng.* **73** (April 1954), pp. 346-347.

13. D. Leed and O. Kummer, "A Loss and Phase Set for Measuring Transistor Parameters and Two-Port Networks Between 5 and 250 mc," *Bell Syst. Tech. J.* **40** (May 1961), pp. 841-884.
14. J. S. Elliott, "A High-Precision Direct-Reading Loss and Phase Measuring Set for Carrier Frequencies," *Bell Syst. Tech. J.* **41** (September 1962), pp. 1493-1517.
15. G. D. Haynie and P. E. Rosenfeld, "An Automated 20-20,000-cps Transmission Measuring Set for Laboratory Use," *Bell Syst. Tech. J.* **42** (November 1963), pp. 2501-2531.
16. J. G. Evans, "Measuring Frequency Characteristics of Linear Two-Port Networks Automatically," *Bell Syst. Tech. J.* **48** (May/June 1969), pp. 1313-1338.
17. W. J. Geldart, G. D. Haynie, and R. G. Schleich, "A 50 Hz-250 MHz Computer-Operated Transmission Measuring Set," *Bell Syst. Tech. J.* **48** (May/June 1969), pp. 1339-1381.
18. J. B. Davis, C. F. Hempstead, D. Leed, and R. A. Ray, "3700–4200-MHz Computer-Operated Measurement System for Loss, Phase, Delay, and Reflection," *IEEE Trans. Instrum. Meas.* **IM-21** (February 1972), pp. 24-37.
19. J. Fitzpatrick, "A History of Automatic Microwave Network Analyzers," *Microwave J.* **24** (April 1982), pp. 46-56.
20. C. H. Young, "Measuring Inter-Electrode Capacitances," *Bell Lab. Rec.* **24** (December 1946), pp. 433-438.
21. L. D. White and H. T. Wilhelm, "A 0.1 to 10 MHz Dielectric Specimen Bridge with Dissipation Factor Accuracy of $\pm 10^{-6}$," *IEEE Trans. Instrum. Meas.* **IM-15** (December 1966), pp. 293-298.
22. L. D. White, "COZY: Accurate Automatic Measurements of Components," *Bell Lab. Rec.* **53** (May 1975), pp. 226-233.
23. S. Rosen, "Measuring System for Carrier Circuits," *Bell Lab. Rec.* **19** (May 1941), pp. 277-280.
24. R. D. Baum, "Streamlining Long-Distance Transmission Tests," *Bell Lab. Rec.* **43** (October 1965), pp. 372-375.
25. T. C. Anderson, "Testing Long-Haul Carrier Systems Automatically," *Bell Lab. Rec.* **52** (July/August 1974), pp. 212-216.
26. J. H. Rowen, "Ferrites in Microwave Applications," *Bell Syst. Tech. J.* **32** (November 1953), pp. 1333-1369.
27. S. Darlington, "Synthesis of Reactance 4-Poles which Produce Prescribed Insertion Loss Characteristics," *J. Math. and Phys.* **18** (September 1939), pp. 257-353.
28. E. G. Andrews, "Bell Laboratories Digital Computers," *Bell Lab. Rec.* **35** (March 1957), pp. 81-88.
29. R. S. Graham, "Relay Computer For Network Analysis," *Bell Lab. Rec.* **31** (April 1953), pp. 152-157.
30. S. Darlington, "The Potential Analogue Method of Network Synthesis," *Bell Syst. Tech. J.* **30** (April 1951), pp. 315-365.
31. M. R. Aaron, "The Use of Least Squares in System Design," *IRE Trans. Circuit Theory* **CT-3** (December 1956), pp. 224-231.
32. A. J. Grossman, "Synthesis of Tchebycheff Parameter Symmetrical Filters," *Proc. IRE* **45** (April 1957), pp. 454-473.
33. G. Szentirmai, "On the Reliability of Ladder Filters," *IRE Trans. Circuit Theory* **9** (March 1962), pp. 91-92.
34. G. Szentirmai, "New Developments in the Theory of Filter Synthesis," *IEE Conference on Electrical Networks*, London (September 1966), pp. 1-28.
35. See special issue on Statistical Circuit Design, *Bell Syst. Tech. J.* **50** (April 1971), pp. 1099-1310.
36. T. H. Simmonds, Jr., "The Evolution of the Discrete Crystal Single-Sideband Selection Filter in the Bell System," *Proc. IEEE* **67** (January 1979), pp. 109-115.
37. R. A. Laudise and R. A. Sullivan, "Pilot Plant Production of Synthetic Quartz," *Chem. Eng. Progress* **55** (May 1959), pp. 55-59.

38. R. A. Sullivan, "Growing Quartz Crystals," *W. Elec. Eng.* **3** (April 1959), pp. 2-10.
39. R. L. Dietzold, U.S. Patent No. 2,549,065; filed November 2, 1948; issued April 17, 1951.
40. J. T. Bangert, "The Transistor as a Network Element," *Bell Syst. Tech. J.* **33** (March 1954), pp. 329-352.
41. J. G. Linvill, "Transistor Negative-Impedance-Converters," *Proc. IRE* **41** (June 1953), pp. 725-729.
42. B. K. Kinariwala, "Synthesis of Active *RC* Networks," *Bell Syst. Tech. J.* **38** (September 1959), pp. 1269-1316.
43. I. W. Sandburg, "Synthesis of Driving-Point Impedances with Active *RC* Networks," *Bell Syst. Tech. J.* **39** (July 1960), pp. 947-962.
44. G. S. Moschytz, "Miniaturized *RC* Filters Using Phase Locked Loop," *Bell Syst. Tech. J.* **44** (May 1965), pp. 823-870.
45. L. C. Thomas, "The Biquad: Part I—Some Practical Design Considerations," *IEEE Trans. Circuit Theory* **CT-18** (May 1971), pp. 350-357.
46. L. C. Thomas, "The Biquad: Part II—A Multipurpose Active Filtering System," *IEEE Trans. Circuit Theory* **CT-18** (May 1971), pp. 358-361.
47. J. J. Friend, "A Single Operational-Amplifier Biquadratic Filter Section," *1970 IEEE International Symposium on Circuit Theory Digest of Technical Papers,* Atlanta, GA (December 14-16, 1970), pp. 179-180.
48. R. A. Friedenson, R. W. Daniels, R. J. Dow, and P. H. McDonald, "RC Active Filters for the D3 Channel Bank," *Bell Syst. Tech. J.* **54** (March 1975), pp. 507-529.
49. P. E. Fleischer, "Active Adjustable Loss and Delay Equalizers," *IEEE Trans. Circuit Syst.* **CAS-21** (July 1974), pp. 527-531.
50. Bell Telephone Laboratories, *Transmission Systems for Communications,* 1st Ed., Informal Notes (1954).
51. C. F. Simone, "Analog Computer Simulation—An Aid to Communication Systems Development," *Bell Lab. Rec.* **42** (June 1964), pp. 182-187.
52. J. Chernak, "Hybrid Computer Techniques and Applications," in *System Analysis by Digital Computer,* ed. F. F. Kuo and J. F. Kaiser (New York: John Wiley, 1966), pp. 315-348.
53. P. E. Fleischer, "Optimization Techniques in System Design," in *System Analysis by Digital Computer,* ed. F. F. Kuo and J. F. Kaiser (New York: John Wiley, 1966), pp. 175-217.
54. S. Narayanan, "Transistor Distortion Analysis using Volterra Series Representation," *Bell Syst. Tech. J.* **46** (May/June 1967), pp. 991-1024.
55. S. Narayanan, "Application of Volterra Series to Intermodulation Distortion Analysis of Transistor Feedback Amplifier," *IEEE Trans. Circuit Theory* **CT-17** (November 1970), pp. 518-527.
56. Y. L. Kuo, "Distortion Analysis of Bipolar Transistor Circuits," *IEEE Trans. Circuit Theory* **CT-20** (November 1973), pp. 709-716.
57. W. R. Bennett, H. E. Curtis, and S. O. Rice, "Interchannel Interference in FM and PM Systems Under Noise Loading Conditions," *Bell Syst. Tech. J.* **34** (May 1955), pp. 601-636.
58. M. L. Liou, "Noise in an FM System Due to an Imperfect Linear Transducer," *Bell Syst. Tech. J.* **45** (November 1966), pp. 1537-1561.
59. S. O. Rice, "Second and Third Order Modulation Terms in the Distortion Produced when Noise Modulated FM Waves are Filtered," *Bell Syst. Tech. J.* **48** (January 1969), pp. 87-141.
60. A. Anuff, M. L. Liou, and P. H. McDonald, "An Approach to Calculating Interchannel Distortion of Linearly Filtered Angle-Modulated Signals Using a Hybrid Computer," *Proc. IEEE* **57** (April 1969), pp. 736-737.
61. A. Anuff and M. L. Liou, "A Time-Domain Approach to Computing Distortion of Linearly Filtered FM Signals," *IEEE Trans. Commun. Technol.* **COM-19** (April 1971), pp. 218-221.

PART III
THE ADVENT OF DIGITAL SYSTEMS

Patents embodying the concepts of pulse-code modulation (PCM) were issued to P. M. Rainey of Western Electric in 1926 and to A. H. Reeves of ITT in 1939. Bell Laboratories developed a secrecy system that sampled and coded a speech signal for transmission over high frequency (HF) radio in World War II, and research on PCM continued in Bell Laboratories after the war. PCM transmission on microwave radio was considered, but the broad band required and the economy of analog systems inhibited the use of PCM for long-haul systems for many years.

Development of a PCM system for exchange-area paired cable was started in 1956 as solid-state devices became generally available. The feasibility of PCM transmission over ordinary cable pairs was demonstrated in the late 1950s. The 24-channel, 1.5-megabit experimental system established the basic parameters of 8-kHz speech sampling and eight-digit coding in a nonuniform near-logarithmic encoder. The first commercial system, T1 carrier, was placed in service in 1962 and was immediately successful. As device technology advanced, further development resulted in lower costs, smaller repeaters, and a succession of improved terminals. By 1972, digital channel banks with toll quality performance were available.

The 96-channel, 6.3-megabit T2 system, intended for medium-length toll transmission and to provide circuits for encoded signals for video telephone service, was introduced in 1972. Its requirement for special low-capacity cable and the failure of the visual telephone service to develop limited its use to a few locations. Further exchange development produced the T1C and T1D 3-megabit systems which provided 48 channels in exchange cable and allowed high usage of cable pairs. In the 1970s, integration of PCM transmission terminals with the trunk terminals of electronic switches, especially with the 4ESS* digital toll switch, produced large economies.

Work on high pulse-rate systems with large circuit capacity and suitable for encoded television pictures of broadcast quality produced experimental high-speed coders and decoders by 1964. The early high-rate systems were not competitive with long-haul analog systems, however, and the first installed high-rate system was T4M, a 274-megabit metropolitan system on coaxial cable, in 1975. High-rate digital transmission was favored from an early date for

* Trademark of AT&T Technologies, Inc.

transmission using the low-loss circular-electric mode in millimeter waveguide. Repeaters based on vacuum tubes proved impractical, but development of an all-solid-state system, WT4, was carried to a successful field experiment in 1974. The experiment demonstrated that 120 (60 two-way) 274-megabit channels were feasible with regenerators spaced at least 30 miles apart. This would have provided about 250,000 voice channels, but a slackening of circuit demand and the advance of optical fiber technology ended the project. A digital radio system, capable of transmitting 274 megabits in a 220-MHz band at 18 GHz, was also designed and tested in the field in the 1970s, but was not used commercially. Digital radio systems at 45 and 90 megabits in the 11- and 6-GHz bands, however, were successful for short toll links.

The invention of the laser in Bell Laboratories in 1958 kindled interest in optical transmission. Early work on transmission via lenses in pipes was redirected as low-loss fibers became available in the 1970s. Digital transmission at 45 megabits was demonstrated in laboratory and field experiments by 1975. First commercial service on a short system was achieved in 1979 and on a route from New York to Washington, D.C., by 1983.

Chapter 18

Pulse-Code Modulation on Cable Pairs—T1 Carrier

I. EARLY WORK

In the fall of 1962, the Illinois Bell Telephone Company installed T1 carrier, the first common-carrier digital telecommunications system in the world. This application of pulse-code modulation (PCM) represented a major departure from the past and marked the beginning of the age of digital communications. This section is devoted to the historical perspective and technical background leading to the development of T1 and the subsequent digital systems.

1.1 What is PCM?

The process of PCM—the conversion of an analog signal to a train of coded pulses—is based on two concepts. The first is that a continuous signal waveform, such as speech, can be represented by and reconstructed from isolated samples; the second is that the samples can be adequately represented by discrete numbers. The samples—the pulses—may be extremely short in duration; it is only their amplitude that is needed. The numbers used as a measure of the amplitude—the code—may be assigned in a variety of ways. In particular, if the numbers are expressed in binary form, each sample can be represented by a series of ones (positive pulses) and zeros (no pulses). A single binary digit is commonly termed a *bit*, and the transmission rate in digital systems is expressed in *bits per second*. The multidigit binary number used to measure an amplitude sample is called a *word*. The signal to be transmitted then consists of a stream of uniform amplitude pulses with intervening zeros [Fig. 18-1]. In the receiving terminal, the pulse samples are reconstructed in their proper amplitude, as determined by the binary number for each sample. Recovery of the original continuous signal is simple. The spectrum of the sequence of samples of varying amplitude contains, along with other frequencies, the original base band frequencies. All that is needed is a fairly simple low-pass filter to exclude all but the desired range.

There are two major purposes behind this rather complex process. First, the pulses, both the amplitude samples and in the binary-coded words, can be very short in duration, permitting time multiplexing of other sampled and coded signals in the intervals between them; and second, the uniform pulses of the

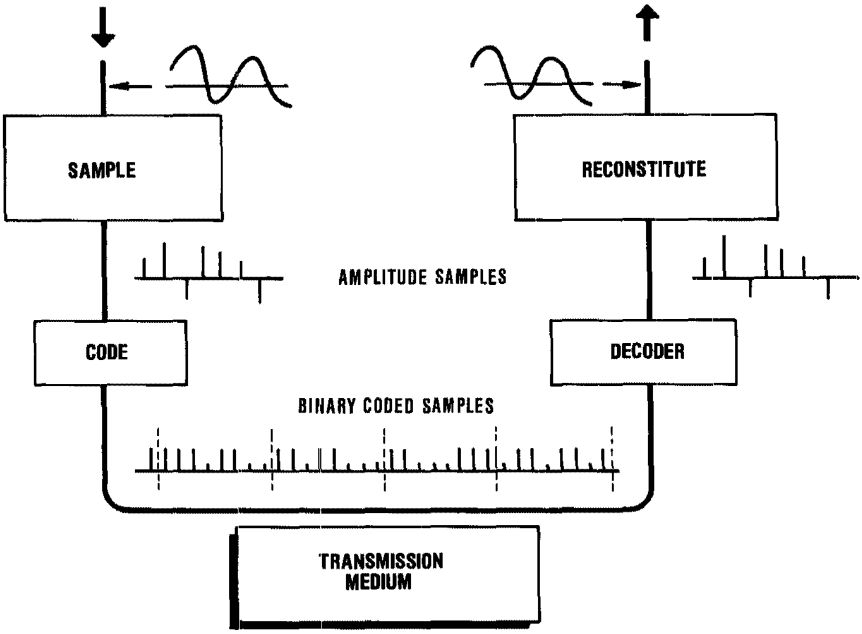


Fig. 18-1. Transmission by pulse-code modulation.

coded samples may be exactly reconstructed at regenerators along a line so that there need be no accumulation of impairments with distance, as there is in all analog systems.

1.1.1 Sampling

The question of how frequently a continuous signal must be sampled to ensure complete recovery from the set of samples is a time-honored problem of interpolation. As such, it has been the subject of treatises by well known mathematicians for more than a century. In 1841, the celebrated French mathematician, A. L. Cauchy came very close to enunciating what later became known as the sampling theorem.[1] (Cauchy [1789-1857], among many other contributions, established and placed the theory of the function of a complex variable on a sound basis. He belonged to that group of eminent French mathematicians whose special genius it was not only to advance basic mathematics but to provide tools of enormous value to applied science and engineering.) The earliest complete demonstration of the sampling theorem is generally attributed to a publication in 1915 by the British mathematician E. T. Whittaker.[2] In engineering terms, he showed that a signal that has no energy at or beyond a frequency f_0 can be determined uniquely from samples taken at a rate twice f_0. For example, if the signal band is confined to less than 4 kHz, it can be reconstructed exactly with 8000 samples per second. (If it surprises the reader to find that $2f_0T$ pieces of data will describe a continuous function completely over the interval T, it should be remembered that the $2f_0T$ coefficients of the

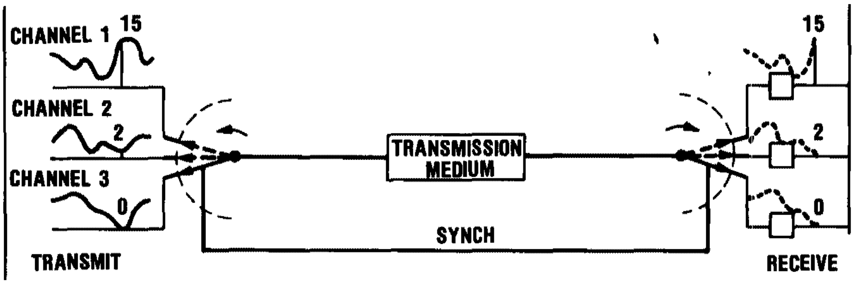


Fig. 18-2. Time-division multiplexing.

sine and cosine terms of a Fourier series do just this, if, as assumed, the function contains no frequencies higher than f_0.)

Communications engineers were largely unaware of the sampling theorem until they gave serious consideration to the transmission and reception of independent speech signals over a shared medium by time-division multiplexing. Time-division multiplexing is the oldest method of sharing a transmission medium by a set of signals [Fig. 18-2]. It was first developed when telegraph engineers realized that the signaling rate capability of at least some of their lines was much greater than the speed of the fastest manual operators. J. M. E. Baudot developed a printing time-division multiplex telegraph system in France in 1874.[3] The relatively low signaling rates were handled by synchronously rotating multisegment distributors. An operator or message tape was assigned a segment, and a five-digit Baudot code was transmitted during the momentary connection. The system was in wide use by 1900, around which time some patents were issued that proposed time-sharing with independent speech waves, but the inventors were unaware of the mathematics of the sampling theorem and the mechanical switches available precluded satisfactory performance with speech.

The sampling theorem was implicit in the work of J. R. Carson in 1920[4] and H. Nyquist in 1928,[5] but time-division multiplexing of speech still lay beyond the technical capability of that period. When the coaxial cable greatly widened the available frequency band in the 1930s, W. R. Bennett again examined the feasibility of pulse transmission and time-division multiplexing.[6] Bennett showed that for sampled (but not coded) speech, small variations in attenuation and phase from an ideal characteristic would produce intolerable crosstalk between channels. Amplitude sampling (pulse-amplitude modulation—PAM) was used for switching within the first speech-interpolation terminals (TAS1) and in some electronic PBXs,[7,8,9] but, for the reasons noted by Bennett, it is not used for transmission over any appreciable distance. From the standpoint of the subject at hand, PAM is important as a prelude to PCM.

1.1.2 Quantizing and Coding

The assignment of a discrete number as a measure of a sample that may have any of a continuum of values is called *quantizing,* and the method of expressing that number in a form that can be carried over a transmission medium is called *coding*. The differences between the actual value of the samples and the discrete numbers cause an impairment called *quantizing noise*. Clearly, the more closely the pulse samples are measured by the numbers assigned, the more perfectly the original signal will be recreated from pulses derived from those numbers, and the lower the quantizing noise [Fig. 18-3]. In binary coding it is apparent that decreasing the step size between successive binary words reduces the round-off error. For the same range of signal amplitudes, this requires a longer word with more binary digits—more bits. The choice of the number of bits to be used depends upon the signal being encoded: voice, video, or voice-band data; the number of tandem conversions between analog and digital expected; and the impairment that can be tolerated.

While other signals have become important, speech has remained the major part of the traffic to be carried. Because of the wide dynamic range characteristic of speech, the step sizes are usually arranged approximately logarithmically to conserve bits and reduce the signaling rate [Fig. 18-4]. This produces small steps for the measurement of low-level signals and larger steps for higher-level signals, making the quantizing noise approximately a fixed percentage of the signal amplitude, independent of its level. A uniform converter with step size equal to the smallest step of a logarithmically tapered converter would require many more bits and consequently more channel capacity and would be more difficult to realize. The logarithmic gradation of coding steps is equivalent to

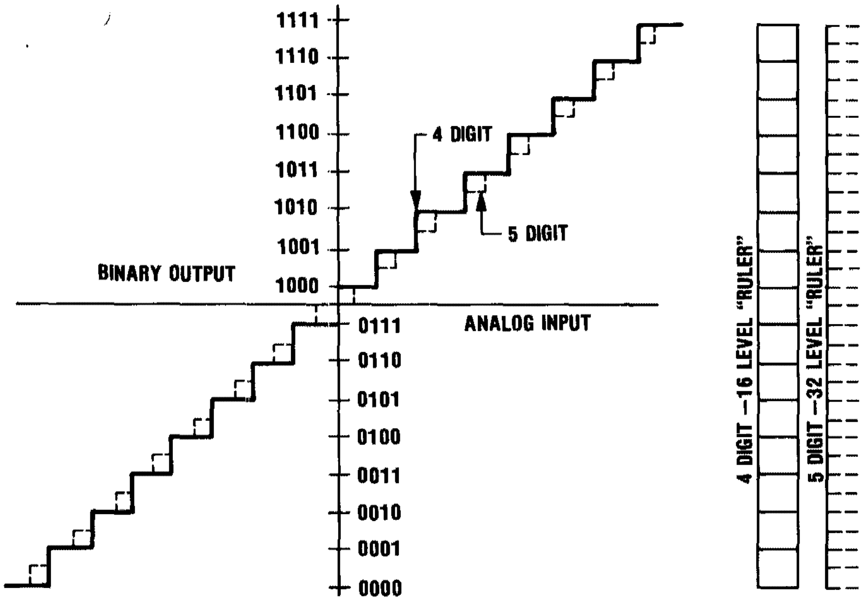


Fig. 18-3. Four-digit (solid lines) and five-digit (broken lines) uniform quantizing of an analog signal. The 32 levels of the five-digit code have smaller round-off error and lower quantizing noise.

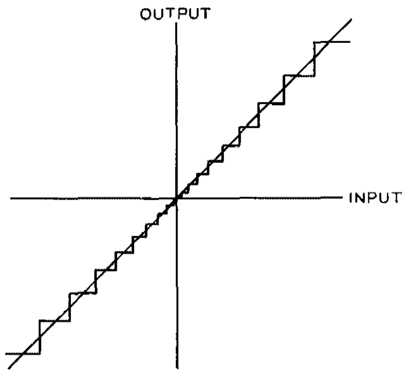


Fig. 18-4. The relation between input and quantized output with quantization logarithmically tapered.

the compression and expansion (companding) of speech volumes in analog systems, such as N carrier. Digital-system companding, however, is on an instantaneous, sample-by-sample basis, rather than at a syllabic rate as in the analog system.

1.1.3 Transmission and Regeneration

In the process of transmission, the pulse train is attenuated and spread out by the medium and corrupted by additive noise. The influence of these factors is removed by regenerative repeaters that are introduced periodically in the transmission medium [Fig. 18-5].[10] The incoming signal is first reshaped. This process consists of offsetting the loss and delay dispersion of the line by amplifying and equalizing the signal to restore the pulse train to the point where a pulse/no-pulse decision can be made. (On a typical cable pair, the individual pulses of a binary code-word may be smeared over several adjacent pulse time slots prior to equalization.) The time of decision is determined by a retiming circuit; the still-distorted signal is sampled where the pulse peak is expected. Depending on the presence or absence of a signal at the sampling moment, the regenerator will produce a pulse or no pulse, exactly reproducing the original signal. Reshaping, retiming, and regeneration are the three R's of PCM repeaters.

If all these functions are performed in an ideal manner, then the only remaining cause of errors is additive noise. If it is assumed that the pulse train has the simplest binary format, with ones represented by a positive pulse of amplitude V, and zeros by the absence of a pulse, zeros will be converted to ones when the noise amplitude at the sample time is positive and greater than $V/2$, and ones will be mistaken for zeros if the noise is negative and greater than $V/2$. If the noise has a Gaussian amplitude distribution with zero mean, which is characteristic of thermal and most solid-state-device noise, the error probability is a function of the ratio of peak signal to rms noise as shown in Fig. 18-6.

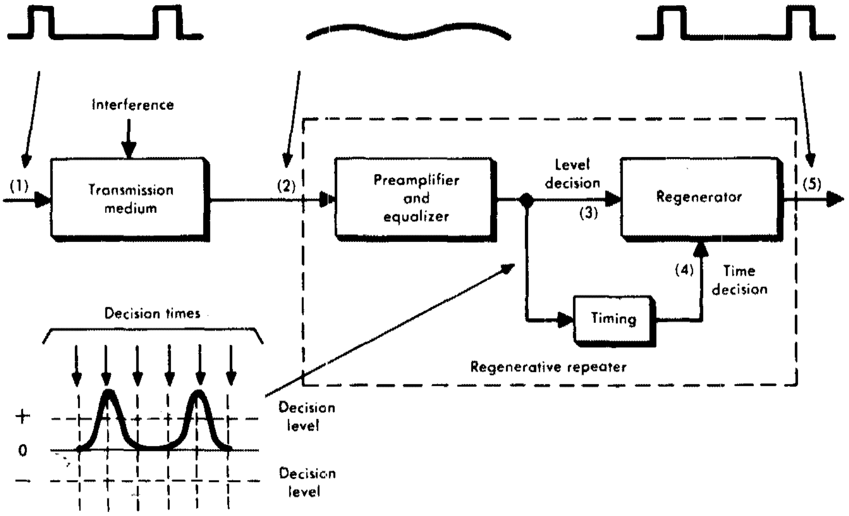


Fig. 18-5. PCM regenerative repeater section, showing the functions of reshaping, retiming, and regeneration.

Two conclusions may be drawn from this figure. First, there is a threshold in the signal-to-noise ratio, around 20 dB, such that an increase of only 1 dB or so reduces the probability of error by an order of magnitude. Second, the signal-to-noise ratio required to make the error rate negligible, say, less than one in one hundred million pulses, is considerably less than that required to achieve satisfactory performance in analog transmission systems. These properties of pulse-code digital signals have great significance for telephony and almost all other telecommunication signals.

For voice signals, an end-to-end system signal-to-noise ratio of about 30 dB or better is required to give satisfactory performance. In analog systems, the noise from each section of a long line adds up pretty much in proportion to the total length of the line. For example, in a multirepeater analog system with *N* repeaters, if the noise introduced in each repeater is random, the noise accumulates as $10 \log_{10}N$. Thus, for a 1000-repeater system (e.g., a transcontinental L3 coaxial system), the signal-to-noise ratio for each repeater section must be 30 dB better than the total system, that is, 60 dB, to achieve the overall 30-dB signal-to-noise figure. By contrast, in a digital system with regenerative repeaters, the increase in signal-to-noise ratio required to accommodate multiple-regenerator sections is extremely small. In 1949, these observations led B. M. Oliver, J. R. Pierce and C. E. Shannon to call regeneration "the payoff in PCM."[11]

This advantage is purchased at a price. The pulse-coded signals require a much greater bandwidth than the corresponding analog signals. For a 4-kHz voice band, amplitude sampling must be at a rate of 8000 per second. For binary coding, the further bandwidth expansion is equal to the number of bits in the analog-to-digital converter code. With the 8-bit code needed for low

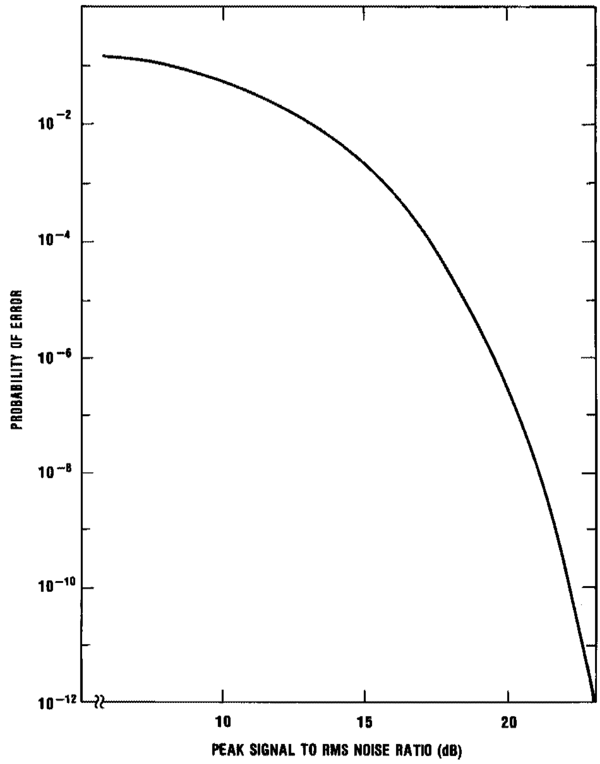


Fig. 18-6. Probability of a decision error in binary PCM as a function of signal-to-noise ratio.

quantizing noise, speech transmission required a pulse stream at 64,000 bits per second. Effective transmission of a pulse stream generally requires about 1 Hz of bandwidth per bit per second. The PCM band was, therefore, about 16 times greater than the normal 3- to 4-kHz analog band. The large bandwidth required was one of the initial deterrents to the introduction of PCM for long-haul transmission.

1.2 PCM Research to 1955

PCM was first disclosed in a patent issued to P. M. Rainey of the Western Electric Co in 1926.[12] Rainey was interested in converting a continuous facsimile signal to a discrete form so that it could be transmitted over telegraph channels. The facsimile telegraph system disclosed in this patent consisted of relays, rotating mirrors, ballistic galvanometers, and light sensitive cells. It is remarkable that this collection of electro-optical-mechanical components embodied the principles of sampling and analog-to-digital conversion at the heart of PCM. Rainey's invention, however, received little attention at the time and was actually forgotten and not brought to light again until many other PCM patents had been issued.

Independently, A. H. Reeves, working in the Paris Laboratories of the International Telephone and Telegraph Company (ITT) rediscovered PCM in 1938.[13,14] Reeves attempted to overcome the problems of transmitting pulse samples of varying amplitude, the method that Bennett later showed to be impractical, first by varying the timing of fixed-amplitude pulses. However, he and his associates soon realized that the effect of noise and distortion would affect the timing as well as the amplitude, and he then proposed a code with pulses and zeros, uniform in time and amplitude. He guessed that 32 levels would be adequate for constant speech volume but did not suggest nonlinear (logarithmic) coding. His disclosure proposed electronic circuitry for the application of PCM to speech transmission, and he considered it most likely to be used on microwave-radio links, where the wide band required might be easily obtained, and on which ITT was also experimenting in the 1930s. Apparently, the invention was not reduced to a working system as, indeed, it would have been difficult to accomplish with the prewar devices available.[15]

1.2.1 Coding and Secrecy—Military Systems

Somewhat before the entry of the United States into World War II, a group of circuit-research engineers at Bell Laboratories was presented with a problem that, in effect, made PCM mandatory. The privacy systems used on the prewar transatlantic radio circuits merely subdivided the voice band and then permuted and inverted the subbands in various ways, always occupying the same bandwidth. Even the privacy offered was limited, unless the subband scrambling was frequently changed. The problem, as war approached, was to develop a truly secret method for the transmission of speech.

In World War I, the secrecy problem for telegraphy had been solved by G. S. Vernam.[16] The method consisted of adding a random binary key, modulo 2, to a synchronous binary message. Modulo 2 addition of the numbers 0 and 1 proceeds as in ordinary arithmetic, except that $1 + 1 = 0$. Thus, if M is the message sequence and K is the key sequence, then the transmitted sequence is $M + K$. If the key is known at the receiver, it is added to $M + K$ to give $M + K + K = M$, since $K + K = 0$. If the key is completely random, all messages decoded without it become equally probable. Recovery of the original message can be performed only if the key is known.

The need in 1940 was for secure encryption of speech on an ordinary voice channel. Digitizing ordinary speech and transmitting in a Vernam code was impractical because of the greatly increased bandwidth required. Fortunately, a method for compressing the speech band was at hand in the form of the vocoder, invented in Bell Laboratories by H. Dudley in 1936.[17] The vocoder provided a bandwidth reduction of speech by a factor of ten, where the preservation of reasonable intelligibility, but not speech quality, was sufficient. If a vocoder signal could be transmitted by digitizing samples over an ordinary voice channel, a Vernam speech-encryption system could be furnished for wide usage.

Details of a vocoder-based secrecy system were worked out by engineers of the Bell Laboratories circuit-research group in 1941. There was no government contract at first, and the early investigations had no classified status, but concern for privacy or even cryptanalyst-resistant secrecy on radio systems had been a growing concern in Bell Laboratories as the war approached. The scheme that evolved transmitted ten vocoder spectrum channels and a pitch channel. The bandwidth of each of these 11 channels was restricted to about 25 Hz, and hence each required sampling at a rate of 50 times per second. Two-level binary coding would still have required too large a band. It was found by trial that the spectrum channels could be adequately represented by a voltage scale with six steps, but the range of pitch values needed a finer representation. This range was met by a further vernier subdivision of each of the six primary steps into six smaller steps. This 36-level quantizer was the first to use nonuniform step sizes. What was found suitable, therefore, was six-level or hexary-PCM transmission, in which one hex digit sufficed for each of the ten spectrum channels, and two hex digits were provided for the pitch channel [Fig. 18-7].

Complete transmission required sending 12 six-valued pulses 50 times per second. The total number of pulses per second was thus 600, which was well within the capability of radio voice channels at that time. Since a principal application was to be on transoceanic high-frequency radio circuits, which were plagued by multipath transmission and with propagation time differences across a voice channel of as much as two or three milliseconds, it seemed best to send the 12 pulses simultaneously in separate frequency subbands of the voice channel. To send the group of pulses sequentially would have allowed only 1.67 msec per pulse; parallel transmission allowed 20 msec for the reception of each pulse. The six levels were sent in the individual frequency bands by transmitting one of six discrete frequencies within each band. The random key for both ends was obtained by recording six-level samples of uniformly distributed random noise and playing synchronized copies of the record at the

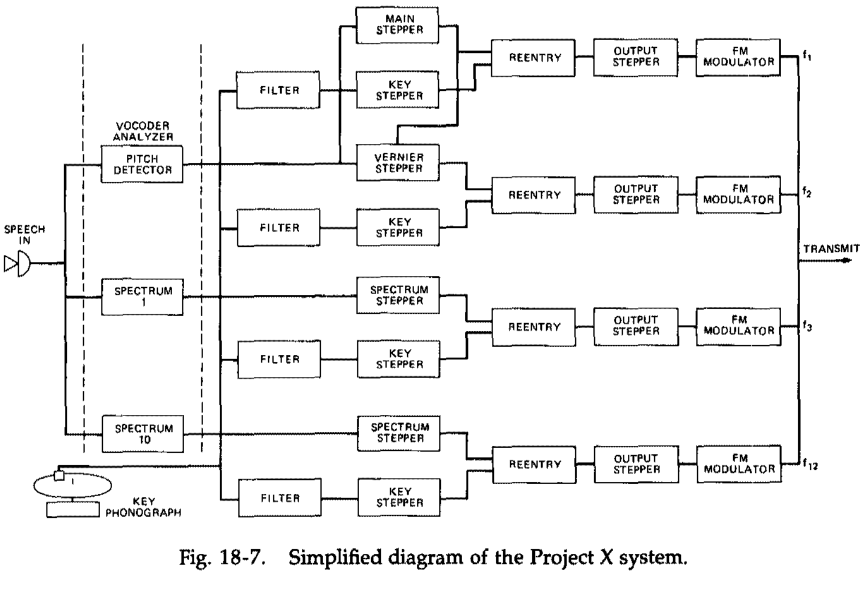


Fig. 18-7. Simplified diagram of the Project X system.

transmitter and receiver. The system, which was later known as the 6-Sally system, or Project X could be described as a 12-channel hexary PCM-FM-FDM (frequency-division multiplex) design. (See another volume in this series, *National Service in War and Peace (1925-1975)*, Chapter 5, Section 4.3.) Bennett cites actual instances of system usage, including the initial portion of the transcript of a conversation between Admiral W. Leahy and Prime Minister Winston Churchill during World War II.[18]

In terms of 1980s technology, Project X was a monster, weighing "a couple of pounds less than a sawmill."[19] Nonetheless it met a need and included several firsts: (1) the realization of Vernam-enciphered telephony, (2) the use of nonuniform quantization, (3) the use of PCM for speech transmission, (4) the use of multilevel frequency-shift keying (FSK), (5) the useful realization of speech-bandwidth compression, (6) the use of FSK and FDM to combat fading.

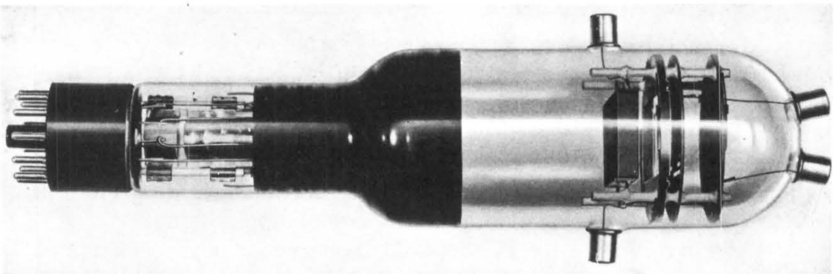
Another digital secrecy system using PCM applied to the complete voice band was developed by H. S. Black and his colleagues during World War II. Intended for use over military microwave links and called AN/TRC-16, it was an eight-channel system using 8000 samples per second with five-bit (32-level) approximately logarithmic quantization. Although built for the United States Government, it was not used during the war because the technology of recording and reproduction was not sufficiently advanced to provide a truly random synchronized key at the high pulse rate of 320 kilobits per second.[20]

1.2.2 Research Toward Commercial Systems

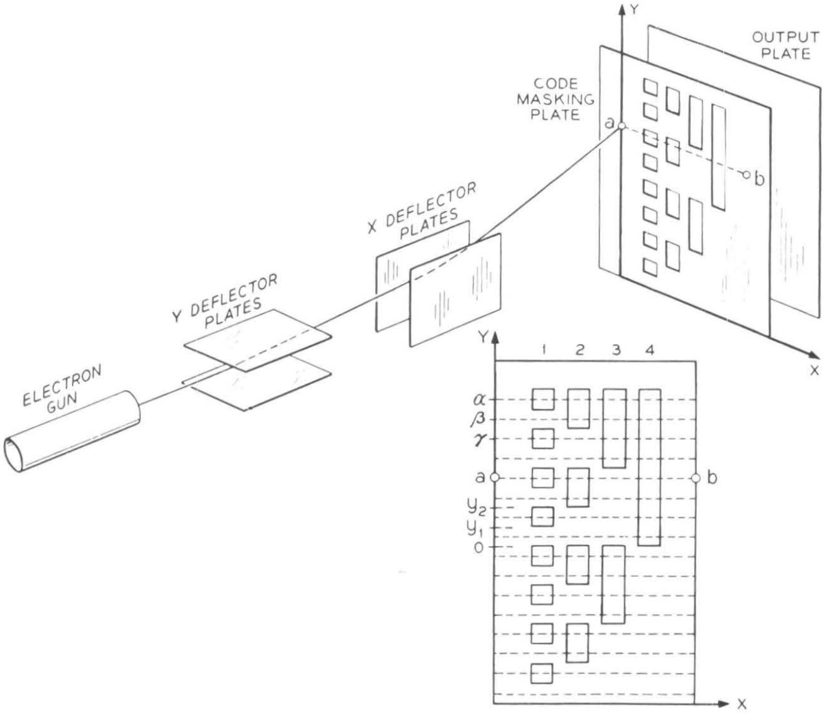
Before the end of World War II, PCM was explored for possible application to commercial telephony. This work was security classified because of the sensitivity of the subject with regard to military applications. The classification was later removed, but many of the early results remain unpublished. Two laboratory models were eventually demonstrated and described in the literature. One of these was a 96-channel system proposed for use on microwave radio. It represented the fruition of an extensive series of experimental and analytical investigations of the problem.[21,22] Part of the work involved subjective tests to determine the number of quantizing steps and their best spacing for speech signals. Circuits operating with the speed and accuracy required to perform the sampling, quantizing, and coding operations were developed. Coding was accomplished with an electron-beam coding tube in which the beam was deflected by the sampled voltage of the signal and then swept across an aperture plate [Fig. 18-8]. To reduce the effect of ambiguities at the vertical transitions, F. Gray devised a binary code in which adjacent numbers differed in one, and only one, binary digit.[23] Gray codes subsequently found widespread application in digital work.

1.3 Areas of Application

By the late 1940s, the fundamental properties of PCM were well understood and possible applications were being examined. The ability to regenerate pulses and to make performance essentially independent of distance at first prompted the study of the possible use of PCM for long-haul transmission. While coaxial cable was known to have a very broad band, only the lower frequencies had



(a)



(b)

Fig. 18-8. Coding by an electron-beam coding tube. (a) An early electron beam coding tube to transform speech samples into pulse codes, 1947. (b) The electron beam deflection system and code plates.

low loss, and these were used with high bandwidth efficiency by single-sideband AM. Much of the work immediately after the war was, therefore, carried out with application to microwave radio in mind. For a time, when radio bandwidth seemed abundant and channel capacity objectives were modest, PCM appeared competitive with FM. But, for higher-capacity systems, FM did not require the wide bandwidth of PCM and was much easier to implement with the technology available. Another possible medium was circular-electric-mode millimeter waveguide, on which research work was resumed after the war. But, while PCM did indeed become the preferred modulation method in the later waveguide project, the research work on the waveguide medium was in too early a stage to warrant a substantial effort on the modulation scheme.

For a time in the late 1940s, PCM looked like a solution looking for a problem. In 1948, a committee was appointed to study and recommend a field of application but came up with no persuasive answer. The research effort was phased down and pursued at a much reduced level. Finally, attention was directed to the only remaining area of application—the wire pairs of the exchange-area transmission plant.

1.3.1 PCM on Exchange-Area Pairs

In many ways, this was an attractive area for PCM. In theory at least, a broad band was available. Both multichannel analog carrier telephony and television had been transmitted over ordinary pairs in the late 1930s. PCM was highly resistant to noise and crosstalk, which were major problems for analog transmission on pairs. Finally, the invention and development of transistors promised to reduce the cost of terminals, especially digital terminals, to a level where carrier on short trunks would be increasingly economical. The high cost of vacuum-tube analog terminals had been a major factor in the failure in the late 1940s to prove-in N1 carrier on exchange trunks.

The failure of N1 carrier to provide low-cost exchange trunks had been very disappointing, and interest in an effective multichannel carrier system for the exchange area remained very high in the local telephone companies. By the end of 1952, the Illinois Company, for example, had 680,000 miles of trunk pairs in place in Chicago and was placing 30,000 additional pair-miles per year. Furthermore, more than 80 percent of this was in Nos. 22 and 19 gauge conductors, requiring respectively more than twice as much and four times as much copper per mile as the No. 26 gauge pairs widely used in subscriber loops. A broad survey in 1952 of over one million local trunks in all operating companies and further specific studies in 1953 and 1954 of the situations in New York, Boston, Seattle, Philadelphia, and elsewhere all pointed to the need for lower cost and flexibility in providing local trunks. The companies were aware that new technology, such as electronic switching, might require major changes in plant layout and were reluctant to continue making large investments in cables that might have relatively short lives.

The surveys also established the characteristics desired. Conversion to dial operation was in its final great surge. Better-quality transmission with lower and more uniform loss was needed. However derived, the trunks would have to work with a wide variety of control signaling arrangements. Any carrier system had to work on existing as well as new pairs. Large route capacity was not desired, at least not if it also meant higher costs. The companies felt that the use of carrier, in general, should defer the need for new cable for 10 to 15 years, and this indicated that a system with only six to eight circuits would be optimum. (This was the subject of much debate later as the 24-channel T carrier system took shape.) Finally, the surveys confirmed early estimates that the potential market was very large. AT&T and Bell Laboratories estimated that an economical exchange-trunk carrier system would find a market ten times greater than any earlier carrier system. (Actual use would first confirm and then far exceed the estimate.) The surveys indicated, for example, that if a carrier system could be made economical at six rather than ten miles, the demand would triple. A quotation from an internal AT&T report following the Boston-area survey conveys the sense of urgency that existed in 1954. "A very important factor to consider today is that such a carrier system is needed just as soon as it is practical to develop it. Speed is essential!"

While the surveys were being conducted, Bell Laboratories initiated a study of alternative ways of meeting the need that was being defined. The conclusions of the study were presented to AT&T on March 1, 1955. Four options were presented and analyzed: (1) adapt the newly developed, solid-state, Type P rural-subscriber carrier system to the exchange area by adding appropriate signaling capabilities, (2) transistorize the vacuum-tube N carrier and add necessary signaling, (3) use finer-gauge wires with voice-frequency gain, and (4) use pulse techniques.

The major conclusion was that the first three options, although feasible and deliverable on a shorter schedule, promised only moderate savings, whereas savings with pulse techniques could be substantial. Attempts to realize PCM terminals using early point-contact transistors had proved disappointing. The new devices were not yet developed to the stage for successful use. Some vacuum tubes specially designed for digital rather than linear applications had also been tried, but this too proved to be a dead end. Nevertheless, solid-state technology was advancing rapidly in the mid-1950s, and Bell Laboratories indicated that a PCM carrier system, although it had not as yet been demonstrated to be feasible on exchange pairs, could probably be made available in about five years. It was to be seven years, but there was no occasion to regret the decision.

The advance of transistor technology was basic to the project. Distances between terminals would be relatively short and terminal costs would dominate the economics. Junction transistors, first demonstrated by Bell Laboratories in 1951, were much more stable than point-contact versions and were expected to make solid-state PCM terminals feasible and reliable at low cost. Much of the digital signal processing could be common to all, or at least large blocks, of the channels. Thus the cost per channel in a PCM terminal decreased as the number of channels increased, favoring a larger number than the 12 used in N carrier (or the 6 to 8 favored by the local companies). In contrast to the analog options considered, where only a few repeaters at wide spacing would be needed, the high digital rate in a multichannel PCM system would require regenerative repeaters every mile or so along the pairs. Here too, the new transistors were essential to the realization of a small, low-power, inexpensive, and highly reliable unit.

The recommendation to proceed with a PCM system was enthusiastically received, and exploratory work, which was already under way, was accelerated.

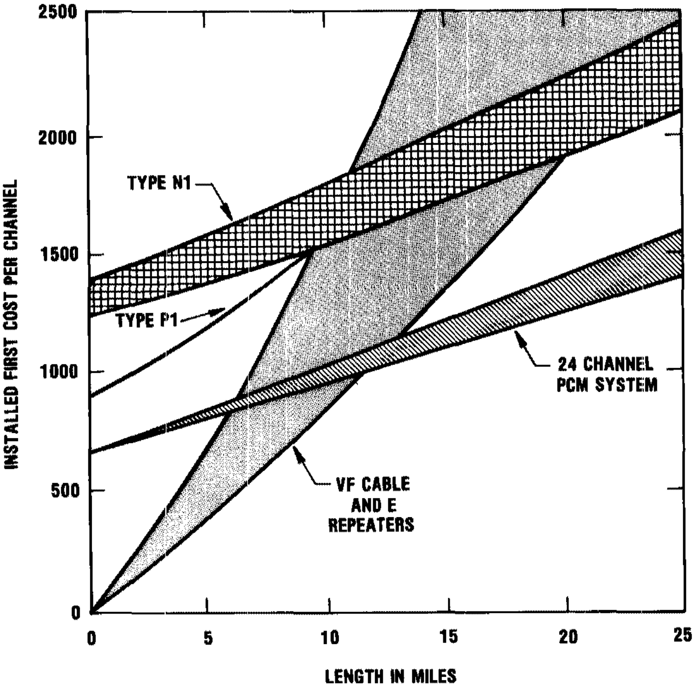


Fig. 18-9. Comparison of installed first cost for exchange trunks, 1955. The range of costs for each facility is for different route conditions.

The economic analyses were also continued and refined based on the early laboratory work. The comparisons made in early 1955 were quantified and published in an engineering prospectus in early 1956 [Fig. 18-9]. This continued to show a substantial advantage for PCM trunks for distances over ten miles. The installed-terminal cost per channel of a 24-channel PCM system was estimated to be about half the cost of a modified N carrier terminal. It was also recognized that a PCM system would be easier to engineer, install, and maintain than an equivalent analog system, and that it could provide lower-cost signaling, which was always an expensive adjunct to the analog carrier terminals.

II. DESIGN AND REALIZATION

Despite the wealth of excellent theoretical work and laboratory research in PCM up to 1956, the developers of the commercial system faced some large unexplored areas and major problems. Because of this, even though the service need was urgent, the initial development was exploratory in nature, with no firm commitment to a service schedule. The transistors available at the start were alloy-junction devices which, while a big improvement over the earlier point-contact devices, were costly and still had limited performance capability and reliability. It was necessary to develop and characterize new devices to meet both the performance requirements and cost objectives. Finally and most crucially, the cable plant environment had to be characterized for transmission performance, noise, crosstalk, and physical conditions. The new system required knowledge of the properties of the paired cable medium up to the megahertz region for a new type of signal where, with few exceptions, only voice or low carrier frequencies had been transmitted before. Therefore, it was deemed essential at the outset to equip an early experimental line and test it in a typical field environment. This would provide an early check of the system concepts, the circuit design approaches, and the overall performance that could be expected.

2.1 System Features

A broad set of system features emerged from the research, system engineering, and first year of exploratory work.[24]

In the terminals:

- Sampling would be at a rate of 8000 per second to provide the equivalent of a 4000-Hz speech channel.
- Twenty-four encoded voice channels would be furnished. This was a larger block than the optimum indicated in the company surveys, but the early work confirmed that much of the most expensive equipment in the terminal could be made common to many channels, thus reducing the cost per channel. The objective was also supported by a feeling that the views based on the current plant growth might be too conservative.
- The quantizing characteristic was to be nonlinear to match the wide dynamic range of speech. Analytical, as well as subjective, studies indicated that seven-bit binary encoding with approximately logarithmic spacing of levels would provide a signal-to-noise ratio of about 30 dB over almost a 40-dB range of signal levels.[25] With these parameters, several conversions from analog to digital to analog (A/D/A) could be accepted as the digital lines were placed in tandem with intervening analog systems. In this real-world network situation, each A/D/A translation would introduce quantizing noise whose total had to be held to low limits. Under normal operating conditions, system performances would be defined almost entirely by the quantizing noise and by crosstalk within the terminals.
- Supervisory signaling was built into the system by assigning an eighth bit to each channel for that purpose. When four-state signaling was required with some switching machines, the least significant of the seven speech digits was borrowed for the brief signaling interval. The built-in signaling feature led to very inexpensive and effective implementation compared to the signaling systems required for analog carrier systems. In the analog systems, signaling costs per channel were a significant fraction of the total cost for speech multiplexing. The large advantage in signaling costs for digital systems proved to be a major factor in their economic success.

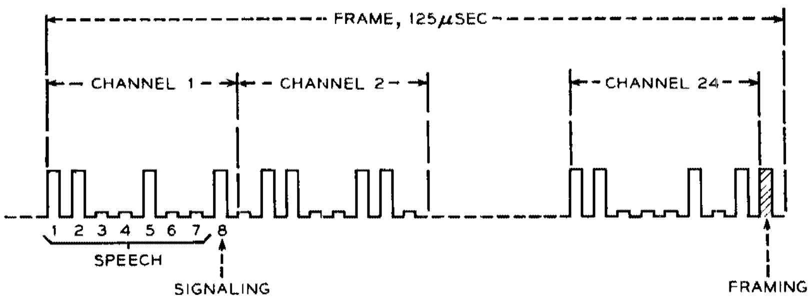


Fig. 18-10. The experimental PCM system digital word and frame format, 1957.

In the repeatered line:

- The line pulse-repetition rate was 1.544 megabits per second. This resulted from the choice of eight-kilobit sampling, eight bits per sample (seven information bits, plus one for signaling), and 24 channels. The channels were sampled and encoded in sequence. After the 24th channel, an additional bit was added to mark the end of each sequence of channel samples (a frame) [Fig. 18-10]. The digital rate per channel was, therefore, 64 kilobits per second (the digital-signal-zero [DS0] rate) and the line pulse rate was $8000 \times (24 \times 8 + 1) = 1.544$ megabits per second (the DS1 rate).
- The system was to operate on Nos. 19 or 22 gauge and possibly No. 24 gauge cable pairs in existing cables.
- Nominal repeater spacing was 6000 feet, so that repeater locations would usually coincide with existing manholes for loading.
- Repeaters were to be powered along the line from the terminals.
- The repeaters were to achieve timing from the transmitted-pulse train.

2.2 Terminal Design

By 1959, the terminal organization that emerged from the early development was pretty well settled [Fig. 18-11].[26] The voice signal on a two-wire line was split into transmit and receive paths by means of a hybrid circuit. The PCM system, like other carrier systems on cable, was a physical four-wire system, with separate, identically equipped pairs, for the opposite directions of transmission. On the transmit side, the signal was passed through a low-pass filter to limit the signal band to about 3.5 kHz and severely attenuate frequencies at and above 4 kHz. Every 125 microseconds (i.e., 8000 times a second), the channel transmitting gate was activated by a channel pulse to place a signal sample on a bus common to one-half the channels—the odd or even group transmitter PAM bus. Dividing the channels into 2 groups of 12 and processing a sample from each group alternately provided a recovery interval for each compressor-encoder that otherwise would have been unmanageably short. One signal sample at a time was passed through a compressor to narrow the am-

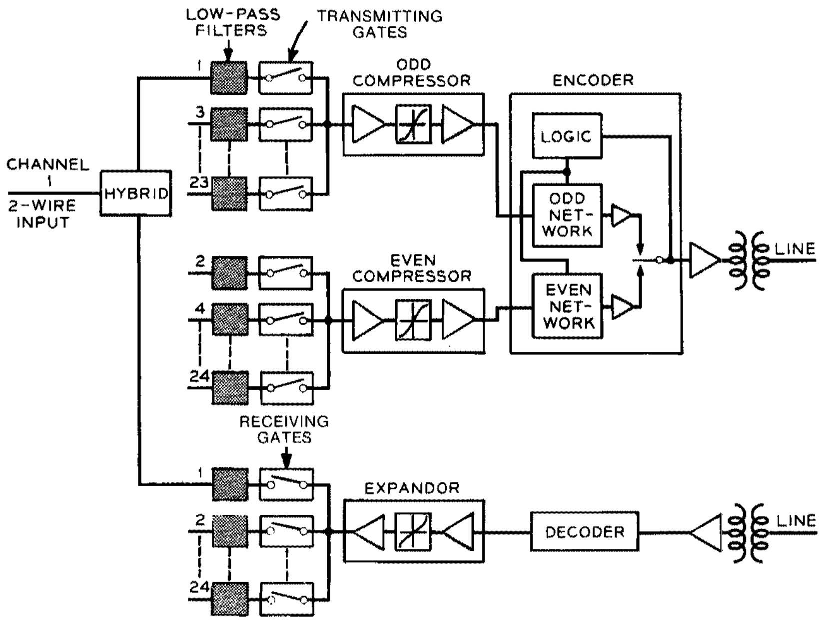


Fig. 18-11. The organization of the experimental PCM terminal, 1959. Signaling, maintenance, and power details are omitted for simplicity.

plitude range of the pulses to be coded and to give the desired preferential gain to low-level signals. Each compressed sample was then coded into a seven-digit binary number by means of a uniform-step-size coder. The combination of an instantaneous compressor and uniform-step coder yielded the approximately logarithmic coding desired. The signaling status of the channel (i.e., on hook, off hook, dial pulsing) was denoted by the eighth bit, appended to the seven information bits to make up the eight-bit channel word. The sequence of 24 eight-bit words thus required 192 bits. At the end of the sequence, a single additional bit, the 193rd, was added to identify the end of the frame for the receiver. With 193 pulse positions possible in the 125-microsecond frame, each digit was assigned a time slot equal to 0.65 microsecond, and the eight-bit channel word occupied 5.2 microseconds.

At the receiver, the coded pulse train was sorted out. The frame was identified and the receiving terminal locked into synchronism by an alternating pattern of ones and zeros in the 193rd time slot. Signaling pulses were directed to the individual channel-signaling units and the information-bearing voice-band pulses fed to a decoder. The decoder output was a reconstructed PAM signal whose amplitude samples were ideally within one-half of a coder step of the transmitting-coder input signal. The PAM pulses were then passed through an expander, whose input-output characteristic was the inverse of the compressor, thereby achieving an overall linear channel. The decoder and expandor operated very rapidly, and a single unit was common to all the channels. Finally the time-multiplexed PAM pulses on the common receiving bus were sequentially gated to their respective receiving low-pass filters to yield the original voice signal, plus the inevitable quantizing noise. A wide-band power amplifier in the expandor, common to all the channels, raised the level of the PAM signals prior to gating, so that no additional per-channel amplification was required.

The sequence outlined is ideal; that is, that was the way things would happen if all the terminal circuits behaved exactly as intended, and the regenerative repeatered line introduced no digital errors or excessive timing jitter to change the relative position of the pulses. There was a great deal of innovation and invention in the introduction of PCM carrier, and it was absolutely essential; but approximating the idealized state with practical components and within economic constraints was what good engineering design (and 90 percent of the development effort) was all about.

2.2.1 Multiplex Gates and Filters

The circuit required to produce the time-multiplexed PAM pulse train for processing by the coders was conceptually very simple. High-level PAM pulses were necessary to minimize the effects of crosstalk from the large control signals used to turn the sampling gates on and off, but no per-channel amplification was used to save cost. This led to the use of energy sampling, rather than voltage sampling, since the latter approach resulted in attenuation equal to the percentage of time the gate was closed, whereas energy sampling was theoretically lossless. The operation of channel-sampling and time-multiplexing may be understood by reference to Fig. 18-12.

Between sampling times, the output capacitor, C_1, of the input filter accumulated a charge proportional to the signal level. During the time the channel-gate pulse was present, the diode-bridge switch was closed and the full charge on C_1 was transferred to the holding capacitor, C_c, common to all 12 even or odd channels. The coupling inductor, L, and the two capacitors were chosen so that one-half their resonant period was equal to the channel-pulse width, to achieve so-called resonant transfer. With all switches open, the voltage on capacitor C_c remained constant during the encoding period. This was essential for the code to represent the sample properly. Finally, on completion of the encoding cycle, the clamp switch was closed for 1.95 microseconds to remove the sample charge from C_c in preparation for the arrival of the sample from the next channel. Closing the gates in the proper sequence formed a train of short pulses of different amplitudes, time multiplexed for transmission to the common compressor-encoder.

The design of a filter terminating in a periodically operating switch presented a new problem. For the two-wire to four-wire hybrid to operate properly with good separation of the directions of transmission, the filter impedance had to present a good match to the hybrid at all signal frequencies. First attempts yielded very complex filters with as many as 15 elements. But further analysis showed that the switch could be simply represented by a resistance and inductance that were only slightly frequency dependent. A filter meeting both transmission and impedance constraints was then realized with only five elements.[27]

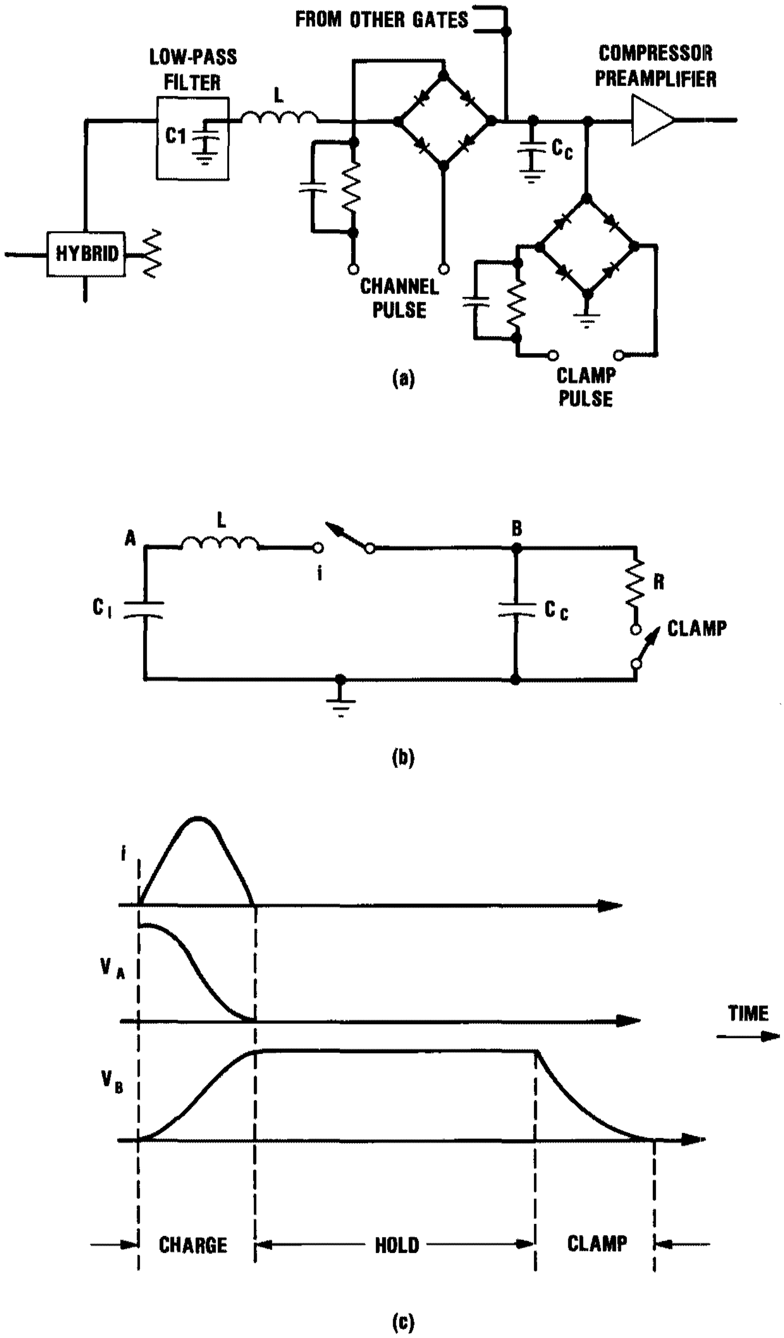


Fig. 18-12. Transmitting sampling gate and resonant transfer. (a) Multiplex gate and clamp schematic. (b) Equivalent circuit. (c) Voltage and current time functions.

2.2.2 *Common Circuits*

The common circuits—the compressors, expandors, encoders, and decoders—operated on the multiplexed PAM pulse train to produce the PCM line signal at the transmitter and recovered the PAM signals in the receiver [Fig. 18-13].[28] The coding process took seven time slots to form the seven signal-information-digit code. If a single compressor-encoder were used, this would have left only the eighth-bit interval of 0.65 microsecond to clear and recharge the holding circuit. With dual compressors, operating alternately on odd and even channels, nine-bit intervals were made available for recycling. This reduced the bandwidth requirements and minimized interchannel crosstalk from pulse to pulse. The encoder had separate analog circuitry to face each compressor, but used a time-shared logic circuit to operate alternately on each half of the coder. At the output of the dual encoder, the binary code words from the odd and even sides were combined into a single PCM bit stream for transmission over the line.

2.2.3 *Compression and Expansion*

Because of the nature of speech and variations between talkers, nonuniform encoding was essential to the success of the PCM system. In a linear encoder of any reasonable number of steps, the quantizing noise would have been excessive for weak speech. With approximately logarithmic encoding, quantizing noise for almost all speech volumes was held to acceptable limits with only seven digits. Seven digits provided 2^7 or 128 levels, 64 for positive-signal excursions and 64 for negative-signal excursions [Fig. 18-14]. For uncompressed samples in a uniform coder of 64 positive levels, the maximum quantizing noise at the lowest coding level could be ±50 percent of the sample, while the quantizing error for the highest level signal would be only ±0.8 percent. With the compressor used, the weak sample was amplified by a factor of 20 making a half-step error only ±2.5 percent of the sample value. Samples at the highest level were attenuated, resulting in an increase in the quantizing error from 0.8

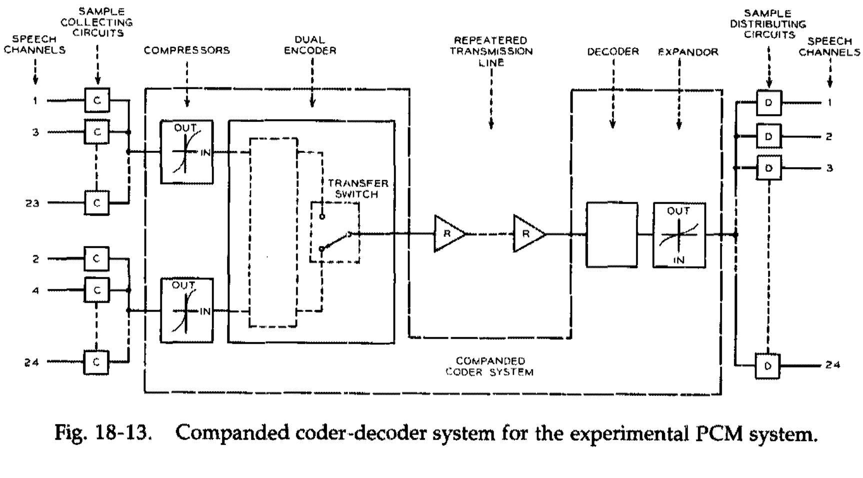


Fig. 18-13. Companded coder-decoder system for the experimental PCM system.

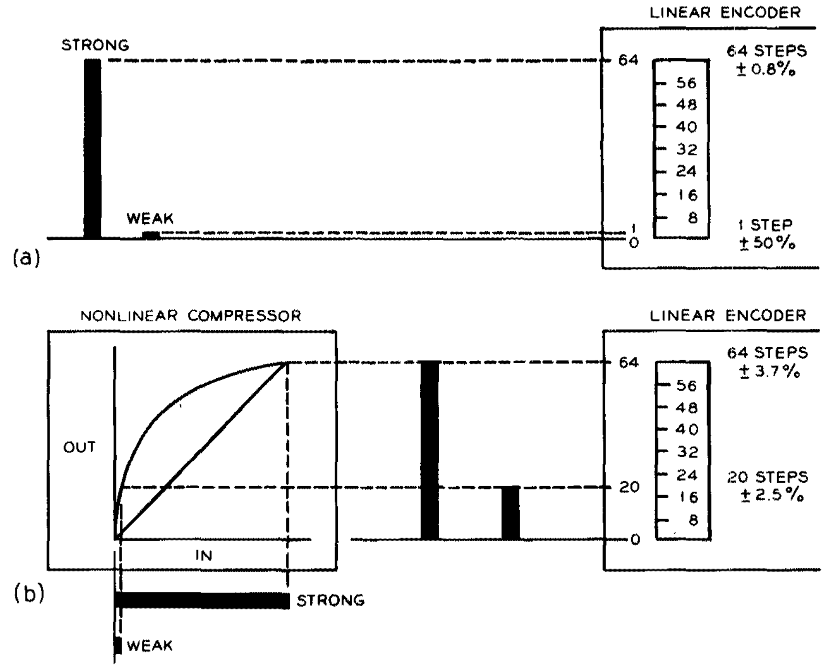


Fig. 18-14. Coding of speech samples. (a) Uniform coding yields high quantizing noise for weak signals. (b) Approximately logarithmic nonlinear compression yields near-uniform quantizing noise for all speech levels.

percent to 3.7 percent; but, as a result, quantizing noise was acceptable and approximately uniform over a large range of speech volume.

There are several ways of achieving a nonuniform coding characteristic. In the experimental system, the method chosen was to compress the PAM pulses before encoding them in a uniform step coder in the transmitting terminal and, inversely, expand them after decoding in the receiving terminal. The companding action was instantaneous in the sense that the PAM pulses were compressed and expanded on a pulse-by-pulse basis by fast-acting circuits, in contrast to the slow-acting, speech-volume-controlled syllabic compandors used in the overseas radio and N carrier analog systems. In the PCM terminals, encoding and decoding of the compressed pulses was then accomplished in uniform-step circuits.

There are, of course, an infinite number of logarithmic characteristics. In the one chosen, the degree of compression was defined by the parameter μ [Fig. 18-15]. There were a number of reasons for desiring a high degree of compression, i.e., high μ. A high μ provided a large companding improvement for weak signals, reduced idle circuit noise, and increased the permissible signal volume range. On the other hand, a high value made the overall circuit loss less stable and accurate tracking of the compressor and expandor for overall linear operation more difficult. A μ value of 100 was chosen for the experimental

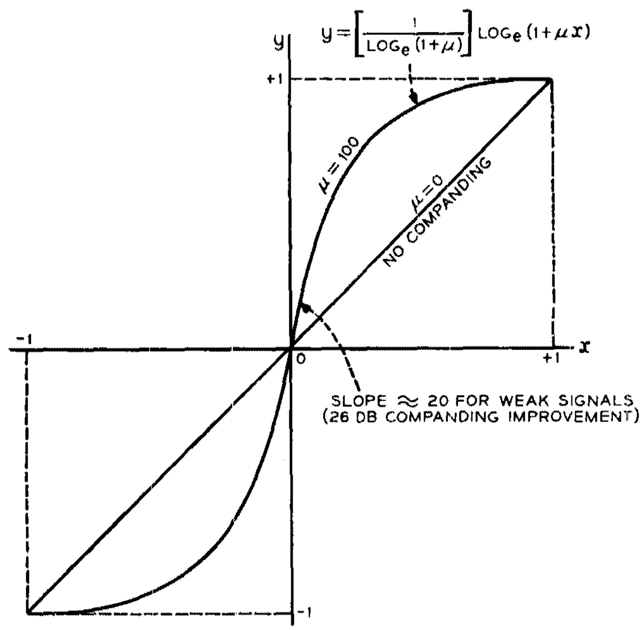


Fig. 18-15. Compandor characteristics defined by the parameter μ. In the compressor, x is the input and y is the output. In the expandor, they are reversed.

system, resulting in a weak signal companding improvement of almost 26 dB, while keeping the other problems within manageable bounds.

The desired companding characteristic was approached by exploiting the inherent voltage-current characteristic of semiconductor diodes. In a fairly simple circuit, it was possible to approach the desired characteristic quite closely [Fig. 18-16]. It was, however, necessary to keep the diodes at a constant temperature of 120 degrees C ±0.1 degree in a temperature-controlled oven to maintain the desired characteristic and to assure close tracking of the compressor and expandor. The quantizing noise performance obtained was quite close to the theoretical bogey [Fig. 18-17]. A 30-dB signal-to-quantizing noise was realized over a speech-level range of about 40 dB. Because of inevitable imperfections in gates, compandors, and coders/decoders, actual performance objectives were set a few decibels lower.

2.2.4 Uniform Coder and Decoder

The process of encoding consisted of generating reference currents proportional to the binary values represented by each digit of the code and, in effect, subtracting them (or attempting to subtract them) from the current generated by the sample, until the closest fit was found. In the experimental and first

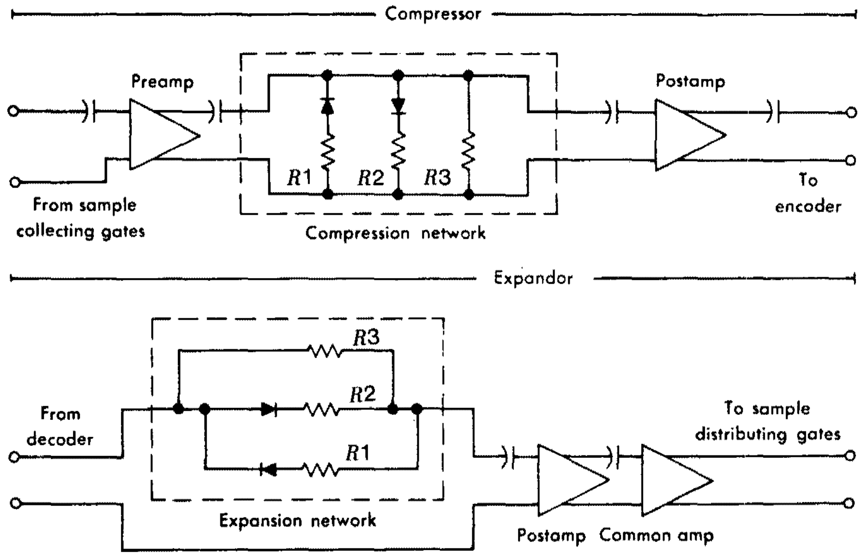


Fig. 18-16. Compressor and expandor realization for approximate logarithmic companding.

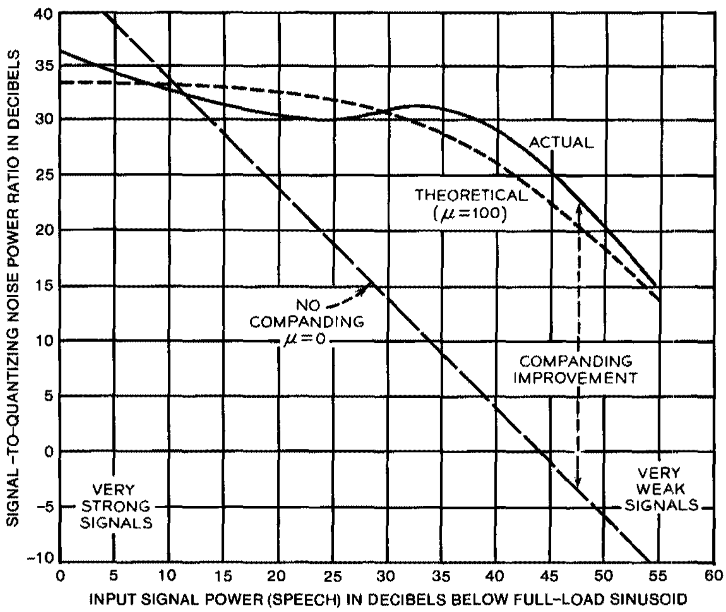


Fig. 18-17. Quantizing noise performance of the experimental compandor.

service system, this process was accomplished in a so-called network encoder [Fig. 18-18].

The input signal to the coder was biased from a highly stable DC source so that the positive and negative PAM samples were all of one polarity at the encoder. The pulses were then, in effect, measured against a scale with amplitudes from 1 to 127, with a zero-value input corresponding to level 64. (The discussion following assumes that the bias left all pulses positive in value and that the encoding process was to measure and transmit the binary value of the positive pulses. Actually, the bias left the pulses negative; but this makes the subtractions and code inversions quite confusing. The discussion is simpler with the assumption made and, clearly, any similarly consistent process of coding and decoding reconstructs the PAM pulses properly.)

Reference currents with the values 2^0 (1 unit), 2^1 (2 units), 2^2 (4 units), up to 2^6 (64 units) were generated from a stable voltage reference and precision resistors. The coder logic circuit first closed the switch to subtract 64 units of current from the sample. If the result was positive, as sensed by the summing amplifier, indicating a sample larger than 64 steps, a pulse (a one) was transmitted. In this case, the switch was left closed and the sample was reduced by 64 units. If the result of the first subtraction was negative, no pulse (a zero) was transmitted, indicating that the sample was smaller than the most significant digit of the transmitted code. This process was repeated for 32, 16, etc., units of current, until the complete sample code was determined.

Decoding was the inverse of coding. For each seven-digit code word, an amplitude sample was constructed. This was done by a digital-to-analog converter [Fig. 18-19]. The time position of each digit represented its weighting. The incoming pulses were read in sequence into a memory register and read out simultaneously through appropriate weighting resistors to generate the correct PAM pulse in a summing resistor. As noted above, circuit-speed prob-

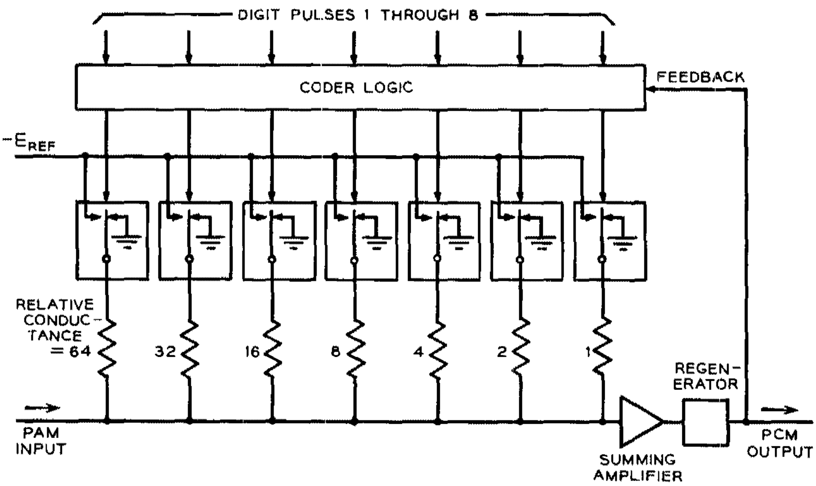


Fig. 18-18. Simplified block diagram of a network encoder.

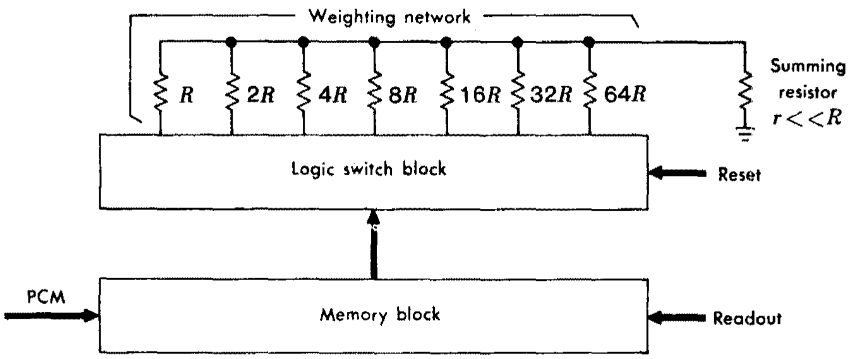


Fig. 18-19. Simplified block diagram of a decoder.

lems required duplicate compressors and coding networks at the transmitter. There were no similar timing restrictions at the receiver, so a single decoder, expandor, and amplifier common to all channels was sufficient.

2.2.5 Timing

Timing of the various functions in the terminal was under the control of a master crystal-controlled clock at the transmitting terminal. The sine-wave output of the master clock was reshaped to produce the steep-sided pulses required to turn the various digital functions in the terminal on and off at precisely the right time. A stable resonant circuit in each repeater and in the receiving terminal produced timing waves slaved to the master clock. A framing generator produced an alternating one-zero pattern in the 193rd pulse of each frame. This pattern, with a strong basic 4-kHz component, uniquely identified the beginning of a frame. Detection of the framing sequence at the receiver was essential to the process of steering the received signals to the proper channels [Fig. 18-20].

2.2.6 Terminal Realization

In real life, of course, things were never so simple as would appear from the previous description. Requirements imposed on the circuit elements to achieve the desired overall coding and decoding characteristics were quite stringent. Absolute uniformity of step size and infinitely sharp transitions between adjacent levels could only be approximated. Compressors and expandors did not track exactly. The channel multiplex gates did not completely open or close at precisely the right time and generated a range of DC pedestals depending on network differences. All these imperfections resulted in variations in the compandor advantage and degradation of the recovered signal. Many months of effort were expended before the designers were able to show that all channels were performing close to the objectives. (In the real world, complete success in complex developments is only approached asymptotically.) The experimental terminal was then deemed to have satisfactory performance.

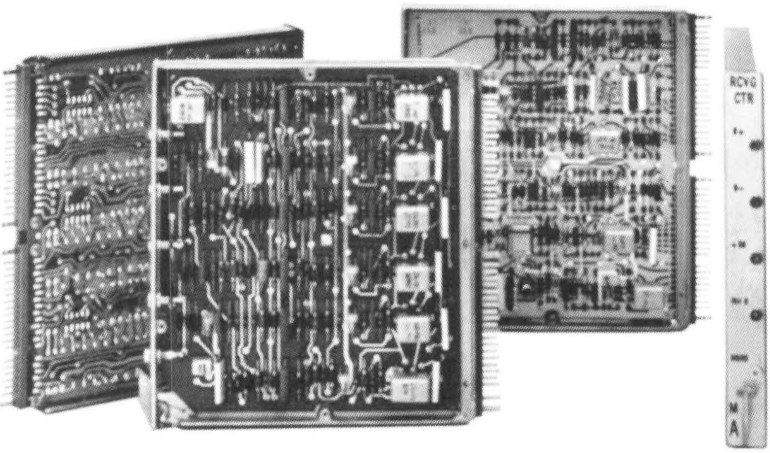


Fig. 18-20. Circuit boards used for the experimental PCM system timing. *Left to right*, digit generator, transmitting channel counter, master clock and framing generator, and receiving channel counter.

2.3 Design of the Repeatered Line

The terminal design was a large part, but only a part, of the job. It remained to transport the 1.5-megabit pulse train over the exchange pairs (typically No. 22 gauge) between the terminals and deliver them reasonably intact. Line attenuation at 1.5 megabits was high, and the plan was to provide regenerative repeaters at the normal locations for loading coils, that is, every 6000 feet or so. While the economic viability of short-haul carrier systems was largely determined by terminal costs, clearly, with repeaters spaced as close as 6000 feet (N carrier spacing was typically 8 miles) small, reliable, inexpensive repeaters were necessary.

One of the significant research contributions to the favorable climate for PCM in the mid-1950s was the demonstration of a simple two-transistor regenerative repeater [Fig. 18-21].[29] This circuit worked very well under laboratory conditions. The development group knew, of course, that the real cable plant would present some additional problems, and considerable effort was invested in designing a repeater to cope with the actual conditions. They had also expected from measurements and analysis of crosstalk that some cable pairs in close proximity would not be usable. Nevertheless, it came as something of a shock when, of the five pairs equipped for the first field experiment, only one worked really well; two worked, but were somewhat impaired; and two did not work well at all. Even though the pairs had been selected to present a varied spectrum of conditions, the error rates were much higher than expected. With the performance achieved, an enormous amount of measurement, pair selection, and plant rearrangement would be necessary to make even limited installation feasible. In a paper exercise, ten systems, each 20 miles long, were installed with all preinstallation qualification measurements eliminated. It was proposed that it would be acceptable to have one of the ten fail to operate and

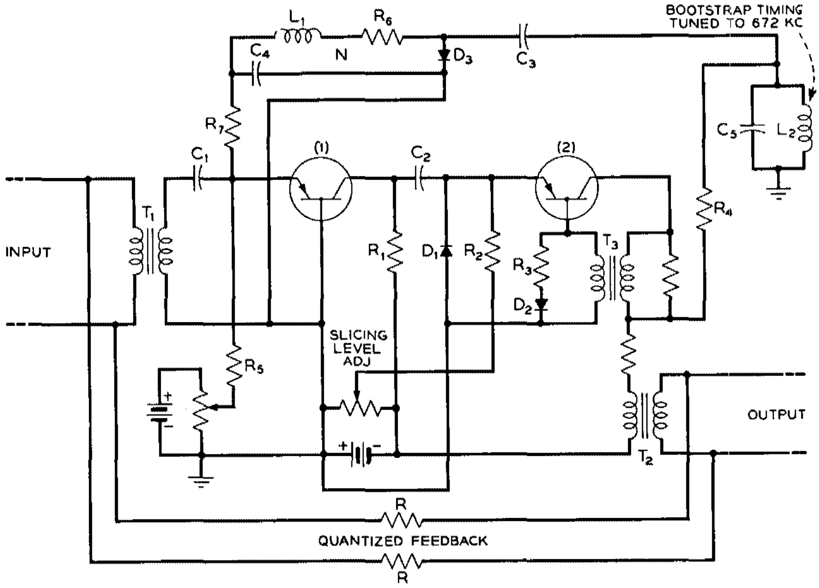


Fig. 18-21. Circuit diagram of the research regenerative repeater.

require reworking. Analysis indicated that, to meet this criterion, the trial repeater would somehow have to be less sensitive to crosstalk by 25 dB. Clearly, if PCM were to be practical, some major improvements were necessary.

2.3.1 The Environment

The exchange plant, where PCM carrier was to be applied, is characterized by several hostile features. Repeaters were to be placed in manholes that could fill up with water, the cable ducts might be adjacent to steam pipes, and high voltages could be induced in the pairs from nearby power and traction systems, especially under power fault conditions. Impulse noise, generated by switching systems in central offices, appeared on the pairs transmitting PCM signals. But the most severe problem was crosstalk from PCM systems on other pairs in the same cable. All the sources of interference, coupled with the high signal attenuation and dispersion, served to limit the number of systems that could be operated in the same cable.

2.3.2 Pulse Detection and Regeneration

For pulse regeneration, it was first necessary to decide whether or not a pulse was present in a time slot. To speak in terms of the frequency domain, the rectangular pulses contained a wide range of frequency components, which were propagated at different velocities over the pairs. At a repeater, the originally clean pulses were greatly attenuated and smeared over several time slots. Even after amplification and delay equalization, the pulse had a rounded form, with tails of significant amplitude into the adjacent time slots [Fig. 18-22]. The

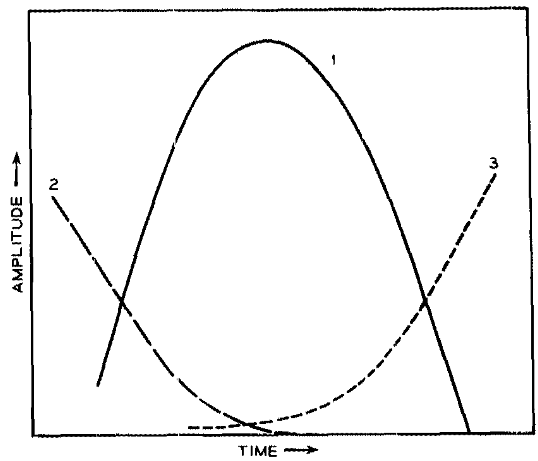


Fig. 18-22. Typical received signal pulse after amplification and equalization. (1) Signal from pulse in time slot. Interference from pulses in (2) preceding and (3) following time-slots.

detection process consisted of deciding at a particular time and with a reference voltage whether a pulse was present or not. The space enclosed between the desired signal and the interfering tails from pulses in adjacent slots was the working area for detection. The voltage and time crosshairs of the decision circuit had to intersect within this area to detect a pulse successfully. This area was referred to as an eye and the chart of all superimposed eyes that can result from various pulse patterns as an eye diagram [Fig. 18-23]. In a repeater, the timing for the decision point was obtained from a resonant clock circuit excited

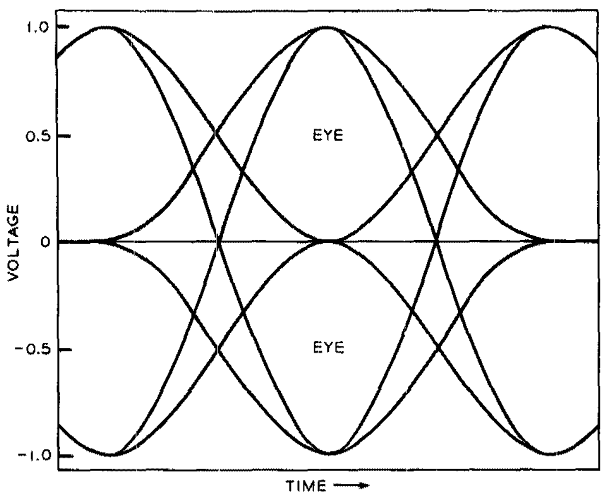


Fig. 18-23. Eye diagram.

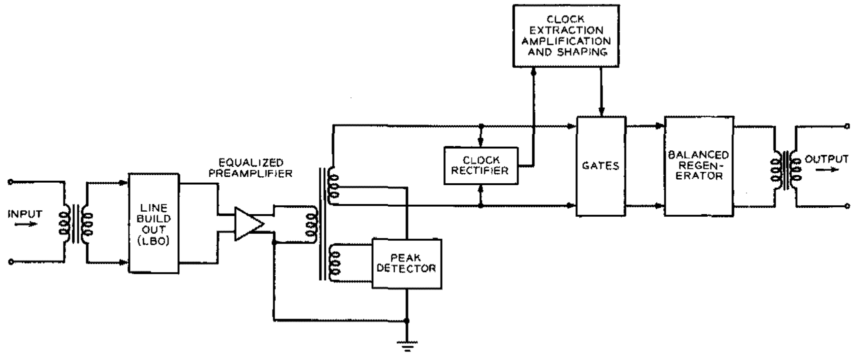


Fig. 18-24. PCM repeater configuration. The clock extraction circuit and peak detector formed the timing and amplitude reference for a pulse/no-pulse decision.

by the incoming pulse train and the reference amplitude by measuring the amplitude of the incoming signal in a peak detector and generating a voltage threshold designed to be one half the pulse height [Fig. 18-24].

The eye opening was reduced by frequency dispersion, thermal and device noise, and crosstalk. Interference coupled in from pulse trains on other pairs reduced the eye opening, both by increasing the amplitude of the interfering tails and reducing the amplitude of the desired signal. As the number of interfering signals increased and/or the coupling between the interfering and disturbed pair increased (lower crosstalk coupling loss), the eye opening was reduced in area [Fig. 18-25]. An early objective was to have an error rate no

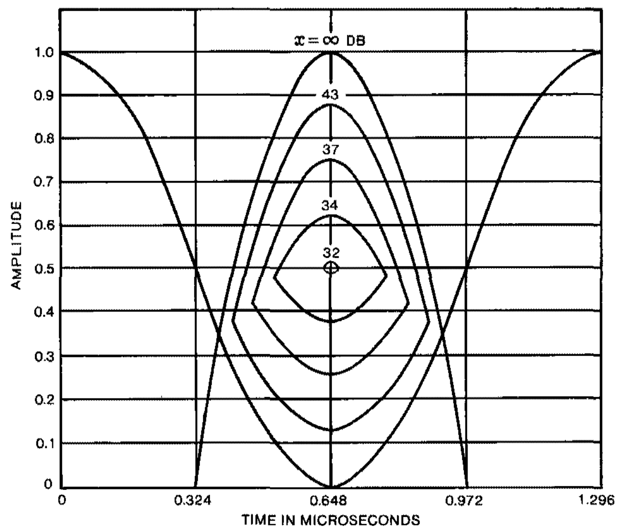


Fig. 18-25. Eye size for various values of near-end crosstalk-coupling loss.

worse than one error, in the most significant digit per channel per minute. This would produce one audible click per channel per minute and corresponded to a pulse error rate of about one error in 10^6 time slots.

As the lessons of the laboratory work and early field experiments were learned, several measures were taken to improve the ability to detect the presence or absence of a pulse in the presence of interference. A fixed reference voltage, equal to half the nominal pulse, would have been easiest to implement, but the received signal varied by 6 to 12 dB from location to location and from variations in the cable environment. The threshold was therefore made dependent on the average received-pulse amplitude. The clipping level in the clock path was also made proportional to the received signal level. Otherwise, the phase of the recovered clock signal would be dependent on the received amplitude. Fortunately, the same signal-dependent voltage, equal to about half the peak received signal, served for both the voltage threshold and clock circuit. The pulse train was clipped to prevent interference and noise from entering the tuned timing circuit in the absence of desired pulses. Timing uncertainties due to clock pulse width were minimized when very sharp, mid-slot, timing spikes were generated from the sinusoidal-recovered clock signals.

2.3.3 *Bipolar Transmission*

The largest improvement, however, in the ability to detect the presence or absence of a pulse was a change from unipolar to a bipolar signal format. This consisted of transmitting successive ones as alternating positive and negative pulses instead of the single polarity train originally used. The unipolar binary system was converted to a bipolar ternary system (zeros were still zero voltage, but ones could be either positive or negative). The change reduced the critical crosstalk frequency. Analysis confirms what a casual inspection of single polarity and bipolar pulse trains suggests: the principal frequency component in a unipolar train is at the pulse rate (1.5 megabits per second), but in bipolar transmission the principal energy is at a frequency only one-half that rate [Fig. 18-26].

The effect on crosstalk was profound. At 1.5 megabits, the line attenuation and the "1-percent-worst-case" crosstalk interference link was such that the disturbing signal had as much influence on the decision point as the desired signal. Crosstalk from additional pairs would further aggravate the situation [Fig. 18-27]. With the bipolar signal and the principal signal energy concentrated at 0.75 MHz, both the line loss and the critical crosstalk coupling were reduced; that is, the signal was stronger and the interference less. An improvement of 15 dB or more was realized, and the situation was changed from an impractical approach, for the conditions at that time, to the one that led to the practical production system.[30,31,32] (The phenomena of crosstalk in digital systems in cables and its influence on timing and amplitude decisions is extremely complex and defies precise analysis. Crosstalk from all pairs and at all frequencies affects the timing, for example. For engineering purposes, a description was developed at the single frequency most important to timing and on a statistical basis. The advantages of bipolar transmission, called *alternate mark inversion*, had been recognized in earlier work in England.)[33]

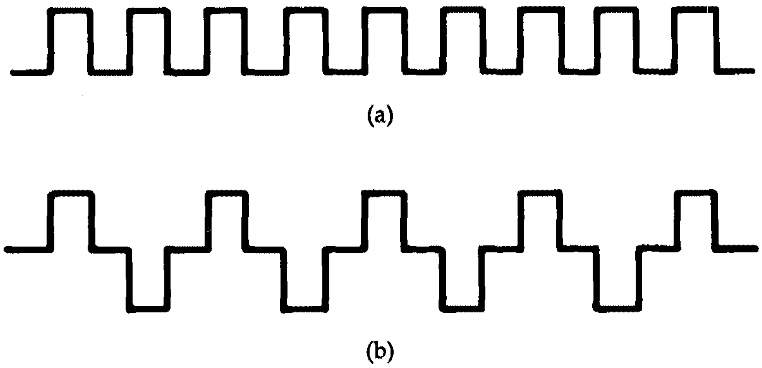


Fig. 18-26. Pulse transmission plans. (a) Unipolar. (b) Bipolar. Bipolar transmission reduced the critical crosstalk frequency to one-half the pulse rate.

Bipolar transmission had advantages in addition to the crucial improvement in crosstalk interference. The bipolar pulse train contained no energy at DC. This permitted economical simultaneous transmission of DC for powering the repeaters along with the signal pulse train. Finally, with bipolar transmission, in the event of an error, when a pulse was converted to a zero or vice versa, the alternating sign property was violated. This allowed convenient in-service monitoring of the line's integrity.

2.3.4 Repeater Realization

In addition to the improvements from automatic threshold control, precise control of timing, and bipolar transmission, repeaters were fitted with line-

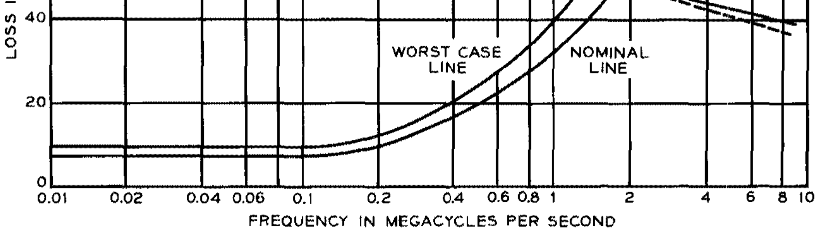


Fig. 18-27. Line loss and crosstalk loss. At 1.5 megabits, the interference signal from a worst case crosstalk link was about equal to the attenuated signal.

buildout (LBO) networks to bring short sections up to the standard loss. While the threshold control circuit automatically adjusted for line losses between roughly 27 and 35 dB at 772 kHz, larger variations in loss due to length and gauge differences were handled by the LBO circuits. A family of 12 networks was used to cover shorter sections and for No. 19 to No. 24 gauge cable. To reduce the effects of impulse noise generated by switching machines, the repeater spacing adjacent to an office was reduced to 3000 feet, half the normal distance.

All these developments, of course, considerably increased the complexity of the repeater, and the early two-transistor research design was abandoned at an early stage. The design that emerged from the field experiment and a subsequent prototype trial required seven diffused-base junction transistors and ten logic diodes for each one-way repeater. A two-way repeater consisted of two such circuits, with a total of 135 components in a package 1 by 3 by 5-3/4 inches [Fig. 18-28]. The repeaters required 1 watt from a power supply common to both directions.

2.3.5 Timing Jitter

The zero crossings of the timing wave extracted at each repeater were perturbed from several sources. These variations were termed *jitter*. The sources of jitter were systematic or nonsystematic, depending upon their relationship to the pulse pattern. Nonsystematic sources, such as noise, produced random phase variations that differed from repeater to repeater. Systematic sources, due to pulse-shape and pulse-pattern effects, were the same at each repeater. Pattern-dependent jitter could accumulate systematically in a chain of repeaters

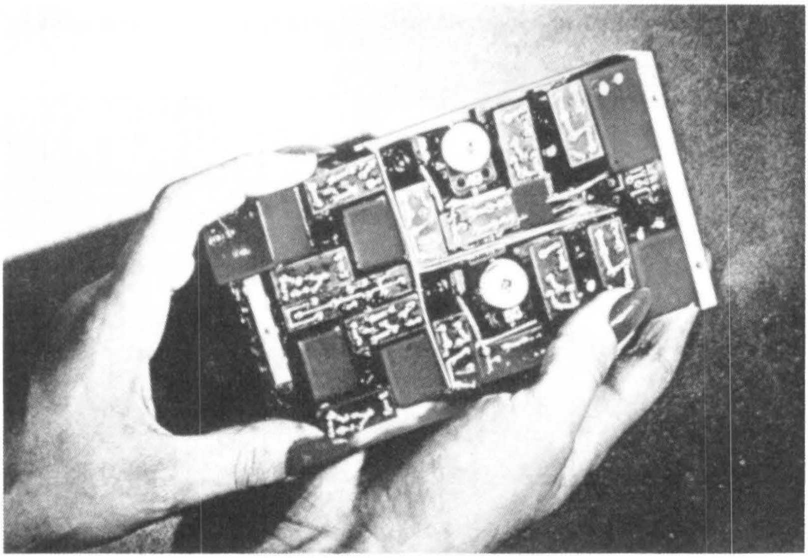


Fig. 18-28. The experimental two-way PCM repeater with cover removed, 1959.

and had to be limited to avoid errors caused by sampling the eye too far away from the peak opening. In addition, jitter in the recovered PAM signal translated into distortion in the reconstructed analog signal.

The subject of timing jitter was studied extensively in the experimental system, the following T1 development, and for many years after.[34,35,36] The problem at first appeared quite serious; but, as understanding improved, better methods for timing extraction were developed. A simple model for jitter accumulation was developed and used to predict and corroborate experimental results. The repeater was designed to minimize timing jitter, and the receiving terminal equipment further reduced the effect of jitter on the recovered analog signal to insignificant levels.[37] Jitter was reduced to a level where it was not a serious problem for speech. At the same time, a phase-locked loop-timing extraction and jitter reducer was designed to be used in later applications for signals such as video, where jitter requirements were more stringent.

2.3.6 *Fault Location*

When a repeatered line fails, the line must be removed from service and the fault isolated to the offending repeater section for subsequent off-line repair. It was obviously desirable to minimize useless trips to manholes by pinpointing the source of the problem by tests at the terminals, preferably from one end. In addition, it was useful to be able to identify a marginal repeater by appropriate testing before an actual failure occurred. A system was developed to determine the health of the line by taking advantage of the properties of the bipolar system in which the average DC component of the signal was zero. Purposely introducing a DC unbalance shifted the decision threshold for pulse or no pulse in a repeater. In this manner, the repeater margin against errors could be determined. In the bipolar system, a DC unbalance was created by inserting bursts of deliberate bipolar violations at an audio rate, causing the repeater output to have a tone at the same rate. At each repeater location, the output was coupled to a narrow audio-frequency band-pass filter whose center frequency was unique to that location. The outputs of these filters were transmitted back to a maintenance location over a common fault-location wire pair. The response to the DC offset could be tested at each repeater in turn by sequencing through the assigned frequencies of the violation bursts. Absence of a return audio signal at a particular frequency was indicative of a failure at a location closer to the terminal than the repeater corresponding to that frequency. Further, if a repeater was in a marginal condition, increasing the density of the violations would increase the offset and cause a repeater to make errors at a stress level lower than that for a normal repeater.

2.4 Field Experiments, Prototype Trials, and Final Development

An initial field experiment was held on several cable pairs between Summit and South Orange, New Jersey in 1956. The results were rather devastating, as they revealed substantial inadequacies of the experimental equipment and the unlikelihood that unipolar transmission could be made to work at all. The experiment, however, also pointed the way to the improvements needed and led to the adoption of bipolar transmission at an early stage. As changes were made, they were tested in the field as well as in the laboratory. All aspects of the repeatered line were again tested in the Summit–South Orange test bed in 1959 to confirm the performance of the bipolar system and the improved repeaters.

Inevitably, as problems were overcome with more complex circuits, costs crept up until the cost advantage over N carrier, initially projected at 50 percent, had almost vanished. (N system developers, of course, were reducing the cost of their system and trying to extend its application into the exchange area during the same period). Strenuous cost-reduction efforts on the digital system, however, resulted in a fragile, but presumably demonstrable, cost advantage of about 15 percent for the PCM system.

A difficult cost contest between an established system, well shaken down and in production, and a new one still in development occurred again and again from 1925 to 1975. It was not easy to prove in a coaxial system over carrier on pairs or early microwave radio over coaxial cable, although television transmission capability helped in both cases. Management wisdom consisted less in accepting hard cost estimates on somewhat soft concepts than in seeing where a particular technology was headed, and which developments opened avenues unavailable to the existing systems.

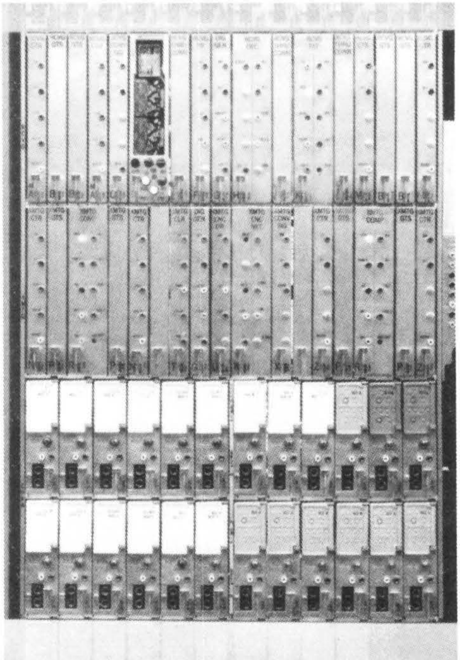
In 1961 and 1962, a prototype commercial PCM system was manufactured by Western Electric and tested by Bell Laboratories with the participation of New Jersey Bell between Newark and Passaic, New Jersey. Extensive cable-pair measurements were made to validate engineering rules for system installation.[38] A full commercial-service trial was conducted on the prototype. Bell Laboratories established a plant school to train telephone-company craft people to operate and maintain this new and unfamiliar system. The ability of the craft personnel to operate and maintain the system during the early commercial service phase was evidence of the success of this endeavor.

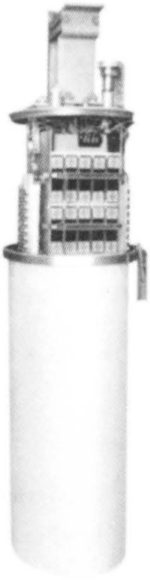
A trial on aerial cable was conducted in Ohio in 1962. This trial was essential to show that the system could operate in the presence of lightning surges and static interference and that pole-mounted repeater housings could survive the elements. Extreme cable-temperature variations in the aerial trial dictated some minor changes in engineering rules; but, with these established, the system was approved for aerial cable.

2.4.1 Production and Initial Installation

While the field tests were still under way, a group was established at the Bell Laboratories Merrimack Valley, Massachusetts branch to complete the final design for manufacture. This group, working closely with their Western Electric counterparts at the same location, refined the design and settled the innumerable details necessary for a commercially manufacturable and service product. Plug-in circuit packs, equipment bays, and manhole-apparatus cases were designed with production, service, and maintenance requirements as primary considerations. The Passaic–Newark prototype trials were conducted with components manufactured by Western Electric from these designs [Fig. 18-29].[39,40,41,42]

The entire system was called the T1 carrier system and the 1.5-megabit lines of the original design continued to be called T1 lines. The original 24-channel terminal was designated the D1 channel bank. The first T1 system from regular





(a) (b)

(c)

Fig. 18-29. The T1 carrier system. (a) Production terminal common equipment and channel units in place for one 24-channel D1 bank. (b) T1 line-repeater apparatus case with cover partly removed. (c) Apparatus cases, each containing 25 two-way T1 repeaters installed in a manhole.

production was installed and turned up for service between the Dearborn office in downtown Chicago and Skokie, Illinois, about 13.6 miles away, in the fall of 1962 [Fig. 18-30].

An intercompany task force (Bell Laboratories, AT&T, Western Electric, and Illinois Bell) was formed to assure a smooth transfer. Installation and maintenance procedures were honed and excellent performance achieved. Concrete evidence of customer satisfaction was given when Illinois Bell ordered more than 300 additional systems for installation in 1963. Acceptance throughout the Bell System was quick, and production took off rapidly. By 1967, over 16,000 systems and one-third of a million T1 carrier lines had been delivered, providing 4,600,000 voice-circuit miles of capacity [Fig. 18-31]. It was estimated that the operating companies had saved $80 million on an investment of $280 million compared to the next best option over the four years of operation. The development to first production had required less than 200 staff years, and development cost was about $5 million. By 1981, T1 carrier, with several successive generations of terminal and repeater designs, was providing over 100,000,000 voice-circuit miles.

2.5 Impact of T1 Carrier on Digital Communications Evolution

It was recognized in the mid-1950s that what was to become the T1 carrier system was the harbinger of a new digital era. (There were prophets at even

Fig. 18-30. The first T1 carrier terminals installed in the Illinois Bell Dearborn office in Chicago, 1962.

earlier dates and skeptics until the 1960s. But when T carrier installations soared, few doubted that it was only a matter of time until digital transmission would spread to all parts of the plant.) The growth of T carrier far exceeded initial expectations. The introduction of new integrated-circuit designs resulted in smaller size, decreased power, improved performance, and lower costs. The cost of a repeater dropped by a factor of three in about a decade—in current dollars, during a period of steep inflation! By 1983, more than half of all exchange trunks were digital. The synergy of digital transmission and digital switching paved the way for the introduction of the digital 4ESS* electronic switch in 1976. More than three-quarters of all toll-connecting trunks were digital by 1984, and virtually all of these trunks were expected to be digital before 1990.

As of 1984, a similar relationship, between local digital switching and digital loop systems, based on T carrier techniques, appeared to be driving subscriber loops in the digital direction. Digital microwave radio systems were being installed in substantial numbers for distances up to about 300 miles. Long-haul systems remained stubbornly analog (but with extensive digital overlays) for

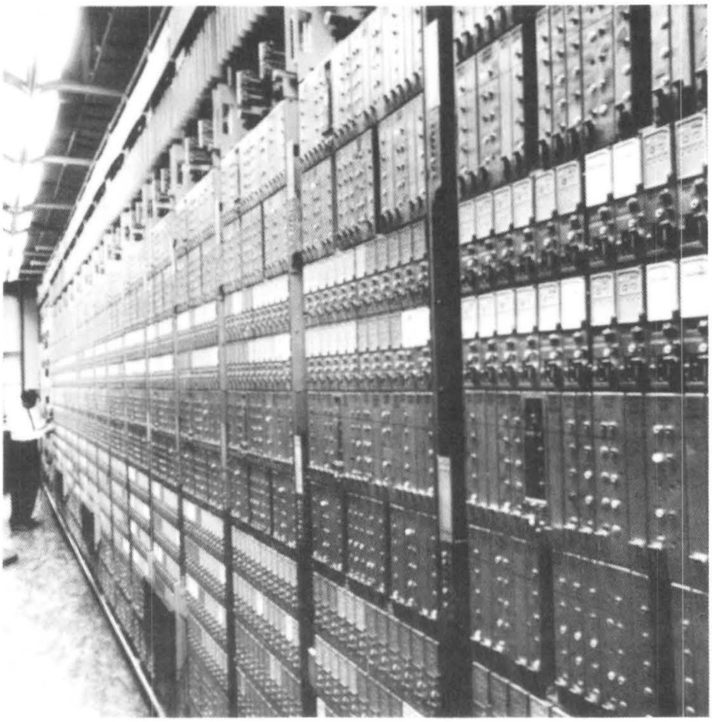


Fig. 18-31. Terminals for more than a thousand T1 carrier telephone circuits installed in a Passaic, New Jersey central office, 1967.

---

* Trademark of AT&T Technologies, Inc.

many years because of low analog-system costs. (See Chapter 9.) But in the mid-1980s, the production and installation of large numbers of digital systems on optical fibers seemed at last to have broken the long-haul barrier. It was projected that by the end of the 1980s coast-to-coast pervasive digital connectivity would be achieved. An enormous advantage of digital transmission is that any signal can be digitally encoded and transmitted without interaction with other signals or with degradation of signal-to-noise with distance. The T1 carrier developers could not, of course, envision the explosive growth of fiber optics, but they did recognize the unification of services implicit in the coded digital stream.[43]

REFERENCES

1. A. L. Cauchy, "Analyze Mathématique," *Comptes Rendus* **12** (January-June 1841), p. 283.
2. E. T. Whittaker, "Expansions of the Interpolation Theory," *Proc. Royal Soc. Edinburgh* **18** (1915), pp. 181-194.
3. K. W. Cattermole, *Principles of Pulse Code Modulation* (New York: American Elsevier Publishing Company, 1969), p. 19.
4. J. R. Carson, "Method of Multiplex Telephony Based on a Division of Time," unpublished memorandum (1920).
5. H. Nyquist, "Certain Topics in Telegraph Transmission Theory," *Trans. Amer. Inst. Elec. Eng.* **41** (April 1928), pp. 617-644.
6. W. R. Bennett, "Time Division Multiplex Systems," *Bell Syst. Tech. J.* **20** (April 1941), pp. 199-221.
7. J. M. Fraser, D. B. Bullock, and N. G. Long, "Over-All Characteristics of a TASI System," *Bell Syst. Tech. J.* **41** (July 1962), pp. 1439-1454.
8. R. M. Averill and R. C. Stone, "No. 101 ESS: The Time-Division Switch Unit," *Bell Lab. Rec.* **41** (December 1963), pp. 425-431.
9. A. M. Gerrish, R. V. Lohmiller, and G. E. Saltus, "The DIMENSION PBX: A New All-Electronic Customer Switching System," *Bell Lab. Rec.* **53** (September 1975), pp. 316-323.
10. M. R. Aaron, "PCM Transmission in the Exchange Plant," *Bell Syst. Tech. J.* **41** (January 1962), pp. 99-141.
11. B. M. Oliver, J. R. Pierce, and C. E. Shannon, "Philosophy of PCM," *Proc. IRE* **36** (November 1948), pp. 1324-1331.
12. P. M. Rainey, U.S. Patent No. 1,608,527; filed July 20, 1921; issued November 30, 1926.
13. A. H. Reeves, French Patent No. 853,183; filed October 3, 1938; issued October 23, 1939.
14. A. H. Reeves, U.S. Patent No. 2,272,070; filed November 22, 1939; issued February 3, 1942.
15. See reference 3, pp. 1-27.
16. G. S. Vernam, "Cipher Printing Telegraph Systems for Secret Wire and Radio Telegraphic Communications," *Trans. Amer. Inst. Elec. Eng.* **45** (February 1926), pp. 109-115.
17. H. Dudley, "Remaking Speech," *J. Acoust. Soc. Amer.* **11** (October 1939), pp. 169-177.
18. W. R. Bennett, "Secret Telephony as a Historical Example of Spread-Spectrum Communications," *IEEE Trans. Commun.* **COM-31** (January 1983), pp. 98-104.
19. See reference 18, p. 103.
20. H. S. Black and J. O. Edson, "Pulse Code Modulation," *Trans. Amer. Inst. Elec. Eng.* **66** (1947), pp. 895-899.
21. W. M. Goodall, "Telephony by Pulse Code Modulation," *Bell Syst. Tech. J.* **26** (July 1947), pp. 395-409.
22. L. A. Meacham and E. Peterson, "An Experimental Multichannel Pulse Code Modulation System of Toll Quality," *Bell Syst. Tech. J.* **27** (January 1948), pp. 1-43.

23. R. W. Sears, "Electron Beam Deflection Tube for Pulse Code Modulation," *Bell Syst. Tech. J.* **27** (January 1948), pp. 44-57.
24. C. G. Davis, "An Experimental Pulse Code Modulation System for Short-Haul Trunks," *Bell Syst. Tech. J.* **41** (January 1962), pp. 1-24.
25. B. Smith, "Instantaneous Companding of Quantized Signals," *Bell Syst. Tech. J.* **36** (May 1957), pp. 653-709.
26. R. H. Shennum and J. R. Gray, "Performance Limitations of a Practical PCM Terminal," *Bell Syst. Tech. J.* **41** (January 1962), pp. 143-171.
27. C. A. Desoer, "A Network Containing a Periodically Operated Switch Solved by Successive Approximations," *Bell Syst. Tech. J.* **36** (November 1957), pp. 1403-1428.
28. H. Mann, H. M. Straube, and C. P. Villars, "A Companded Coder for an Experimental PCM Terminal," *Bell Syst. Tech. J.* **41** (January 1962), pp. 173-226.
29. L. R. Wrathall, "Transistorized Binary Pulse Regenerator," *Bell Syst. Tech. J.* **35** (September 1956), pp. 1058-1084.
30. See reference 10.
31. F. T. Andrews, Jr., U.S. Patent No. 2,996,578; filed January 19, 1959; issued August 15, 1961.
32. J. S. Mayo, "A Bipolar Repeater for Pulse Code Modulation Signals," *Bell Syst. Tech. J.* **41** (January 1962), pp. 25-97.
33. R. H. Barker, British Patent No. 706,687; filed June 20, 1951; issued April 7, 1954.
34. M. R. Aaron and J. R. Gray, "Probability Distributions for the Phase Jitter in Self-Timed Reconstructive Repeaters," *Bell Syst. Tech. J.* **41** (March 1962), pp. 503-558.
35. C. J. Byrne, B. J. Karafin, and D. B. Robinson, Jr., "Systematic Jitter in a Chain of Digital Regenerators," *Bell Syst. Tech. J.* **42** (November 1963), pp. 2679-2714.
36. J. M. Manley, "The Generation and Accumulation of Timing Noise in PCM Systems—An Experimental and Theoretical Study," *Bell Syst. Tech. J.* **48** (March 1969), pp. 541-613.
37. See reference 32.
38. H. Cravis and T. V. Crater, "Engineering of T1 Carrier System Repeatered Lines," *Bell Syst. Tech. J.* **42** (March 1963), pp. 431-486.
39. A. Hamori and N. E. Lentz, "T1 Carrier System Description," Paper No. CP 62-1469, presented at American Institute of Electrical Engineers Conference (1962).
40. D. F. Hoth, "The T1 Carrier System," *Bell Lab. Rec.* **40** (November 1962), pp. 358-363.
41. H. T. King and D. B. Penick, "T1 Carrier System Terminals," *Bell Lab. Rec.* **41** (June 1963), pp. 223-228.
42. K. E. Fultz and D. B. Penick, "The T1 Carrier System," *Bell Syst. Tech. J.* **44** (September 1965), pp. 1405-1451.
43. M. R. Aaron, "Digital Communications—The Silent (R)evolution?" *IEEE Commun. Mag.* **17** (January 1979), p. 18.

Chapter 19

Further Development on Cable Pairs

I. 1.5 MEGABIT SYSTEMS

1.1 T1 Line Development

Following the successful initial development of T1 carrier, a series of further developments of the line repeater followed, capitalizing on new technology as it became available. By 1970, a new repeater was designed and placed on trial using both silicon and hybrid integrated circuits. Like the old versions (coded 201 and 205), the new repeaters were made in two types (coded 208 and 209) to provide the different surge protection needed in buried and aerial cables. In addition to much smaller size, the new repeaters had several improved features and operating characteristics. The new design was in substantial production by 1972.

With the first production-repeater design, loss measurements were required on all cable pairs within the pair group to be used in a cable section. Based upon the average pair loss, a line-buildout (LBO) network was selected from among 12 designs available to bring the loss up to a standard value. The same "length" LBO network was then used for all pairs in that section. As a result, equalization was only approximate. Furthermore, even though this method appeared to be simple to administer, in practice, a significant number of wrong LBO networks were inserted, resulting in marginal line performance.

In the new repeater, an automatic line-buildout network (ALBO) inserted the correct buildout loss, depending only on the received average signal power. This not only compensated accurately for any difference in the loss of each pair from the average, but offset changes due to temperature more effectively than adjustments of the threshold level (which was still retained). The stocking and error-prone selection of separate LBOs was eliminated [Fig. 19-1].

The old repeater also had two undesirable characteristics that sometimes caused difficulty in operation and maintenance. In the absence of an input pulse train, such as could occur during installation prior to service or because of an upstream line failure with loss of signal, the repeater was apt to oscillate. The uncontrolled oscillation then interfered with the ability to monitor the performance of that line and troubleshoot on the other lines that shared the

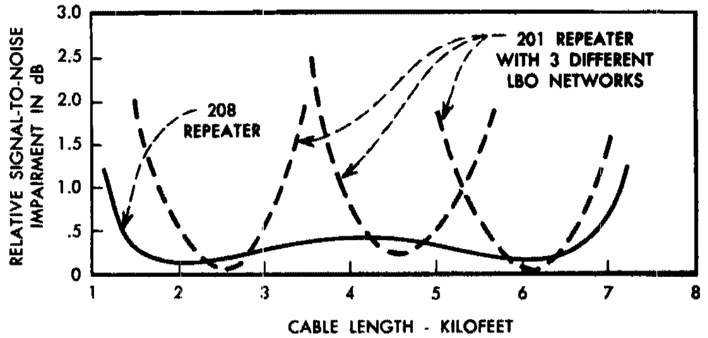


Fig. 19-1. Performance of the automatic line buildout in T1 carrier compared to selected buildout networks.

common fault-location equipment. The second problem involved the sensitivity of the original repeater to cable imperfections, such as splices and bridge taps located close to the repeater output. (The taps should have been, but were not always, removed.) At times, this sensitivity resulted in marginal repeatered-line performance. The new repeaters eliminated both undesirable characteristics, i.e., they did not oscillate during the absence of an input signal nor were they sensitive to cable imperfections. The result was that installation, testing, and maintenance were made significantly easier, and the new repeater operated with more margin than the old.[1]

1.1.1 Space Saving and Heat Dissipation in Manholes

The new repeater designs were much smaller than the original design, which helped to conserve space in increasingly congested metropolitan manholes. The size reduction was made possible by realizing most of the repeater in hybrid integrated circuits (HICs). Four silicon circuits were used, one each for the ALBO, the preamplifier, the timing circuit, and the regenerator. Two of the hybrid circuits, one for each direction of transmission, were mounted on a printed-wiring board along with discrete components, such as transformers and large-value bypass capacitors, to form the new repeater [Fig. 19-2].[2]

With the substantial reduction in the size of the new repeater and without an equivalent reduction in power dissipation, the power density was higher. As a result, it was necessary to design the new repeater to operate over an ambient temperature range from −40 degrees C to +85 degrees C, which was substantially greater than the operating range of the earlier repeaters. This was made possible because of the close matching and similar thermal-tracking characteristics of the silicon integrated circuits, and the excellent temperature stability of the thin-film resistors and ceramic capacitors. In addition, the repeater and apparatus case were designed to provide high thermal conductivity to the outside environment.

T carrier equipment located in offices, while differing in its electronics from earlier carrier systems, was not fundamentally different in its environment or

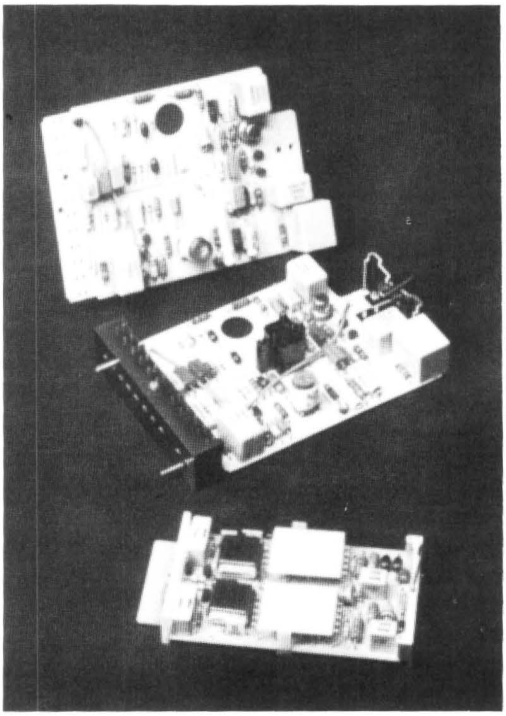


Fig. 19-2. *Top*, the 1962 T1 line regenerative repeater separated into two parts and, *bottom*, the design realized with hybrid integrated circuits in 1969.

packaging requirements. Indeed, the lower power dissipation of transistors made the equipment easier to house. Over the years, T carrier office equipment shared in the same general advances in power reduction, compactness, and ease of manufacture that characterized other electronic designs. The most difficult physical design problems were encountered in the outside plant area, especially in manholes, where, in the early deployment, about 80 percent of the T1 repeaters were located.

Manhole use meant that the repeater housing could be exposed to a variety of pollutants and continuous submersion. Thus, all apparatus cases had to be watertight and materials had to be highly resistant to corrosion. In order to ensure that water would be excluded, repeater cases, like the trunk cables to which they were spliced, were pressurized by compressors located in the central offices. Most often the cables themselves were the conduits for pressurization, although in some instances, especially in larger cities with numerous cables of various pneumatic integrity, there was an air-pipe network sharing the duct runs with the cables. Out of long experience in this environment, designers had evolved a selection of materials and designs for apparatus cases, especially for the housing of loading coils. The first T1 carrier repeater cases were based on that technology. In the early days of digital systems, it was common to convert cable pairs in existing trunk cables from voice to T carrier by removing a loading-coil case and replacing it with a repeater case of similar external appearance. A repeater case of the 1960s consisted of a repeater card cage mounted on a heavy cast-iron base with a cylindrical galvanized steel cover [Fig. 19-3]. The cover had to be removable, since repeaters could fail and need to be replaced. The lateral space required to slide the cover off about doubled the required wall space. The repeater case, like its loading-coil predecessor, was connected to the trunk cable by a stub cable.

By the early 1970s, the very success of T carrier was creating new problems. Instead of simply adding a few carrier repeater cases to what was primarily a voice-frequency cable, operating companies were often installing cables primarily for T carrier. Under cable-pair utilization rules in effect in the 1960s, there could be up to eight repeater cases (200 repeaters) on a 900-pair cable. Eight cases could generally be placed and spliced in a standard-size manhole without much difficulty. But as T carrier became more popular and its economic advantage relative to voice-frequency (VF) trunks became more pronounced, special screened cables were developed that were optimized for carrier. (In screened cables, the pairs in a single sheath were separated into two shielded, or screened, compartments—the pairs for one direction of transmission in one and the opposite direction in the other. The word *screened* (a British term for

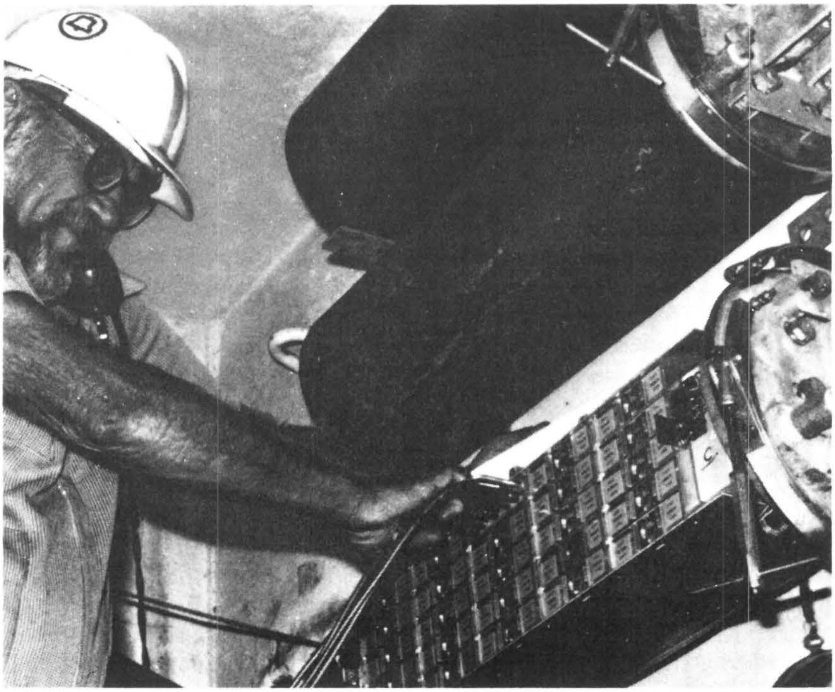


Fig. 19-3. T1 repeater apparatus cases with access equipment, 1963.

shielded) was used because the term *shielded* cable had long been used for cables in which all internal pairs were shielded from external interference, but not from each other.) In screened cables, virtually every pair was available for carrier, and a 900-pair cable could require 18 cases. This number presented problems in both manhole space and splicing arrangements.

In response to these problems, and to take full advantage of the smaller repeaters, a new apparatus case, the 475 type, was designed with first production in 1972. The new cases were not only much smaller but were provided with three small access covers that eliminated the need for the large clearance space required with the old case [Fig. 19-4]. Two of the modules held ten repeaters each, while the third held five repeaters and the fault-location networks. With the new design, the number of cases that could be housed in a given manhole volume was tripled. In the 475 design, all repeaters rested directly against the inner surface of the case's cast-iron well. This placement provided highly efficient cooling, a necessity since the new repeaters dissipated the same power as their forebears in a much smaller volume.

1.1.2 Low-Power Repeaters

While the 208/209 type repeaters represented a major technical advance and significantly improved performance, an urgent need to reduce power dissipation remained. Reduced power would not only lower repeater operating temperatures and thereby improve life and reliability, but it would increase

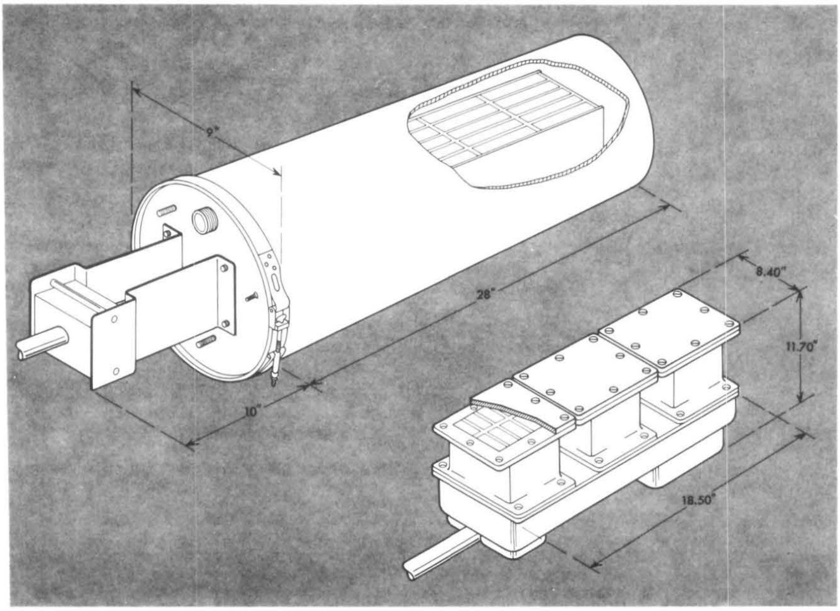


Fig. 19-4. T1 carrier repeater apparatus cases. The apparatus case designed for the 1970 integrated circuit T carrier repeater, *bottom*, compared to the original, *top*.

the length of spans permissible between powering points. Last, but by no means least, it would reduce the cost of primary power, a factor of abruptly increased significance following the oil embargo of the mid-1970s.

A major reduction in power needs was achieved in the early 1970s by Bell Laboratories with a new device technology called complementary bipolar integrated circuitry (CBIC). CBIC allowed high-quality, matched n-p-n and p-n-p transistors to coexist on a single silicon chip. (See another volume in this series, *Electronics Technology (1925-1975)*, Chapter 2.) The complementary chips permitted another redesign of the T1 repeater, completed in 1976, that reduced consumption from 1.5 w to 0.43 w. With the low-power design, the spacing between powering offices could be extended to as much as 37 miles. This flexibility was of increasing importance as T carrier expanded more and more into the suburban and rural areas.[3]

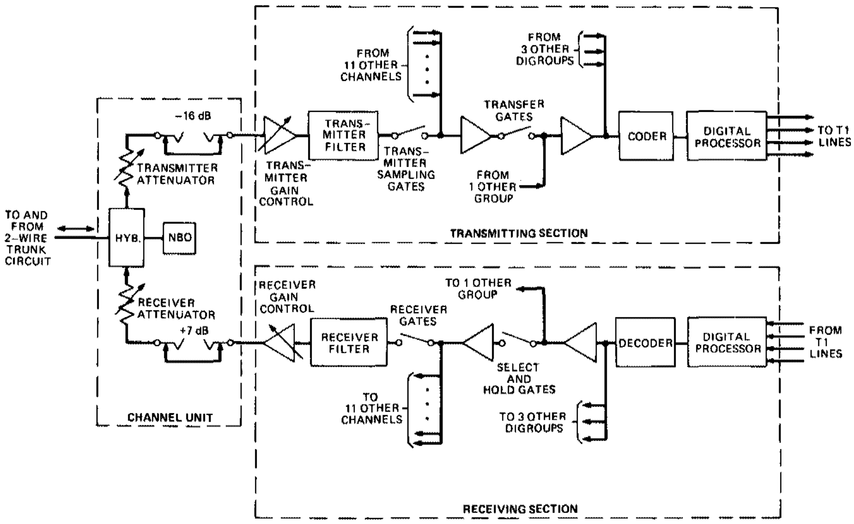
Although the repeater designers took advantage of each advance in device technology by repackaging to exploit the new characteristics, as each new generation of repeater design appeared, compatibility was maintained by designing adapters that permitted placing the new repeaters in earlier-generation apparatus cases.

1.2 The D2 Channel Bank

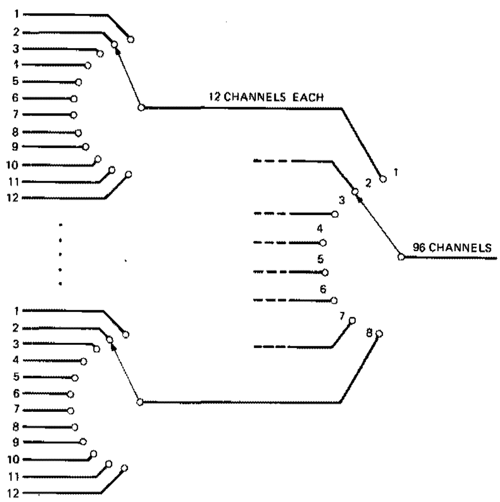
The T carrier system was viewed from the start as the first step toward a nationwide digital network. But the first system design had been tailored to the performance requirements and features of exchange-area trunks. The popularity and phenomenal growth of T carrier in the telephone companies pointed to the next step—a digital channel bank that would meet intertoll performance objectives. Attention was directed to the design of the terminal channel bank because digital lines and regenerative repeaters do not affect voice-frequency parameters, such as frequency response and noise, provided they are working properly; they affect error performance only.

The D2 channel bank was developed in the Holmdel, New Jersey digital transmission laboratory from 1968 to 1970. It was expected that T1 lines would serve for short toll connections, and several other higher-capacity digital lines were in prospect for longer routes. The new channel bank processed up to 96 voice-frequency signals into pulse-code-modulation (PCM) digital form for transmission, thus sharing the cost of some of the common circuits over four times as many channels as the D1 channel bank. Functionally, however, the D2 channel bank was equivalent to four 24-channel terminals; i.e., its output was four 1.544-megabit streams, which could be connected to four 24-channel T1 lines.[4]

The basic functions of the D2 channel bank were similar to those of D1 [Fig. 19-5]. Each channel was limited to 4 kHz by a low-pass filter and sampled 8000 times a second. The samples of varying amplitude were then time multiplexed in two stages. In the first stage, channels in 12-channel groups were sampled in sequence; in the second stage, these samples were interleaved (time multiplexed) with samples from seven other 12-channel groups. Device speed had advanced to the point where a single coder or decoder could handle all 96 channels. Following the coder, a digital processor separated out the four



(a)



(b)

Fig. 19-5. The D2 channel bank. (a) Block diagram of D2 channel bank showing the stages of time-division multiplexing into groups (12 channels), digroups (24 channels), and 96-channel signals for coding. (b) D2 multiplexing plan.

1.544-megabit streams, each stream being made up of the coded samples from two of the original groups. The inverse process of assembling four T1 incoming-signal streams into a single stream for decoding was done in a similar, but somewhat more complex, processor in front of the decoder. By starting with 12 channel groups, and combining two of them for each 24-channel digroup output, the advantage of a reasonable recovery interval for each channel-sampling circuit was retained. Throughout the sampling and multiplexing process, the equipment was arranged so that the bank could be equipped for one 24-channel digroup at a time and so that failure or maintenance activity in one would not disturb the operation of the others.

1.2.1 New Standards and Compatibility

Beyond this set of functions, however, the D2 channel bank designers were faced with some difficult decisions. In toll connections, many trunks could be connected in tandem by switching machines that, at that time and for many years to follow, could handle signals only in analog form. Even without switching, there could be several digital and analog transmission links in tandem. In such situations, repeated analog-to-digital and digital-to-analog conversions would be required, and quantizing noise would accumulate. The noise from each conversion, therefore, had to be held to a very low value. Also, in the future digital network envisioned, a digital trunk could be switched to another digital trunk without the conversions. (A digital time-division toll switch was in the early stages of planning during the D2 channel bank development.) This meant that it had to be possible to reconstruct a single channel signal converted into digital form by one channel bank back to analog in any other bank. A high degree of standardization and uniformity was therefore required.

The D1 channel bank, as designed around 1960, was not adequate, from either the standpoint of the signal-to-noise ratio or uniformity. A decision was therefore made to design a channel bank to standards that would be good in all applications for the foreseeable future, but that consequently would not be compatible with the existing D1 design.[5] It would not be possible to have D1 and D2 channel banks at opposite ends of a T1 line. The number of channels, sampling rate, and output bit rate were chosen to match those of the D1 channel bank, thus permitting the use of the T1 line to interconnect D2 channel banks. The most important differences were in the number of digits per coded sample and in the companding law. These differences, in turn, caused other attributes, such as the signaling and framing format, to be different. Since the 8-kHz sampling rate was retained, it was expected to be possible to interconnect D2 channel banks with other channel banks using the same rate by digital processing.

1.2.2 Eight-Digit Coding

For the signal-to-noise requirement to be met, all eight digits were used to code speech samples, compared to the seven-digit code used in the D1 channel bank. Instead of every eighth bit being dedicated to signaling, as in D1, the eighth bit was used for signaling in only one frame out of six. This signaling rate was adequate for the D2 bank to operate with all the signaling systems handled by D1, while maintaining performance close to that of full eight-digit PCM. The actual format in the D2 bank was to code each sample into eight-digit PCM in five out of six frames and into seven-digit PCM in the sixth frame, when the eighth bit was used for signaling. The resultant quantization noise was about 4 dB better than seven-digit coding and 2 dB above that of full eight-digit PCM. It was anticipated that even this degradation would be eliminated in time with the adoption of common-channel interoffice signaling (CCIS). In CCIS, which was later widely installed as anticipated, interoffice control signaling was over dedicated data channels, completely removing signaling requirements from transmission terminals. The D2 channel bank format allowed for both the current and anticipated CCIS methods.

The frames that were to convey signaling were identified by coding the sequence of the ones and zeros in the 193rd digroup framing bit [Fig. 19-6]. Since for some switching systems, two signaling paths were required, it was necessary to identify a superframe of 12 frames in which the 6th and 12th were used for the two signaling paths.[6]

1.2.3 Companding and Coding

The companding in the original D1 channel bank, based on the nonlinear properties of diodes (see Chapter 18, Section 2.2.3) was not adequate for the tight requirements placed on D2. In the D1 compandor, the companding characteristic had been approximated by a logarithmic expression whose steepness was defined by a parameter, μ (in D1 μ had a value of 100). At the time of the D2 development, and earlier, various other approximations had been proposed. The one proposed by the Bell System was based on the μ-law formula with a μ of 255.[7] The Europeans, led by the British, favored a slightly different approximation following the so-called A-law.[8] Both approximations became linear for very small signals, as a logarithmic function cannot be carried to zero, and both could be represented by linear piecewise segments [Fig. 19-7].[9,10] Furthermore, both were digitally linearizable. This was achieved by aving the step size for each segment related to the adjacent segments by a power of two. With this relationship, and with fairly simple digital processes,

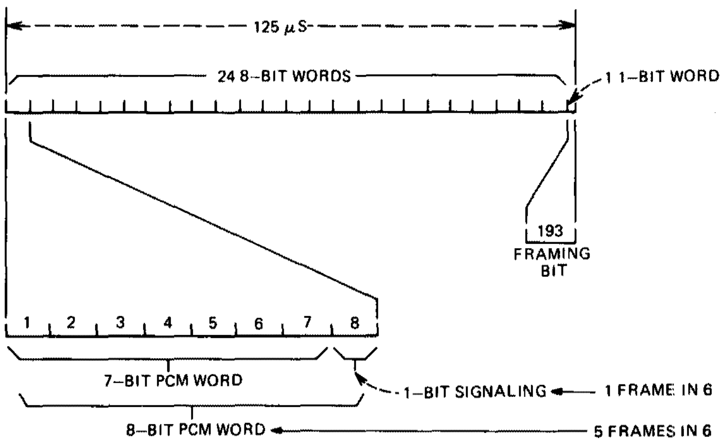


Fig. 19-6. The D2 bit stream format.

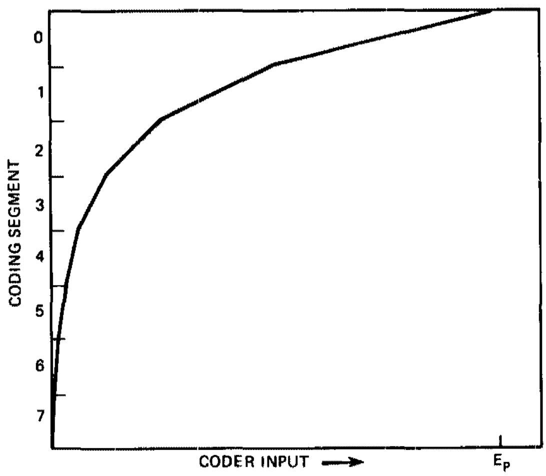


Fig. 19-7. Fifteen-linear-segment approximation to the $\mu=255$ compression characteristic. The remaining segments are in a symmetrical negative portion. The two segments from zero input are colinear.

it was possible to translate the binary code words representing the compressed amplitudes into other codes representing linear, uncompressed signals without incurring any additional quantizing noise. In the linear digital form, signal processing, such as gain change, echo suppression, and conferencing arrangements could be carried out by purely digital processes.[11,12]

The nonuniform-step coder consisted of an arrangement of tandem stages with one stage for each of the eight digits. Each coding stage produced a digit and a residual amplitude pulse. The residue was the input to the next tandem stage [Fig. 19-8]. The first stage determined the polarity of the sample and each of the remaining stages produced one digit of the binary code representing its magnitude. (This was equivalent to weighing an object in a balance scale by a sequence of successively smaller known weights.) The nonlinearity was generated by controlling the corners and relative slopes of the segments. Unlike the D1 channel bank, where the compression was due to the nonlinearity of active devices (diodes), the segments were determined in passive networks by highly precise tantalum thin-film resistors.[13] The approach used in the D2 bank

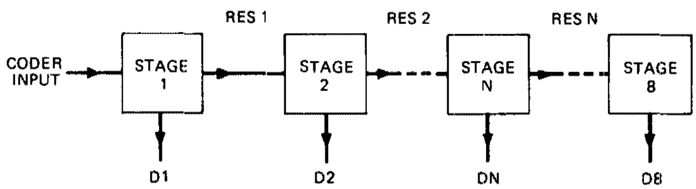


Fig. 19-8. The D2 channel bank tandem stage coder.

eliminated the need for field adjustments of the companding action. In fact, careful design eliminated all field adjustments of common circuits.

1.2.4 Testing

A novel and important maintenance feature of the D2 design was a built-in digital test panel. This allowed initial setup and maintenance measurements to be made by one person and with a minimum of external test equipment. The test circuit included a digital signal generator that produced a PCM signal that was the digital equivalent of the 1-mw, 1000-Hz sine wave often used to test analog transmission equipment [Fig. 19-9]. By looping the output of the bank back to the receiving side and using the "digital milliwatt," only one person was needed to align a D2 channel bank, and it could be done independently of the alignment of the far-end bank. The procedure took advantage of the fact that digital transmission lines have zero loss for the encoded signals passing over them. A technician at the D2 bank could use the test panel to simulate reference-level incoming signals and adjust the bank's channel-gain controls. These adjustments could be made before the channel bank was connected to a digital transmission line, thus assuring that all the voice channels would be be properly aligned when any two individually adjusted banks were connected.[14]

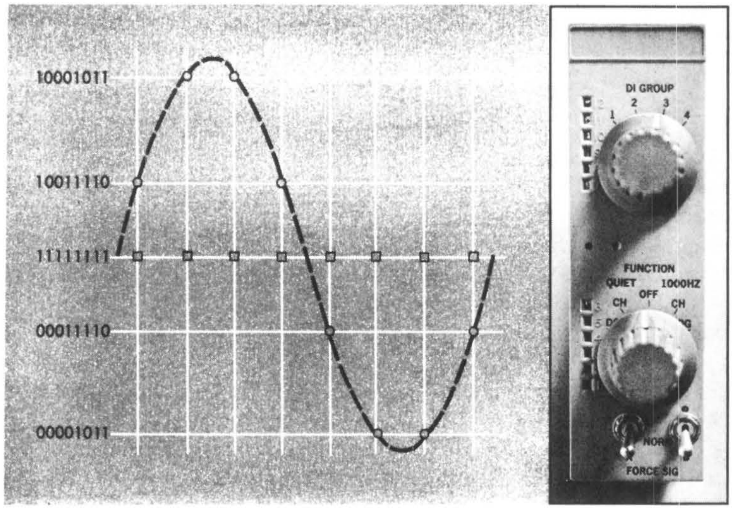
1.3 The D3 and D1D Channel Banks

1.3.1 The D3 Channel Bank

The same advances in device technology that made the D2 design possible made it apparent that an improved and less expensive digital channel bank for the exchange area application was also possible. Exploratory work with this purpose in mind started in 1968 in the Merrimack Valley transmission group, which was responsible for the D1 channel bank. The availability of small-scale integrated circuitry, especially in the hybrid tantalum-film/silicon-chip form, made improvements in the common digital circuitry and the voice-frequency filter look attractive.

With over 150,000 D1 channel banks in service, end-to-end compatibility with D1 was initially accepted as a firm requirement. As originally conceived, therefore, the new channel bank (later designated D3) would have been incompatible with the D2 toll bank. In retrospect this may seem strange, but it should be recalled that the D1 design was in large-scale production and would continue to be so for some time to come. There were and would continue to be many partially filled bays in the field. Competitive D1-compatible banks were beginning to appear in the market. The D3 channel bank was originally conceived as a lower-cost D1. Performance improvements were sought, but toll performance was not one of the original objectives.

As the development progressed, the limitations imposed by the D1 features, compared to what had become technically possible, became increasingly apparent. Companding by diodes in the D1 design had always been a maintenance burden, especially where voice-band data transmission was involved. The tightly controlled and adjustment-free segmented companding was adopted at



(a)

(b)

Fig. 19-9. The D2 channel bank digital test panel. (a) The panel generated the digital equivalent of a 1-mw, 1-kHz test tone. (b) With the test panel, a channel bank could be aligned independent of the far end.

an early stage. Once this was part of the plan, the advantages of $\mu = 255$ companding and the improved performance from eight-bit coding in five out of six frames were temptingly available and readily achievable—except for the D1 compatibility requirement. After several months of analysis, experiment, and debate, a decision was made to make the D3 bank compatible with D2 and to furnish a relatively inexpensive conversion kit with which existing D1 channel banks could also be made compatible with D3 and D2. Thus, the telephone companies could leave existing exchange links as they were, filling out with D1 banks if they chose (D1 continued in production); buy the improved bank for both ends of a link; or use the new bank in conjunction with old D1 banks upgraded with the conversion kit.[15]

With the format issue settled, the D3 designers set out with zest to exploit the new devices and integrated-circuit techniques. For one thing, integrated circuits were much smaller. Integration had not yet reached the stage where the function of entire panels and bays were reduced to a silicon chip, but 8-inch, plug-in wiring boards could be replaced with 2-inch-square ceramic hybrid integrated circuits. This opened the door to some advantageous changes in architecture. As in D1, the bank continued to handle 24 channels for a single T1 line but required only one coder-decoder. The filters and gates for each input line were moved from eight boards in the common control area to the 24 channel-unit plug-in boards. All the per-channel functions were therefore on a single plug-in board. Channel boards could then be furnished only as required, and troubles affecting one channel could be corrected without affecting any other channel. The channel units could still be tailored to a specific application, such as two-wire or four-wire connecting circuits, or be provided with various special service features. The remaining common-circuit functions, which had occupied 24 circuit boards in the D1 channel bank, were reduced to three plug-in units.

These measures, along with the new devices, resulted in a two-to-one reduction in both size and power consumption as well as lower costs. Several other features that were pioneered in the D2 channel bank and reduced the cost of operation and maintenance were included in D3. Built-in test gear with the digital milliwatt, a hot spare for fast substitution of defective plug-in units, built-in carrier-group alarms to inhibit switches from making connections to faulty or out-of-service banks, and built-in power converters on a per-bank basis were made part of D3. Line-up adjustments were essentially eliminated. The only adjustments were those necessary to the specific application, i.e., those required for direct, tandem, or toll-connect trunks, and these were made by prescription setting of switches, which were expected to remain fixed for the duration of that application or for the life of the equipment.

The improved features made the D3 channel bank highly attractive for new exchange-area installations; its market was assured. At this stage, however, it still fell short of meeting toll requirements, principally in its lack of access jacks, normally provided for toll testing and patching, and in the characteristics of the voice-frequency filter. But it was getting very close to toll quality, as its designers were well aware. The access-jack deficiency was largely removed by the growing use in the 1970s of switched maintenance-access arrangements. The new bank used an active filter to limit the VF band. (See Chapter 17, Section 3.3.) The active filter was smaller and comparable in cost to a conventional passive design but with more potential for cost reduction. When it turned out that the cost difference between filters meeting exchange and toll requirements was trivial, the D3 developers adopted the better filter and gleefully announced that they, too, had a toll-quality bank.

A successful field experiment of the D3 design was carried out in Lawrence, Massachusetts in November 1970, and preproduction field trials were held in New York City in early 1972. The first regular service installation was in Calumet City, Illinois in March 1973. With space, power, cost, and performance advantages, the D3 channel bank took over the bulk of the market for new banks as production capacity was built up.

1.3.2 The D1 Conversion Kit, D1D

The need to furnish a conversion kit to upgrade existing D1 channel banks emerged as soon as it was accepted that the advantages of making the D3 channel bank compatible with D2 rather than D1 outweighed the disadvantages. The concept of a conversion kit and the electronic design were straightforward enough, but the execution was challenging because of some severe physical constraints. The conversion was to require changes in a minimum of plug-in units only; no frame-wiring changes were to be required. The new units could consume no significant additional power. Performance was to be as least as good as the original D1 channel bank and preferably better.

In this spirit, it was not possible to relocate the channel-sampling gates to the channel units; changes were confined to the common control circuits. When the design was complete, the change was accomplished with a kit of 12 plug-in circuit boards to replace 12 of the original 31 boards in the D1 common control. Several of the new boards used state-of-the-art integrated circuitry and special assembly techniques, such as multiboard plug-in units, to save space [Fig. 19-10]. With the kit installed, the D1D banks had the improved companding, 8-bit coding, the signaling capability of D2 and D3, and improved alarm controls to prevent switching-machine seizures of trunks in defective units. In addition, the modified banks could be slaved to remote-clock signals for timing (a feature needed for some of the new switching interfaces and not provided in the D1 bank) and lined up and tested from one end. Field adjustments were reduced by two-thirds, with none remaining in the common equipment. In every respect, the D1D conversion kit met the objectives set, and performance was made comparable in most respects to the new D3 design at a fraction of the cost of an entire bank.[16] Shipments were made both as kits for upgrading installed banks and as complete banks with the new plug-in units in place. Deliveries started in 1973, and, by 1980, over 30,000 D1D channel banks were in service.

II. SIX-MEGABIT AND THREE-MEGABIT SYSTEMS

2.1 T2 Carrier System Development

As soon as the T1 system was well established, attention was directed to higher-capacity systems designed for use over longer distances. In 1963, pre-

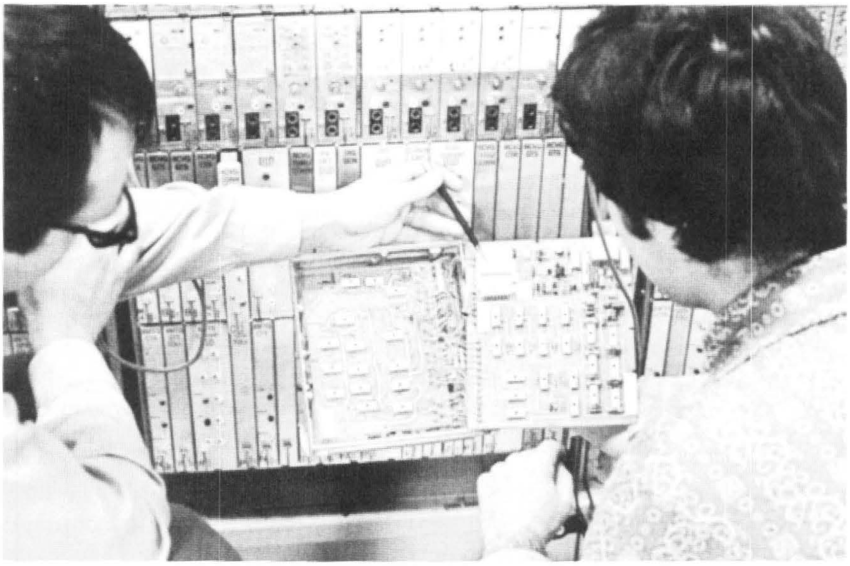


Fig. 19-10. The D1D conversion kit. A multiboard plug-in unit has been opened showing hybrid integrated circuits and dual-in-line packages carrying silicon integrated circuit chips.

liminary studies led to a proposal to develop a system with four times the capacity of T1, that is, 96 PCM channels, at about 6.3 megabits. The system was to be for transmission over pulp- or polyethylene-insulated pairs and good for applications in the short- and intermediate-range toll plant, as well as in exchange areas. This initial work evolved into the T2 carrier system.

An early hope was for single-cable operation at the nominal 6000-foot T1-regenerator spacing. The regenerator design was extremely difficult, as the pair loss at the 6.3-megabit digital rate required was about 70 dB. In 1967, field tests with an experimental regenerator quickly revealed that, not only was the single-cable operation not feasible, but application even to existing twin cables was not promising.

Nevertheless, the enthusiasm for longer digital links remained very high. Much of the push during the late 1960s and early 1970s came from the expectation that some form of T2 could provide a badly needed intercity link for PICTUREPHONE* visual telephone service, which was in intensive development at that time. With digital picture processing, it was expected that a reasonable-quality monochrome picture, equivalent to one in about a 1-MHz analog band, could be transmitted over the 6.3-megabit links. The options were limited. On existing types of paired cable, the loss at the high bit rate required regenerator spacings so short that maintenance of longer routes appeared impractical. This led to the development of a new low-capacitance (LOCAP) cable with lower loss and lower crosstalk, designed specifically for T2.

* Service mark of AT&T.

Further surveys continued to point to moderately long intercity links as the principal field of use. As a result, a great deal of attention was directed to the maintenance problems that would attend such lines. Automatic protection switching to a working spare was needed, and means were developed for the rapid sectionalization and localization of line failures.

2.1.1 System Description

T2 carrier was designed to transmit 96 PCM channels at 6.312 megabits (the second digital signal rate, DS2) over distances up to 500 miles. The input was to be either four time-multiplexed T1, 1.544-megabit signals or a digitally encoded picture signal that would occupy the entire bit stream. Operation was over separate cables for the opposite directions of transmission, with repeaters spaced up to 2.8 miles apart on LOCAP cable. The high bit rate and high loss required a regenerator using the best available high-speed devices. While the principles of operation remained the same—to reshape, retime, and regenerate the signal—the greater speed and precision needed resulted in a regenerator considerably more complex in circuitry than the T1 regenerator [Fig. 19-11].[17] Crosstalk requirements were so stringent that it was not possible to mount two oppositely directed regenerators on one card, as was done in the T1 repeater. The separate regenerators were mounted in separate apparatus cases, one for

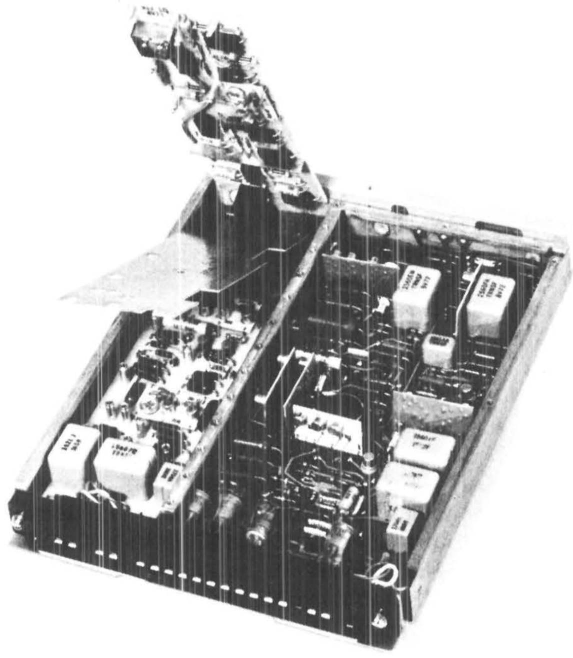


Fig. 19-11. The 6.3-megabit T2 regenerator. Two regenerators were required for signal transmission in both directions.

each cable. Automatic protection switching to a working spare line was provided. The development program included the design of a multiplexer, the M12, to combine the T1 signals, as well as the realization of the new cable.[18]

2.1.2 *Low-Capacitance Cable*

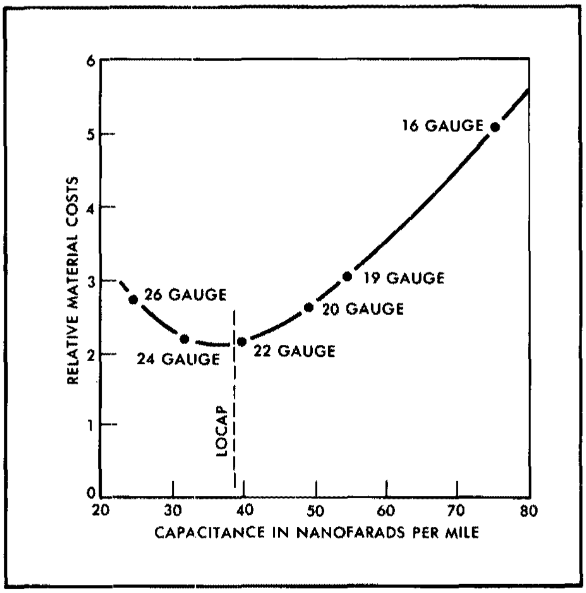
There was more than one possible approach to a lower-loss cable. The loss desired was about one-half that of a standard No. 22 gauge polyethylene-insulated-conductor (PIC) pair. Pair loss depended on both the series resistance (directly related to the gauge) and the shunt capacitance (a function of the insulation material properties and thickness). A brute-force approach was to scale up both conductors and insulation to reduce loss. This would have required No. 16 gauge conductors with quadruple the amount of copper and would have been bulky as well. However, the loss caused by the high resistance of finer-gauge conductors could be offset by lowering the capacitance with even thicker insulation. At some point, the saving in copper costs would be more than offset by higher insulation and sheath-material costs. An optimum was found with No. 22 gauge conductors and expanded (foamed) polypropylene insulation [Fig. 19-12].

Fifty-two pair cables were constructed, with seven units of seven LOCAP pairs each plus three interstitial pairs. Within each unit, each pair had a different and carefully controlled twist length. The seven-pair unit was constructed by stranding, or winding, six of the pairs around the seventh. The design minimized crosstalk between the pairs of a unit and between units as well. The interstitial pairs were used for maintenance and for voice-frequency communications along the route (i.e., an order wire). LOCAP cable was successfully tested in a field trial in Willow Grove, Pennsylvania, in 1971, and was used in subsequent T2 installations [Fig. 19-13]. Designs for 104 and 204 pair cables were also developed.[19]

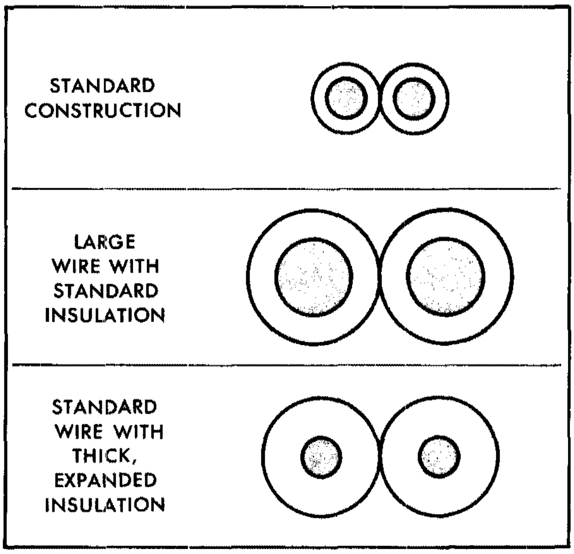
2.1.3 *The M12 Multiplexer*

The M12 Multiplexer accepted 1.5-megabit (DS1 rate) signals from up to four T1 lines or any of the variety of D-type channel banks that produced the DS1 signal. (By the early 1970s, there were in service or in planning not only D1, D2, and D3 banks, but versions specially designed for data, e.g., the T1WB4. (See Chapter 22, Section 3.4.1.) It multiplexed these signals into the 6.3-megabit signal, accepting all pulse patterns and synchronization plans and demultiplexed them at a receiving terminal, exactly reconstructing the 1.5-megabit patterns received [Fig. 19-14].

Like the D2 channel bank, the M12 multiplexer had to accept 1.5-megabit streams that differed in rate. But, where the receiving D2 channel bank dealt with the asynchronism in the process of reconstructing analog voice signals, the M12 multiplexer had to accept mixed-rate streams, multiplex them for transmission over the T2 line, and deliver the original signal to a T1 line or D channel bank at the receiving end unchanged in pulse content and rate. The ingenious method for accomplishing this was dubbed pulse stuffing. The multiplex output, 6.312 megabits, was chosen to be slightly higher than any possible sum of the pulse rates expected from the four DS1 inputs (note that

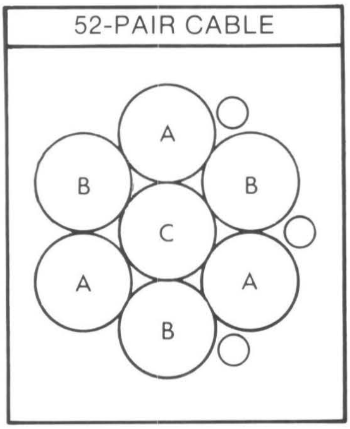


(a)



(b)

Fig. 19-12. Low capacitance (LOCAP) cable design. (a) The least expensive cable meeting the design specifications for the T2 digital carrier system had No. 22 gauge copper wire insulated for a capacitance of 39 nanofarands per mile. (b) LOCAP cable used one-quarter the copper and was smaller in diameter than a No. 16 gauge design for the same loss.



(a)

(b)

Fig. 19-13. Low capacitance (LOCAP) cable. (a) Fifty-two pair LOCAP cable had seven units of seven pairs each. Units of the same twist and stranding structure (*A, B,* and *C*) were separated to reduce crosstalk. (b) Twin 52-pair LOCAP cables being installed on first T2 route from Dallas to Longview, Texas, 1972.

$4 \times 1.544 = 6.176$, not 6.312). The extra capacity was used by inserting (stuffing) extra pulses into each incoming signal until its rate was raised to that of a locally generated clock signal.

In the actual process, the combined signal was formed by a digit-by-digit interleaving of pulses, detected and made much shorter in duration by the multiplexer. After every 48 information digits, 12 from each DS1 signal, a control digit was inserted, forming a multiplexer frame. Any deficiency of pulses necessary to read out the frames at the (higher) local-clock rate was made up by inserting pulses at preselected positions with respect to the frame-control pulses. At the receiver, the known stuffing positions were examined and any added pulses removed before demultiplexing down to the separate DS1 bit streams [Fig. 19-15]. The process of reframing and regenerating the DS1 stream then recreated the 1.5-megabit signal, including its original rate.

By this means, differences in the DS1 signals up to 5.4 kilobits per second,

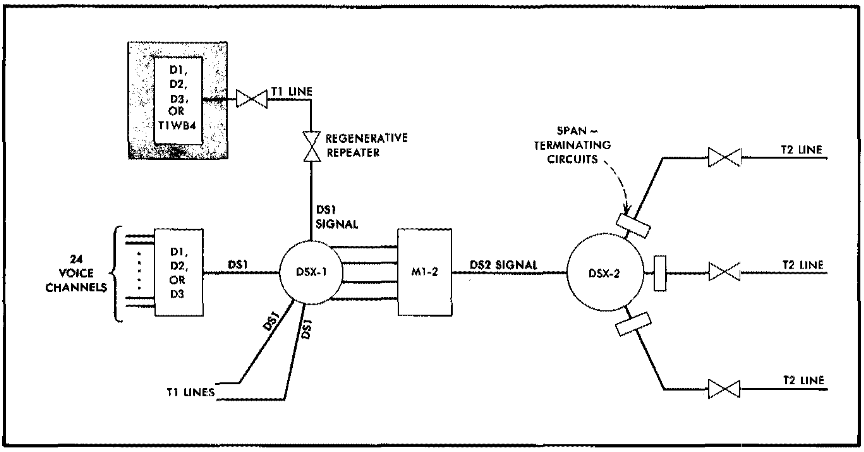


Fig. 19-14. Multiplexing for the T2 6.3-megabit line. DSX-1 and DSX-2 are cross-connects at the 1.5-megabit DS-1 and 6.3-megabit DS-2 rates.

which was much larger than the rate tolerance of the incoming signals, could be accommodated. Pulse stuffing was efficient in that it did not require a great deal of buffer storage, and it was generally used in the successive higher stages of digital multiplexing. Finally, it preserved flexibility: it did not preclude other methods of digital-network synchronization. If lower-speed signals were made synchronous for use in a switched digital network, pulse stuffing could still be used for the higher-speed transmission facilities.[20]

2.1.4 *Toll-Line Maintenance*

As on all toll lines, continuity of service was extremely important in T2. Unlike the exchange area, troubles were likely to be far from an attended office and alternate facilities much less available. T2 provided automatic protection switching to a working spare, with one protection pair for up to 23 working lines. The cable facility was divided into maintenance spans about 100 miles

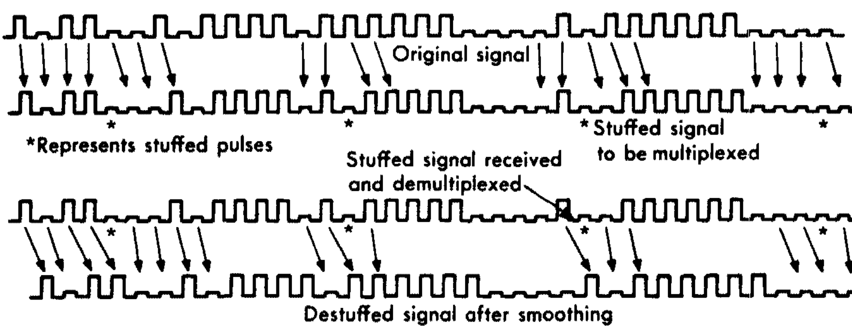


Fig. 19-15. Pulse-stuffing synchronization.

long, with performance monitors installed at the end of each span. The line signal was bipolar, that is, successive pulses alternated in polarity. Violation of the polarity alternation was sensed and used to raise alarms and actuate the protection switch if the problem was severe. When a monitor sensed a violation, it removed it and transmitted a violation-free signal so that the violation would be detected only once. This localized the trouble to the maintenance span preceding the monitor sensing the polarity violation.

Further trouble localization was provided by means of a test signal transmitted from a maintenance office. Each repeater location had a narrow band-pass filter tuned to a particular voice frequency and coupled between the regenerators and a fault-location line (one of the auxiliary pairs in the cable). The test signal was generated with a voice-frequency component, which was changed to match the frequency assigned to each repeater site in succession. If a repeater failed, it would not return the tone unique to that location, and the regenerator or preceding cable section was then known to be in trouble [Fig. 19-16]. By this means, maintenance people could be dispatched to specific locations to make repairs or replace equipment. Finally, T2 carrier was designed with interfaces to a standard Bell System telemetry system (called E2 telemetry). This allowed remote monitoring of unattended maintenance offices. Remote monitoring was also used for centralized administration. With the telemetry, nearly all the indications available at the terminal bays could be monitored at a single central location. It was thus possible to keep track of the performance of the digital network over an entire area and to take coordinated action in dealing with major problems.[21,22]

2.1.5 *Type T2 Carrier in Service*

The first T2 carrier installation was made between Dallas and Longview, Texas (125 miles) in 1972. Others followed in Indiana and Ohio. The system

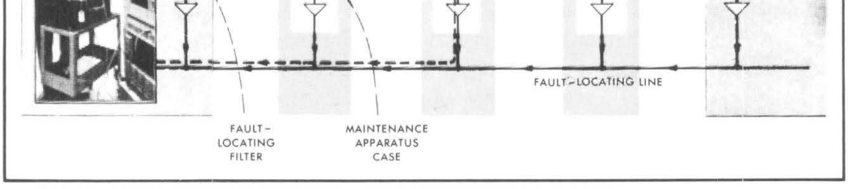


Fig. 19-16. T2 carrier system fault location. Test signals with a voice-frequency component (f_1, f_2, etc.) unique to each repeater helped to localize failures.

met its design objectives and worked well. Within the next few years, improved regenerators were designed with integrated circuits requiring less space and power. By 1980, almost four million voice-circuit miles had been installed, but this was only about four percent of the digital-circuit mileage in service. A major restraint in the hoped-for market was the failure of the PICTUREPHONE visual telephone service network to materialize.

2.2 T1C and T1D Systems

2.2.1 T1C—A 48-Channel, Three-Megabit System

While T2 carrier proved satisfactory for intermediate-range digital toll links, the need for larger-capacity and more economical systems in the exchange area continued to grow. The incentive to obtain more circuits on existing cables increased in the mid-1970s, when rising copper costs and an increase in the cost of petroleum-based plastics made new cables more expensive. The search for a larger-capacity exchange-area digital system had not ended with the decision to tailor T2 to intercity applications on LOCAP cable. The rapid rate of T1 installations and the rise in new cable costs led to another study of the problem, with interest focused on a 48-channel, 3.15-megabit system, called T1C for T1 conversion.

As indicated by the name, the system was intended as a means for a possible upgrade in the capacity of routes already equipped with T1 carrier. Aside from the bit rate, T1 and T1C had many similarities. Transmission was bipolar with a 50-percent duty cycle. Installations could be on single, screened, or two separate underground or aerial cables; on Nos. 19, 22, or 24 gauge pulp- or PIC-pairs; in pressurized or filled waterproof sheaths. System length between terminals was limited to 50 regenerators, or about 50 miles, the same as T1. Regenerators used automatic line buildout, power was simplexed over the pairs, and maintenance and trouble location was by bipolar violation monitoring in a way similar to T1. The construction of the 3.15-megabit line signal was accomplished in a M1C multiplexer, highly similar to the M12 multiplexer used to assemble the T2 signal, except that only two 1.5-megabit DS1 streams were combined to form the 3.15-megabit DS1C line signal. Asynchronous DS1 streams were combined using pulse-stuffing.[23,24]

As always, the big problem was crosstalk. In particular, for single-cable operation, near-end crosstalk (NEXT) between the high-level output pulses and the low-level inputs was limiting. The line loss, and hence the repeater gain, was about 16 dB greater than T1 at the reference frequency of 1.576 MHz for the bipolar 3.152-megabit signal. This, combined with the greater crosstalk coupling at the higher frequency, increased NEXT by a factor of 100 over T1. The principal system problem lay in finding the increased margin required.

The adverse factors would seem at first glance to place the T1C developers in precisely the position that was found to be impossible in the original T1 carrier development before the earlier system was rescued by the use of bipolar transmission. Fortunately, extensive field experience with T1 had furnished a great deal of information about the outside plant that was not available to the first system's designers. The T1C developers found that a great deal of the crosstalk coupling arose where cable symmetry was lost in manholes, entrance stub cables, apparatus cases, and splices. A large reduction in NEXT was obtained with a new design for the repeater apparatus case. In addition, the new case employed two separate stubs to connect it to the main cable outside the manhole, one for low-level input signals and the other for the high-level regenerated signals, even where only a single cable was used between repeaters. The apparatus case was similar in appearance to the 1970 design but was slightly larger (and looked even more like a hibachi grill) [Fig. 19-17]. Three repeater housings were provided with separate removable covers. Two of the housings had slots for ten repeaters; the third housed five repeaters, a fault-location filter, and loading coils for the fault-location and order-wire pairs. All slots accepted either a T1 or T1C repeater, with regenerators for both directions of transmission on a single card.

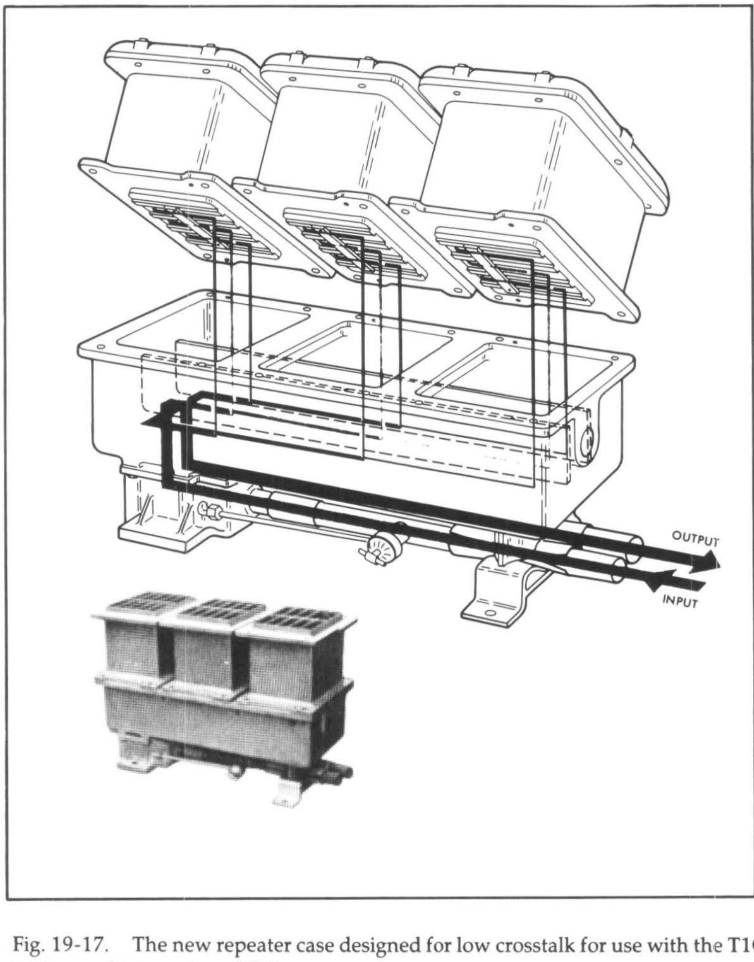


Fig. 19-17. The new repeater case designed for low crosstalk for use with the T1C 3.15-megabit repeaters, 1975.

Besides the two stubs to separate the input and output pairs, the two directions of transmission were kept separated inside the case as much as possible. The repeaters on one side of the housing were inserted with an orientation opposite to that of the repeaters on the other side. This allowed all the low-level input pairs to occupy the center of the housing and the high-level output pairs the outer edges. The separation was maintained to the point where the separate stubs entered the base. By these means, the crosstalk contribution of the T1C apparatus case was made negligible compared to other sources. Shielding of cable splices was also necessary, especially for those close to a repeater where the level differences were large.

In unscreened single cables, it was still necessary to select cable pairs in units (binder groups) with the maximum separation within a sheath. In T1 cables, at least one intervening cable unit without carrier was necessary between units carrying the opposite directions of transmission. Because of the higher coupling at the T1C rate, at least two intervening units were necessary [Fig. 19-18]. As a result, for single-cable operation on 900-pair cable, it was possible to operate up to 150 T1C systems (7200 channels on 300 pairs), compared to a maximum of 200 T1 systems (4800 channels on 400 pairs). A full complement of T1C also released an additional 100 pairs for voice services, but it was not possible to mix voice and carrier pairs in the same binder group. Nor could T1C and T1 signals coexist in the same binder group, but they could be applied

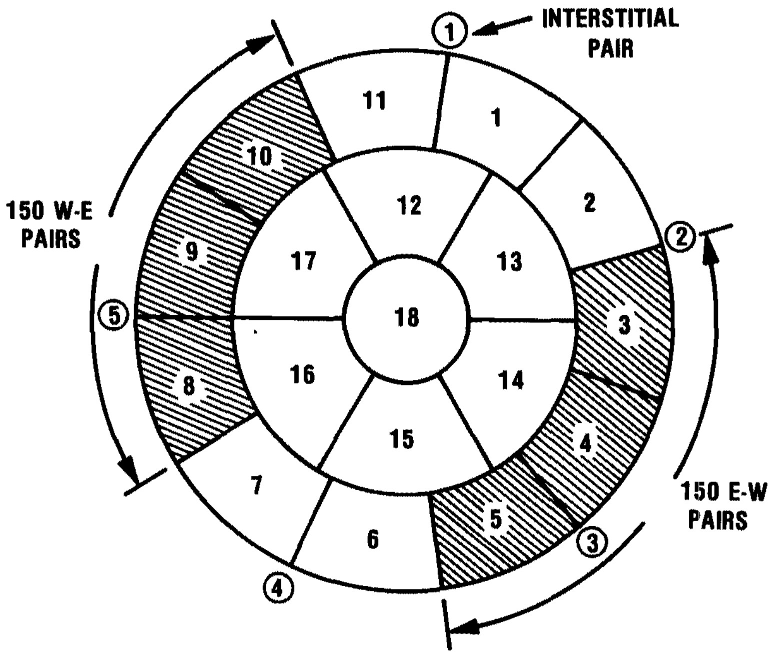


Fig. 19-18. Unit selection for single cable operation with 150 T1C systems—900 pair cable.

to separate binder groups in the same cable, making group-by-group conversions feasible. With two-cable, or screened-cable operation, far-end crosstalk rather than NEXT was limiting, but it was low enough to allow T1C operation on all pairs, with only the interstitial pairs reserved for fault location and order wires. On two 900-pair cables, up to 850 T1C systems (40,800 channels) were possible and, on a single screened 900-pair cable, as many as 400 T1Cs (19,200 channels).[25]

2.2.2 *The T1C Repeater*

The paired regenerators in a T1C repeater were similar in principle to the T1 repeater and bore a strong family resemblance to the later T1 versions. Three custom-designed silicon integrated circuits mounted on ceramic plates, only 1 by 1.3 inches, provided preamplification and equalization, clock extraction, and regeneration. An automatic line-build-out network adjusted the repeater to accommodate incoming signal levels for line losses from 6 to 54 dB at the 1.5-megabit reference frequency. In the spirit of a conversion system, the size and power consumption were equal to or less than T1. The two regenerators on a single card occupied a volume 3 by 6 by 7/8 inch [Fig. 19-19].[26]

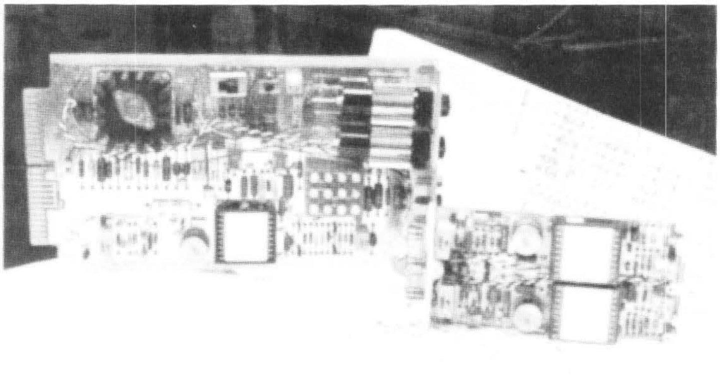
Outside-plant conversions from existing T1 installations to the new apparatus cases, shielding, and two-stub arrangements were made; but, as soon as the design was assured, the local Bell System companies were urged to use the new apparatus cases and plant layouts, even though a 24-channel T1 line was adequate for their immediate needs. Thus, at only slightly greater expense, later clean conversions from T1 to T1C could be made by merely adding the M1C multiplexer to the terminal and exchanging the plug-in repeaters. (The removed repeaters could be used elsewhere, where T1 capacity was sufficient.)

A highly successful field trial of 50 T1C systems in a single cable over a 16-mile span was completed in Atlanta by June 1975. First installations from regular production were by Southern Bell Telephone and Telegraph Company in Miami later that year. By the end of 1980, about 5.5 million voice-circuit miles of T1C were in service, and 15 percent of new growth in local digital capacity was on T1C.

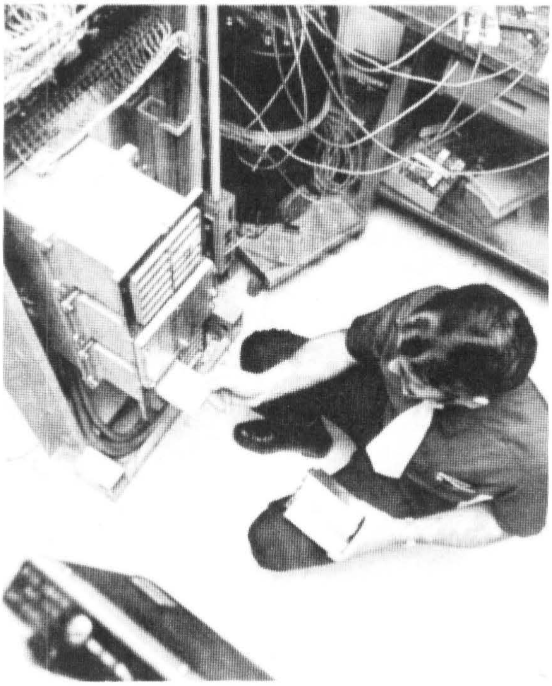
2.2.3 *T1D—T1 Duobinary*

T1C had been made possible by improved apparatus-case design and outside-plant arrangements. By 1980, advances in integrated circuits made more complex signal processing possible. The pattern of zeros and ones in the digital sequence was manipulated, without changing the digital rate, to produce a spectrum with less energy and, hence, lower crosstalk at the critical regenerator timing frequency. The T1 duobinary (T1D) system was a 3.15-megabit system with a line-signal format somewhat like two interleaved T1 signals and with crosstalk at 1.5 megabits only 6 dB more severe than T1.[27] The new format, along with the outside plant improvements realized for T1C, permitted much greater utilization of cable pairs [Fig. 19-20].

In the early 1980s, T1G, a 96-channel system with six-megabit (four-level, three-megabaud) signals, was developed to double the capacity of T1C lines.



(a)



(b)

Fig. 19-19. T1C repeaters. (a) *Left,* T1C office repeaters and *right,* manhole repeaters. (b) Laboratory test of manhole apparatus case and repeaters.

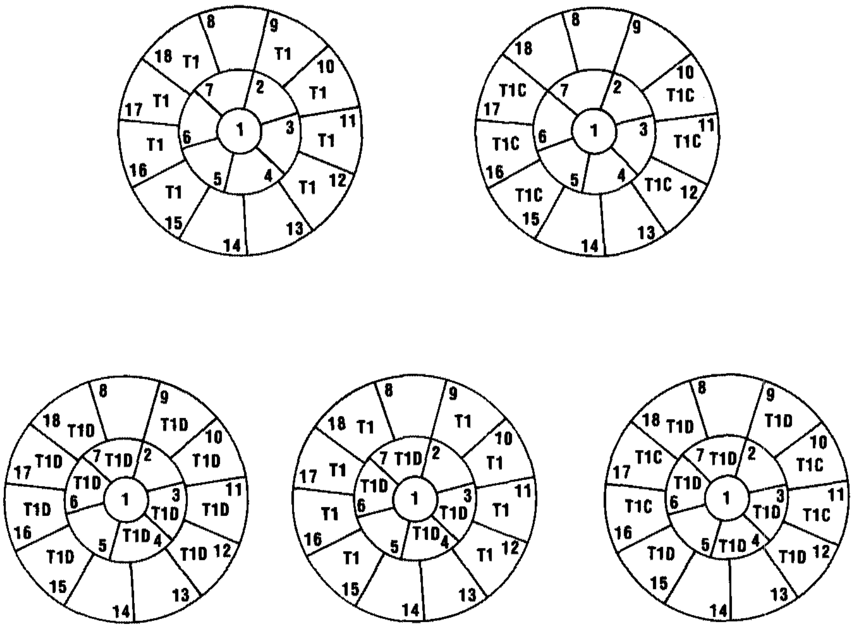


Fig. 19-20. T1, T1C, and T1D unit selection and cable utilization for single cable operation on a 900-pair cable.

The original objective of the T2 development, to transmit 96 channels over ordinary exchange cable at 6000-foot repeater spacing, was thus finally achieved, but only after 15 years of device advances and improved understanding and upgrading of the outside plant. (While it was possible to operate the higher-rate exchange systems on single cables, their use was usually in high-growth areas where two-cable and screened-cable applications were common.) T1D was in commercial production by 1980, and T1G by 1984.

2.3 The D4 Channel Bank

In mid-1973, about a year after the start of the T1C development, studies were initiated on how best to provide terminals for the new 48-channel line. Clearly, a stand-alone multiplexer was needed to provide an interface with the numerous D3 and D1 banks already in service, but the expected market for T1C was so large that a new terminal, tailored to T1C and taking advantage of the latest technology, also looked attractive. For a time, a D3-based system with a built-in multiplexer was considered. As the study progressed, however, the advantages of a more flexible approach became apparent.

Looking ahead a few years, planners could see a situation of growing complexity in the digital plant. Signals could be transmitted over T1, T1C, or T2 systems, as well as coaxial cable or other higher-rate lines. The lines might be terminated in any of three different digital banks (D1/D1D, D2, or D3), two multiplexers (M1C, M12), or a digital terminal for the new toll 4ESS* switch, then under development. The sampling rate and companding law had been standardized (providing D1 channel banks were converted to D1D), but the lines carried signals at different bit rates, the multiplexers combined signals in different patterns, and the channel banks could sample signals in different sequences. The emphasis for the new channel bank shifted from optimizing for one specific application to a design with maximum versatility and flexibility. The outcome was the enormously successful D4 terminal. With the larger-scale integrated circuits that became available, little or no sacrifice in economy was necessary compared to a more specialized bank.[28]

The D4 channel bank was composed of a set of up to 48 individual channel units, each of which handled a single trunk or special service line, and common control circuits—the brains of the equipment. (The term *special services* referred to anything other than plain old telephone service (POTS). They included a great variety of private lines, such as data lines, audio programs, connections to private branch exchanges, and a number of others. They were of growing importance, accounting for a substantial fraction of all lines, and an even larger part of the growth. About 30 different versions of D4 channel units were planned for special services.) It is beyond the scope of this history to go into any detail of the circuitry of the D4 channel bank. The techniques used were basically similar to those developed in the D1, D2, and D3 channel banks. The principal difference was the elaboration and great consolidation made possible by the larger-scale integrated circuits.

A key feature of D4 was the provision of four different digital line interface units, one for each of the basic line equipment combinations or modes [Fig. 19-21]. The simplest and most efficient combination was a D4 terminal transmitting over a T1C line to another D4. The 48 channels were automatically synchronized and only one alarm control was needed. When D4 was to be connected by a T1C line to existing 24-channel banks or T1 lines, an M1C multiplexer was required to combine the far-end 24-channel facilities and synchronize the two groups of 24 channels by pulse stuffing. At the D4 end, a line-interface unit and a synchronizer-desynchronizer (syndes) unit performed these functions. Two alarm-control units maintained independent 24-channel operation. In a third mode, D4 connected to two T1 lines with a line-interface unit that made the channel bank operate as though it were two D3 compatible banks. The D4 bank could then be used with D1D, D2, D3 channel banks; the new digital switch termination; or any combination of these at the far end. Finally, two D4 banks could be linked to provide a 6-megabit, 96-channel signal for a T2 line. At the far end, the T2 line could be terminated in another pair of D4 banks or in an M12 multiplexer for connection to 24-channel banks. In this mode, the line interface ensured that the signal frames were compatible with an M12 multiplexer. Syndes and alarm controls linked the two D4 banks to provide, in effect, four independent 24-channel groups.

During this period and earlier, a great deal of study was devoted to the emerging digital hierarchy, with the intention of defining a sequence of suc-

---

* Trademark of AT&T Technologies, Inc.

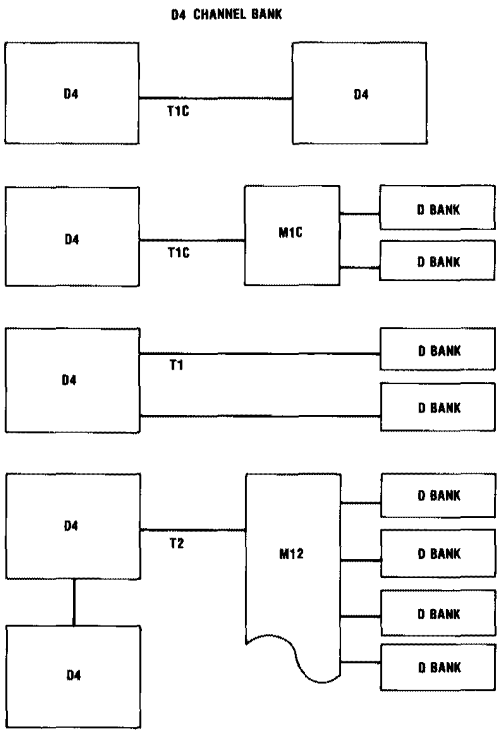


Fig. 19-21. The D4 channel bank connected to three different lines and a variety of far end terminals in four different modes.

cessively higher-rate signal formats that would be compatible and flexible.[29] The successive steps at 1.5 megabits (DS1), 6.3 megabits (DS2) and 45 megabits (DS3) rates were intended to provide recognized service needs or convenient-size steps for new facilities. The D4 3.15-megabit mode was not hierarchy compatible, an advantage that was considered perhaps worth sacrificing in some cases in the interest of economy. The remaining modes were all compatible.

The D4 channel bank was developed through 1974 and 1975 and tested in a field trial in Pennsylvania in 1976. It featured several physical-design innovations, including a printed-wiring back plane into which all units were plugged. Size reduction, made possible by integrated circuits, doubled the number of channels per bay compared to the D3 channel bank, with about one-half the power per channel (i.e., no greater power per bay). Power converters had also been so reduced in size and cost that they were furnished for each bank. D4 was the first fully stand-alone digital bank and could be installed individually if required. Within a few years after its introduction, virtually all D4 units were redesigned to take advantage of very-large-scale integration (VLSI). As a typical example, the changes reduced the integrated-circuit device count in the receiving unit from 51 to 5 per unit. Commercial production started in late 1976, and the first service application was in the Crete, Illinois central office of Illinois Bell. By the early 1980s, D4 channel banks were being produced at a rate of over 2,000,000 channel terminations per year (45,000 banks).

III. INTERFACES

Unlike power or gas utilities whose configurations are relatively static, the basic purpose of the telephone plant is to permit different connections of terminals, customer lines, and interoffice trunks to form end-to-end paths on request. Less obvious is that the facility plant—the complex of switching offices, cables, multiplexes, and transmission lines—is not static, either. Not only does it grow and evolve over time, but it is constantly being rearranged and reconfigured to serve the basic function better. The flexible and efficient use of facilities requires the association of different types of transmission systems and of transmission and switching systems in many different and easily rearranged combinations. Traditionally, this led to stand-alone facilities with well-defined interfaces but with enough options to mate with any other system to which it could be connected.

By the mid-1960s, three very different types of carrier transmission systems were in widespread use: (1) the classic, analog, linear, single-sideband-suppressed-carrier L-multiplexed family, which included the coaxial and FM microwave systems. (While the radio lines used FM, the terminal multiplexes to form the composite signal were of the same type as the coaxial systems.) (2) the analog, compandored, carrier-transmitted, Type N family; and (3) the recently introduced digital system. Interconnection between the types was done at voice frequency, requiring back-to-back terminals, even though access to the individual voice circuits may not have been required.

About 1965, studies and development started on methods for connecting larger blocks of circuits in a way less expensive than complete multiplex terminals. By 1968, an N3-L junction was produced that permitted direct interconnection of N3 and L systems in 12 channel groups.[30] The junction processed the groups by changing the relative magnitude of the carriers and speech signals of an N3 carrier line to make them suitable for transmission over a coaxial or radio link. The resulting line signals, however, were not compatible with L multiplex; for one thing, they were still compressed in volume by the N3 compandor. A special terminal to perform the inverse carrier-speech level adjustment was necessary at the far end before recovering the voice signals in a standard N3 receiving terminal.

The N3-L junction saw only limited field use because of the rapid expansion of the digital systems and the expected decline of N carrier. However, as more transmission links became digital and switching became increasingly software-controlled and finally digital, it became apparent that considerable economies could be realized by more integrated interfaces between transmission terminals and the electronic switches.

3.1 The 4ESS Switch—Voice-Band Interface and Digroup Terminal

The first major departure from the traditional pattern of stand-alone switches and transmission systems came with the development of the digital time-division toll 4ESS switch. Preliminary development on 4ESS began in 1968 and continued through the early 1970s. The new switch imposed many new requirements on the transmission terminals but also presented many opportunities for innovation. Appropriate arrangements for voice-frequency, analog-carrier, and digital-carrier transmission terminals, compatible with the switch and developed in parallel with it, were ready for service when the first switch was cut over in Chicago in January 1976.[31,32]

Two new terminals—the voice-band-interface frame (VIF) and the digroup terminal (DT)—performed the interface functions for analog and digital transmission systems, respectively [Fig. 19-22]. Within the switch, access to the switching network itself was via serial PCM links, each operating at 8.192 megabits and accommodating 120 voice-frequency channels encoded in standard eight-digit PCM format. The 8.192-megabit rate was chosen to exploit the speed of the available devices and was actually capable of carrying 128 PCM channels (a convenient number for binary logic—$128 = 2^7$), but eight of the channels were reserved for maintenance of the link. The remaining 120 channels formed a conveniently sized match to ten 12-channel analog groups, or five 24-channel digital digroups. The DT terminated digital facilities and performed the multiplexing and demultiplexing necessary to connect 1.5-megabit DS1 bit streams to the 8.2-megabit PCM links in the switch. The VIF ter-

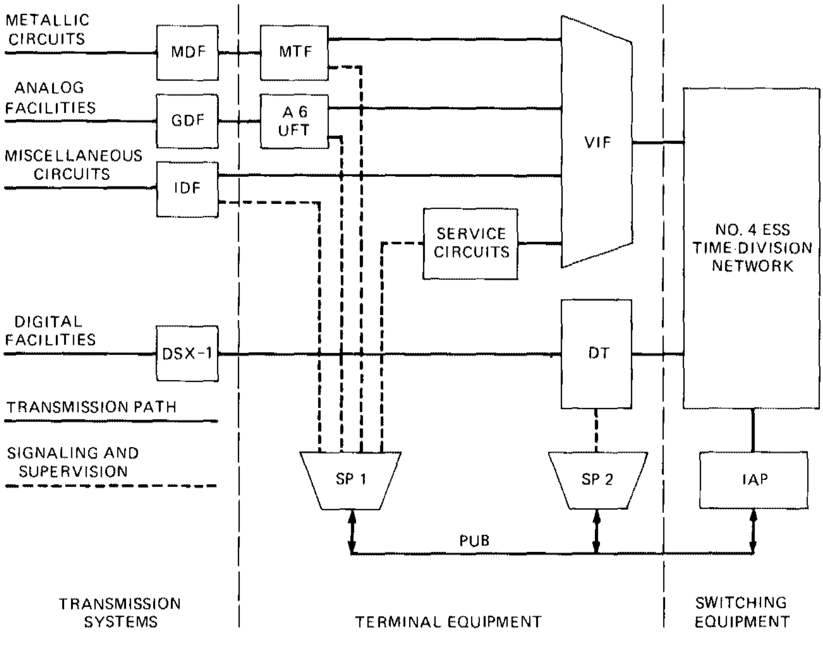


Fig. 19-22. The 4ESS transmission-switching interface system. Analog facilities were terminated in distributing frames (MDF, GDF, IDF) and conventional terminals and converted to digital format in a voice interface frame (VIF). Digital facilities were connected via cross-connects (DSX-1) and a digroup terminal (DT). All terminals were controlled from the switch 1A processor (1AP) via signal processors (SP1 and SP2).

minated four-wire, voice-frequency analog trunks and performed the analog-to-digital and digital-to-analog conversions, as well as the multiplexing and demultiplexing required to match the switch's 8.2-megabit stream.

The amount of equipment required was greatly reduced compared to that needed to connect on a trunk-by-trunk basis, as was necessary in the earlier space-division crossbar toll switches. The advantages of digital-to-digital connection was most strikingly demonstrated by the DT, where there was no need to derive each physical channel. This not only avoided the signal degradation that attended the conversions from digital to analog and vice versa, but eliminated the need for a great deal of per-channel equipment, such as signaling detectors and generators, distributing-frame appearances, and trunk circuits. (In switching-system terminology, a trunk circuit is the interface circuit [in the switching machine] that controls the sequence of actions necessary to connect the communications path established through the switch to an interoffice trunk. It is tied to and controls the signaling on the transmission side and communicates with the central control of the switch to keep it informed of the call-connection status.)

3.1.1 *Voice-Band Interface*

Termination of voice-frequency analog metallic pairs was done on a metallic terminal frame (MTF) that separated the signaling and message signals. Analog-carrier groups were terminated on an A6 unitized facility terminal (UTF) that derived the audio voice channels and separated out the signaling in a conventional way. The unitized part of the terminal designation referred to packaging all the elements (channel banks, signaling, pads, test access) necessary to terminate carrier groups in a single bay, thus avoiding a great deal of interbay wiring and distributing-frame appearances. The signaling information from both metallic and carrier channels, typically DC in the so-called E and M-lead standard format, was connected to a signal processor 1 (SP1) where it was scanned and analyzed for transfer to the switch's 1A central processor (1AP) over a peripheral-unit bus (PUB). The message signals from either voice-frequency or analog-carrier terminals were passed to the VIF for conversion to standard PCM format and transmission over the 8.2-megabit links to the time-division switching network. All these units were two-way. For outgoing signals, the VIF demultiplexed PCM signals, the SP1 distributed signaling information, and the voice-frequency and carrier-terminating frames combined signaling and message signals, as well as the inverse functions.

VIF was similar in many respects to the D2 channel bank, except that the signals were multiplexed into 120 rather than 96 channels and transmitted at the 8.2-megabit, DS120 rate to the switch instead of being sectioned into 24-channel DS1 streams, as in D2. Each input was a four-wire pure message channel; signaling, as already noted, was handled separately. The VIF consisted of seven voice-band interface units (VIUs), a switchable spare VIU, and duplicated frame controllers. Each VIU terminated 120 analog trunks and provided the interface with the switch via a DS120 port. The port consisted of a pair of coaxial cables, one for each direction, using the DS120 8.2-megabit signal format. The total capacity of the VIF was therefore 840 channels (7×120), in a seven-foot triple bay six feet, six inches wide.

3.1.2 The Digroup Terminal

The DT consisted of up to eight digroup-terminal units in a bay seven feet by four feet, four inches, with each unit capable of terminating five DS1 1.5-megabit signals and multiplexing and demultiplexing them for connection to the DS120 port of the switch [Fig. 19-23]. The DT also extracted and added signaling information from and to the DS1 streams, exchanging signaling information with a signal processor 2 (SP2) for transmission to and from the switch's central processor. In addition to the eight normally active 120-channel units, the DT had a switchable spare unit and two nearly identical controllers. The use of one-for-eight protection took advantage of the highly reliable solid-state circuitry and effected considerable hardware savings compared to full duplication. A controller supplied test signals to the units and constantly checked itself as well. When it detected faults in either a digroup unit or itself, the information was reported to the central processor in the switch. Software in the central processor diagnosed the trouble and determined the appropriate reconfiguration to be made. This information was transmitted back to the DT controller which executed the appropriate switch to substitute a spare for a failed digroup unit or the alternate controller for itself.

The full capacity of a DT was 40 DS1-level signals, or 960 channels. All

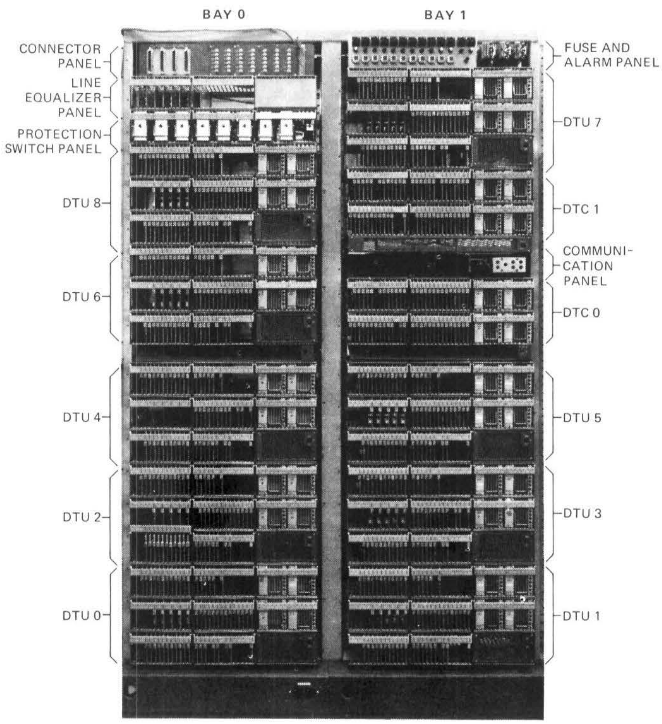


Fig. 19-23. Digroup terminal frame for terminating 960 digital channels on 4ESS switching equipment—seven-foot bays, 1976.

timing was derived from clock signals transmitted from the 4ESS switch network clock, so that the exchange of digital-signal words was done at a rate consistent with the switching function in the switch.

3.1.3 Technology and Floor Space

Both the DT and the VIF made use of the latest technology available, including printed-wire multilayer backplanes and a ceramic-plate photolithographic circuit technique developed for the 1A ESS switch. However, a large number of the VIF functions were analog, which led to the use of a mixture of techniques. The ceramic-plate 1A technology was used for digital functions, and a combination of hybrid integrated circuits (silicon chips and tantalum circuits on ceramics) and discrete components mounted on epoxy boards, for inherently analog circuitry [Fig. 19-24]. As a result, the VIF occupied about half again as much space for 840 channels as the digital terminal required for 960 channels.

This was only part of the space story. Except for small-size cross-connects, the incoming digital lines were connected directly to the DT while analog lines had first to be terminated on metallic or carrier terminals before connection to the VIF. The standard interface and nonblocking characteristics of the 4ESS switch made possible the elimination of traditional distributing frames, with their administrative and daily operational problems. (In earlier space-division switches, it was necessary to distribute connections to the switch terminals so that heavy traffic lines would not overload one portion of the switch while another was relatively idle. In rapidly changing areas, it was necessary to balance

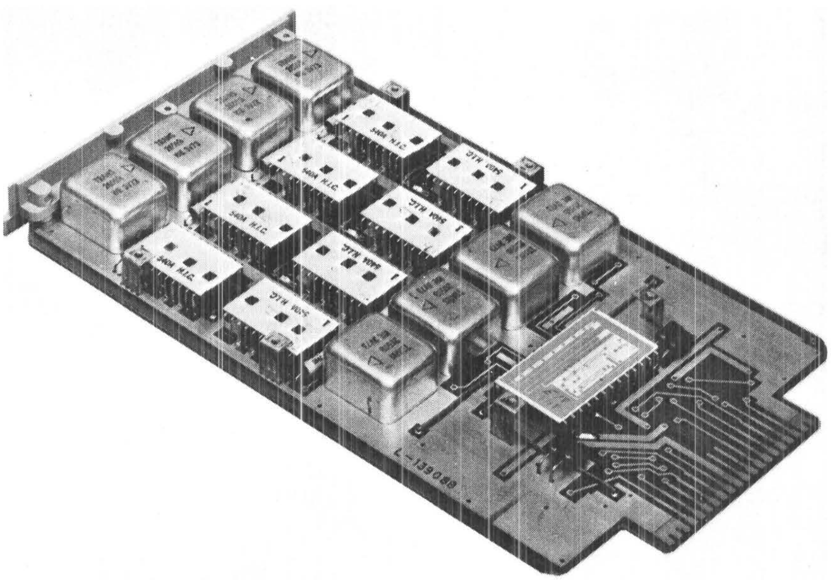
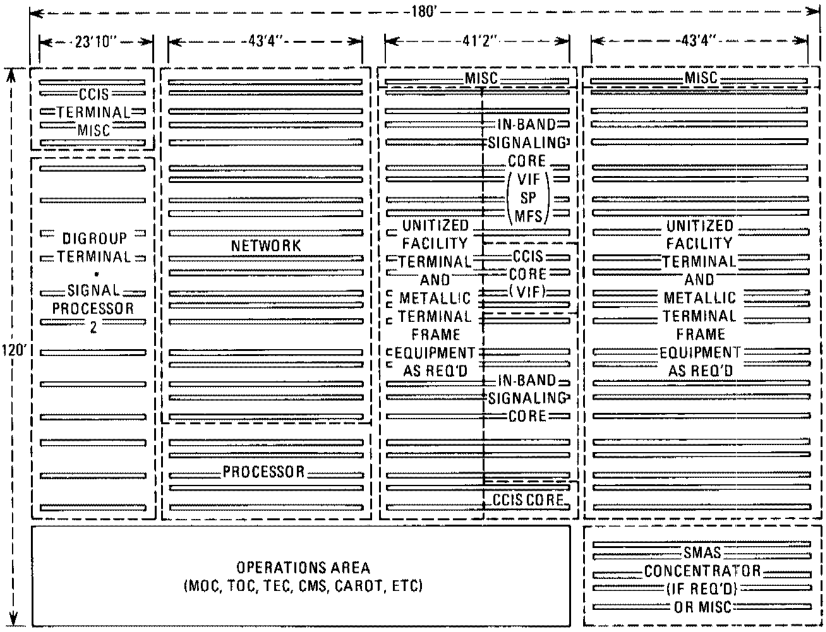


Fig. 19-24. An analog circuit pack from the voice interface unit of a 4ESS switch.



HYPOTHETICAL FLOOR PLAN LAYOUT APPROXIMATELY 82,000 TRUNKS
40,000 ANALOG (40% CCIS) 42,000 DIGITAL (25% CCIS)

Fig. 19-25. The overall floor plan of a 4ESS switching equipment office with 82,000 trunks. Most of the space was occupied by analog transmission terminal equipment.

and rebalance the switch by assigning connections at distributing frames.) With careful floor plan layouts, it was possible to connect the transmission-facility terminating bays to the VIF and switch signal processors with connector-terminated cables. In a large 4ESS switch office (about 80,000 trunks), it was estimated that connectorization and the elimination of the need for office rearrangement saved from 200 to 300 linear feet of distributing frame, over one million frame terminations and half a million cross-connect wires. With the previous methods, necessary to balance the loads on a switch, the floor-space requirement would have been about doubled. However, despite the reduction in space required for the transmission equipment, an 80,000-trunk 4ESS switch office, terminating about one-half digital and one-half analog-trunks, still consisted largely of analog transmission terminal equipment [Fig. 19-25].

3.2 The Digital-Carrier Trunk

The success of the consolidation of the transmission terminations and the inputs to the new digital toll switch, and the application of the advanced 1A central processor to the space-division local electronic switch in 1976, spurred efforts to find economic consolidations for transmission and switching in local offices as well. The outcome, completed in the late 1970s, was the digital-carrier terminal (DCT). The DCT integrated D4 channel-bank functions and the trunk-circuit function of 1A ESS switching equipment. DCT channel units performed the functions formerly associated with both the switching-system trunk circuits and channel-bank channel units. The integrated design considerably lowered the cost of terminating T carrier trunks on 1A ESS switches.[33]

In the DCT frame, signaling and supervisory information for up to 480 channels was obtained by a microprocessor-based controller that communicated directly with the 1A ESS switch central processor [Fig. 19-26]. Since signaling information was interpreted by the central controller software rather than by channel-unit hardware, a single channel-unit design was used in DCT. Various trunk-signaling formats were translated appropriately by the controller, so the single design was suitable in applications that formerly would have required more than 30 different trunk-circuit/channel-unit combinations.

As in conventional channel banks, the DCT common control separated the signaling from the voice information and routed both signals to the channel units. In the channel unit, the voice signal was converted to analog form and sent on to the switching network. The DCT channel units, however, did no detailed processing of the signaling information. The required processing was accomplished via a new common-equipment unit, the digroup controller, which in turn communicated over a data link to a peripheral-unit controller (PUC) in the switch. The digroup controller polled the channel units to detect changes in signaling information and, in the reverse direction, was capable of changing the status of outgoing signaling. The PUC sent signaling status information to and accepted orders from the switch central processor. Commands from the central processor to change a trunk state (e.g., busy or idle) or to receive or transmit address information were formatted and directed to the digroup controller over the data link. The digroup controller and PUC distributed the commands to the proper channel unit.

In addition to savings in space and cost, the DCT permitted the integration of digital line-error and fault monitors into the switch-maintenance plan. Whenever a bit error occurred, it was reported to the PUC, which kept a count of the error rate for each channel digroup. The PUC in turn generated a report to the ESS switch central control, which then forwarded it to an area switching-control center. By this means, it was possible to maintain area-wide surveillance of the health of digital lines terminating on 1A ESS switches and to take action as required.

3.3 The LT-1 Connector and Digital Interface Frame

In the late 1970s, an additional step was taken in the reduction and further consolidation of the equipment required to terminate transmission lines at the 4ESS switch. As already noted, a great deal of the equipment at a large 4ESS switch office was devoted to terminating the analog trunks. In contrast, the termination for digital trunks was compact and efficient. The strategy adopted was to bring all transmission to the switch in digital format as efficiently as possible.

3.3.1 The LT-1 Connector

The first step was to terminate voice-frequency metallic trunks in conventional D4 channel banks. In effect, it was found that, with the new banks and

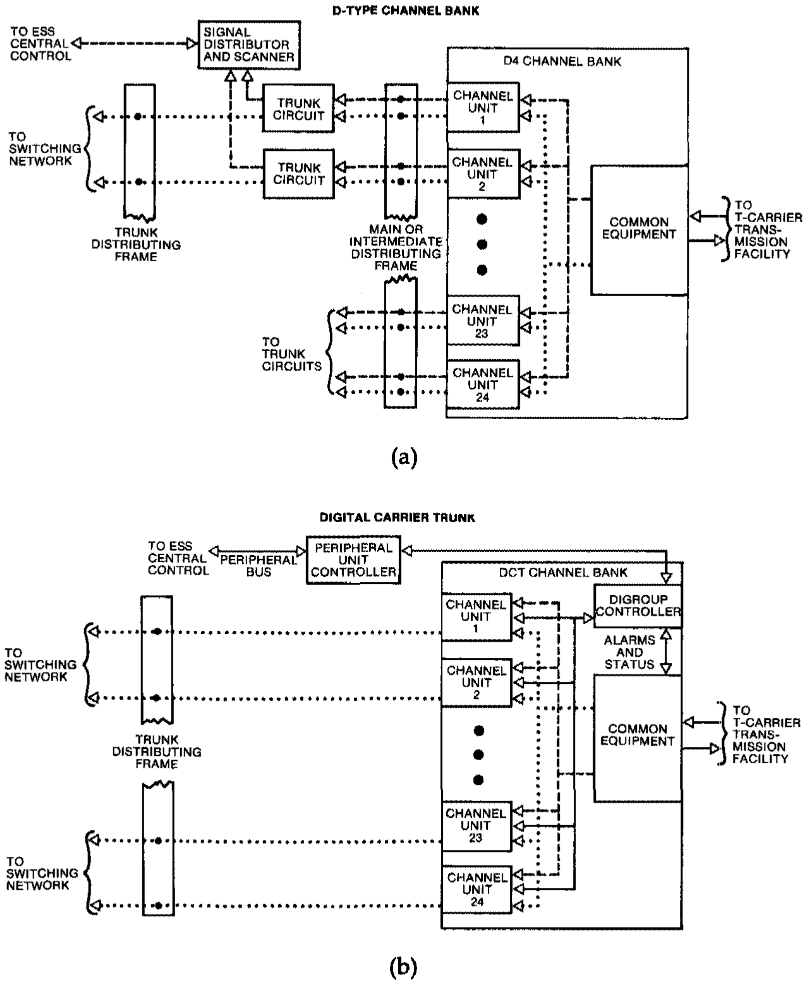


Fig. 19-26. T carrier termination on 1A ESS switching equipment, 1980. (a) Digital channel termination with conventional interface and (b) digital carrier trunk (DCT).

with the advantages attending a digital input to the switch, the prove-in distance for a digital toll-connecting trunk was zero miles. The next step was to convert analog-carrier signals to digital DS1 format as efficiently as possible. Analysis indicated that much of the cost and bulk of the A-type channel bank and subsequent analog-to-digital conversion in the VIF lay in the impedance transformations and the separate detection and generation of signaling over transmission-line type links. The conventional arrangement of bays with analog-carrier channel banks followed by the VIF occupied a great deal of space. As much as 37 linear feet of aisle space was required for the A6 unitized bays that derived the 840 voice-frequency trunks served by a VIF. Because of the substantial amount of analog voice-frequency equipment included, VIFs were also sizable, and a good deal of intraoffice wiring was required to connect the two. Through consolidation of the analog channel-bank function, signaling origination and detection, and digital encoding and decoding in a single unit—the LT-1 connector—two-thirds of the space, all the voice-frequency interbay wiring, 40 percent of the power, and one-third of the cost were saved.[34]

It was theoretically possible to process two groups of analog-carrier trunks at the basic group frequency of 60 to 108 kHz directly to the 1.5-megabit DS1 form without deriving each audio-frequency voice channel as an intermediate step. This was, in fact, accomplished in the early 1980s in the LT-2 connector. LT-2 used a VLSI digital-signal processor to operate on and transform the analog speech and signaling signals at carrier frequency to the desired DS1 format by purely digital processing. The digital processors themselves were very versatile and evolved through several generations with device counts ranging from 45,000 to 150,000 on a single silicon chip. They could be used not only for analog-to-digital and digital-to-analog conversions, as in LT-2, but could convert companding characteristics from μ-law to A-law and vice versa, a capability important in international applications.

3.3.2 *The Digital Interface Frame*

The orderly pursuit of any large-scale development requires the division of the entire project into a set of subtasks that can be carried forward without constant concern about the evolution and status of the parallel tasks. For many decades, systems development in Bell Laboratories had been functionally divided and pursued in separate areas for switching and transmission. One of the reasons for the original architecture of the 4ESS switch was the desire to preserve a clean and stable interface between switching and transmission, so that development improvements could proceed on either side without repercussions throughout both systems for every change in either. This interface, in effect, was the line between the DT and the SP2 on the one hand, and the 8.2-megabit DS120 link between the DT and the time-division network on the other. (See Fig. 19-22.)

As early as 1974, however, even before the first 4ESS switch was in service, advances in device technology had made further consolidation of the transmission and switching interface technically feasible and economically desirable. A number of different integrations were possible, ranging up to a single unit combining all the functions of transmission termination and the first stage of switching. At the same time, designers had demonstrated the feasibility of a common-control echo suppressor. The echo suppressor could operate on a multichannel digital stream and be built for a fraction of the cost per trunk of the existing per-line suppressors.[35,36] The decision was to combine the SP2 and DT function into a new digital interface frame (DIF) and to insert an echo-suppressor terminal in the DS120 link where required [Fig. 19-27].[37]

For all its virtues, the original DT/SP2 division was an awkward arrangement in several respects. The SP2 did not process DT maintenance messages but instead shuttled them between the DT and maintenance programs resident in the central 1A Processor. Communications between the terminal and the processor were slow and valuable processor time was consumed, reducing the

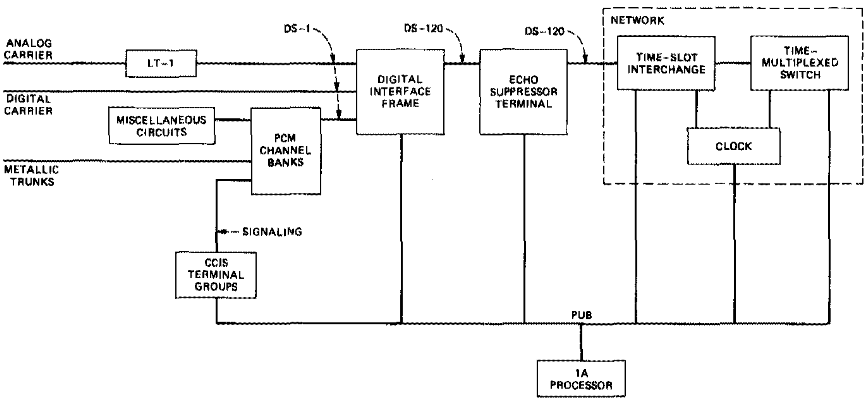


Fig. 19-27. 4ESS switch architecture with common channel interoffice signaling, 100-percent digital input, digital interface frame, and echo suppressor terminal.

amount available for larger call-serving capacity. New large-scale integrated circuits and, especially, newly available microprocessors made a different approach possible. A large part of the intelligence for testing and maintenance was placed in the combined DT/SP2, that is, the digital interface frame, and only simple and short messages were necessary between the new interface terminals and central processor. A great deal of freedom for evolution was possible, and compatibility could be maintained by software changes only.

During the period from 1976 to 1980, the 4ESS switch was virtually completely redesigned to take advantage of the dramatic advances in integrated-circuit and software technology. The ceramic-plate 1A technology was replaced by much larger-scale integration on silicon chips. The DIF and echo-suppressor terminals were important parts of that evolution. (The evolution of echo suppressors through a succession of per-circuit versions and, finally, to the echo-suppressor terminal and the single-chip echo canceler is not covered in this history but is well described in the published literature.)[38,39,40,41] With the new microprocessors and by combining the equivalent of four DTs and the SP2 into a single structure, designers eliminated a number of duplicated functions and a complex terminal-to-central-processor communication interface. The new structure was easier to maintain, occupied less than one-half the space, and required only one-third the power of the equipment it replaced.

REFERENCES

1. P. B. Cunningham, D. L. Durand, J. A. Lombardi, and R. A. Tarbox, "A New T-1 Regenerative Repeater," *Bell Lab. Rec.* **49** (April 1971), pp. 109-114.
2. R. A. Tarbox, "A Regenerative Repeater Utilizing Hybrid Integrated Circuit Technology," *IEEE International Conference on Communications*, Boulder, CO (June 9-11, 1969), pp. 46-5–46-10.
3. P. B. Cunningham and J. C. Smith, "A Low Power Regenerative Repeater for the T1 Digital Carrier System," *National Telecommunications Conference Record* **3**, Los Angeles, CA (December 5-7, 1977), pp. 44:1-1–44:1-5.

4. H. H. Henning and J. W. Pan, "D2 Channel Bank: System Aspects," *Bell Syst. Tech. J.* **51** (October 1972), pp. 1641-1657.
5. D. L. Richards, "Transmission Performance of Telephone Networks Containing P.C.M. Links," *Proc. IEE (London)* **115** (September 1968), pp. 1245-1258.
6. See reference 4.
7. E. J. Anderson, "Considerations in Selection of a Mu=255 Companding Characteristic," *IEEE International Conference on Communications,* San Francisco, CA (June 1970), pp. 7-9–7-19.
8. R. F. Purton, "A Survey of Telephone Speech-Signal Statistics and Their Significance in the Choice of a P.C.M. Companding Law," *Proc. IEE (London)* **B109** (January 1962), pp. 60-66.
9. H. Kaneko and T. Sekimoto, "Logarithmic PCM Encoding Without Diode Compandor," *IEEE International Convention Record—Pt. VIII: Digital Modulation* **11,** New York (March 25-28, 1963), pp. 266-281.
10. C. L. Dammann, "An Approach to Logarithmic Coders and Decoders," *NEREM Record,* Boston, MA (November 2-4, 1966), pp. 196-197.
11. H. Kaneko, "A Unified Formulation of Segment Companding Laws and Synthesis of Codecs and Digital Compandors," *Bell Syst. Tech. J.* **49** (September 1970), pp. 1555-1588.
12. M. R. Aaron and H. Kaneko, "Synthesis of Digital Attenuators for Segment Companded PCM Codes," *Trans. Commun. Technol.* **COM-19** (December 1971), pp. 1076-1087.
13. C. L. Dammann, L. D. McDaniel, and C. L. Maddox, "D2 Channel Bank: Multiplexing and Coding," *Bell Syst. Tech. J.* **51** (October 1972), pp. 1675-1699.
14. F. A. Saal, "D-2: Another Step Toward Nationwide Digital Transmission," *Bell Lab. Rec.* **47** (November 1969), pp. 327-329.
15. W. B. Gaunt and J. B. Evans, Jr., "The D3 Channel Bank," *Bell Lab. Rec.* **50** (August 1971), pp. 229-233.
16. A. L. DeBurro and H. A. Mildonian, Jr., "The D1D Channel Bank: Modernizing the D-1 Workhorse," *Bell Lab. Rec.* **51** (February 1973), pp. 34-40.
17. J. R. Davis, "T2 Repeater and Equalization—Design and Performance," *IEEE International Conference on Communications* **1,** Seattle, WA (June 11-13, 1973), pp. 22-11–22-13.
18. J. H. Bobsin and L. E. Forman, "The T2 Digital Line—Expanding the Digital Network," *Bell Lab. Rec.* **51** (September 1973), pp. 238-243.
19. R. L. Durham, W. G. Nutt, and J. J. Refi, "LOCAP: A Low-Capacitance Cable for a High-Capacity System," *Bell Lab. Rec.* **52** (July/August 1974), pp. 217-221.
20. R. A. Bruce, "A 1.5 to 6 Megabit Digital Multiplex Employing Pulse Stuffing," *IEEE International Conference on Communications,* Boulder, CO (June 9-11, 1969), pp. 34-1–34-7.
21. T. L. Maione, "Testing and Maintenance of the T2 Digital Line," *IEEE International Conference on Communications* **1,** Seattle, WA (June 11-13, 1973), pp. 22-4–22-7.
22. P. Brostrup-Jensen, "T2 Automatic Protection Switching," *IEEE International Conference on Communications* **1,** Seattle, WA (June 11-13, 1973), pp. 22-8–22-10.
23. J. F. Graczyk, E. T. Mackey, and W. J. Maybach, "T1C Carrier: The T1 Doubler," *Bell Lab. Rec.* **53** (June 1975), pp. 256-263.
24. J. D. Moore, "M1C Digital Multiplex," *IEEE International Conference on Communications* **3,** San Francisco, CA (June 16-18, 1975), pp. 39-14–39-18.
25. J. P. Fitzsimmons and W. J. Maybach, "Engineering T1C Carrier Systems," *IEEE International Conference on Communications* **3,** San Francisco, CA (June 16-18, 1975), pp. 39-5–39-9.
26. A. Anuff, J. F. Graczyk, J. J. Ludwick, and E. T. Mackey, "A New 3.152 Mb/s Digital Repeater," *IEEE International Conference on Communications* **3,** San Francisco, CA (June 16-18, 1975), pp. 39-10–39-13.
27. J. B. Singleton and J. A. Lombardi, "T1D System Overview," *IEEE National Telecommunication Conference,* Houston, TX (November 30–December 4, 1980), pp. 39.1.1–39.1.4.

28. W. G. Albert, W. Gaunt, J. F. Oberst, and D. C. Weller, "D4: Up-to-Date Channel Bank for Digital Transmission Plant," *Bell Lab. Rec.* **55** (March 1977), pp. 66-72.
29. M. J. Thompson, "New Developments in Digital Transmission Planning for the Bell System," *IEEE International Conference on Communications*, Minneapolis, MN (June 17-19, 1974), pp. 24A-1–24A-5.
30. R. L. Haner, T. H. Simmonds, Jr., and L. H. Steiff, "A New Interface Between N3 and L Systems," *Bell Lab. Rec.* **46** (November 1968), pp. 339-344.
31. See special issue on No. 4 ESS, *Bell Syst. Tech. J.* **56** (September 1977), pp. 1015-1320.
32. J. F. Boyle, J. R. Colton, C. L. Dammann, B. J. Karafin, and H. Mann, "No. 4 ESS: Transmission/Switching Interfaces and Toll Terminal Equipment," *Bell Syst. Tech. J.* **56** (September 1977), pp. 1057-1097.
33. R. A. Billhardt, P. J. Guarneri, T. W. Madigan, and K. D. Tentarelli, "The Digital Carrier Trunk: A 'Smart' Move Toward Increased ESS Capabilities," *Bell Lab. Rec.* **58** (November 1980), pp. 341-345.
34. S. Dodds and W. J. Mitchell, "From Analog to Digital for the No. 4 ESS—Economically," *Bell Lab. Rec.* **57** (November 1979), pp. 288-292.
35. A. A. Geigel and R. E. LaMarche, U.S. Patent No. 4,029,912; filed December 10, 1975; issued June 14, 1977.
36. R. C. Dreschsler, D. C. Koehler, and R. D. Royer, "Suppressing Echoes Digital Style," *Bell Lab. Rec.* **46** (June 1978), pp. 142-146.
37. K. M. Hoppner, H. Mann, S. F. Panyko, and J. Van Zweden, "No. 4ESS: Digital Interface," *Bell Syst. Tech. J.* **60** (July/August 1981), pp. 1131-1166.
38. P. T. Brady and G. K. Helder, "Echo Suppressor Design in Telephone Communications," *Bell Syst. Tech. J.* **42** (November 1963), pp. 2893-2917.
39. M. M. Sondhi, "An Adaptive Echo Canceller," *Bell Syst. Tech. J.* **46** (March 1967), pp. 497-511.
40. See reference 36.
41. D. L. Duttweiler and Y. S. Chen, "A Single-Chip VLSI Echo Canceler," *Bell Syst. Tech. J.* **59** (February 1980), pp. 149-160.

Chapter 20

Higher-Rate Systems: Coaxial, Waveguide, Radio, and Optics

I. BROADBAND ENCODING—VIDEO AND MASTERGROUPS

1.1 The Beginnings of Higher-Rate Systems

Early in the development work on what was to become the 1.5-megabit T1 carrier system, Bell Laboratories transmission planners and developers began working within a concept that pulse-code-modulation (PCM) digital systems, including systems of much greater capacity than T1, would evolve to a nationwide network capable of transmitting all kinds of signals over any distance. They were keenly aware of the advantages if such systems could be realized: (1) Problems of frequency response, delay distortion, net loss, and other characteristics that plagued transmission over analog systems would be greatly reduced. (2) Quality would be determined almost entirely by the terminals. The impairment due to analog-to-digital and digital-to-analog conversions could be made small and uniform for all channels. Line impairments due to digital errors and timing jitter could be made negligible. Performance would be essentially independent of distance. (3) All kinds of signals—voice, data, or video—could be encoded at an appropriate bit rate. Signals of the same or different rates could be combined into a single pulse stream with no interaction between the different types of signals. (4) Digital systems promised to be extremely efficient for the transmission of the rapidly growing computer-data traffic, already in digital form. (5) Digital transmission was exceptionally well suited to such wide-band media as the millimeter waveguide, then in development (as it was to be to the optical fibers developed later). (6) Finally, since they used much logic and digital processing, digital systems benefited enormously from the rapid advances in integrated circuits.

Investigation of the transmission of high-speed PCM digital signals over coaxial cable began in 1960. Even as Western Electric started production on the T1 system in 1962, Bell Laboratories was operating major parts of an experimental PCM system designed for a transmission rate about 150 times greater.

While a development group in Bell Laboratories at Merrimack Valley in Massachusetts addressed the problems of perfecting the 1.5-megabit T1 system for large-scale manufacture and service, groups at the Murray Hill, New Jersey laboratory started work on systems designed to meet the higher capacity and performance requirements of the long-haul transmission routes, typically served by coaxial systems and TD microwave radio. One line of development led to the D2 channel bank, M12 multiplexer and T2 six-megabit line. Another was aimed at the laboratory realization of a high-speed demonstration system suitable for transmitting broadband signals, such as television, over coaxial cable. The latter work laid the foundation for the development of high-rate digital systems for the next two decades.[1,2] Technical feasibility of the digital approach to long-haul transmission was demonstrated by this system, but it was not until the advent of fiber-optic transmission in the late 1970s that digital, high-capacity, long-haul transmission systems began to be deployed in the United States. Until low-loss fibers became available, analog techniques on coaxial cable and radio stubbornly proved more economical for the long-haul transmission of voice telephony and video. In the meantime, many of the technical advances arising from this high-rate PCM work found application in short or intermediate-length digital systems.

1.2 Early High-Speed Coder-Decoders

An early objective was to design coders and decoders (codecs) that were capable of encoding and recovering the 4-MHz+ band of a high-quality television signal or the 2.5-MHz band of a 600 voice-channel mastergroup signal from a frequency-division multiplex (FDM) terminal. Encoding television required sampling at rates as high as 12 million samples per second (12-MHz sampling) and encoding each sample into a nine digit code. (Coding for television was from a linear measure of the sample, since the amplitude distribution of a video signal did not permit the large advantage from companding realized with speech signals.) Such speed and precision were beyond the capabilities of the transistors and circuit techniques of the early 1960s, and were first achieved using a PCM coding tube that evolved from the PCM research work of the late 1940s.[3] This marvelously precise tube consisted of a cathode and electronic lens structure for generating a ribbon electron beam that was focused on a code plate [Fig. 20-1]. The signal sample to be coded was applied to the vertical deflection plates. Beam deflection was proportional to the sample voltage, and apertures on the code plate allowed current to be collected on output terminals in a pattern defining the binary number proportional to the sample. The digits, which appeared simultaneously (in parallel), could be read out in sequence as a serial PCM signal [Fig. 20-2].

The tube coder was rendered obsolete in a few years by the development of high-speed transistors. A demonstration system assembled in 1964 included both a tube coder, which coded a video signal with a signal-to-noise ratio of 62 dB (within 2 dB of the 64 dB ideal for a 9-bit coder), and a solid-state coder, with a 57-dB signal-to-noise ratio [Fig. 20-3]. Decoding was less of a problem; careful implementation of a ladder-type decoder with available diodes and a precise thin-film resistor network was adequate.[4]

A coder for 600-channel analog mastergroups with a frequency band extending to 3.084 MHz required a 6.168-MHz sampling. For a high-quality television picture, sampling at 9 or 10 MHz would have been adequate, but, for convenience, a sampling rate was chosen at twice the mastergroup rate, or

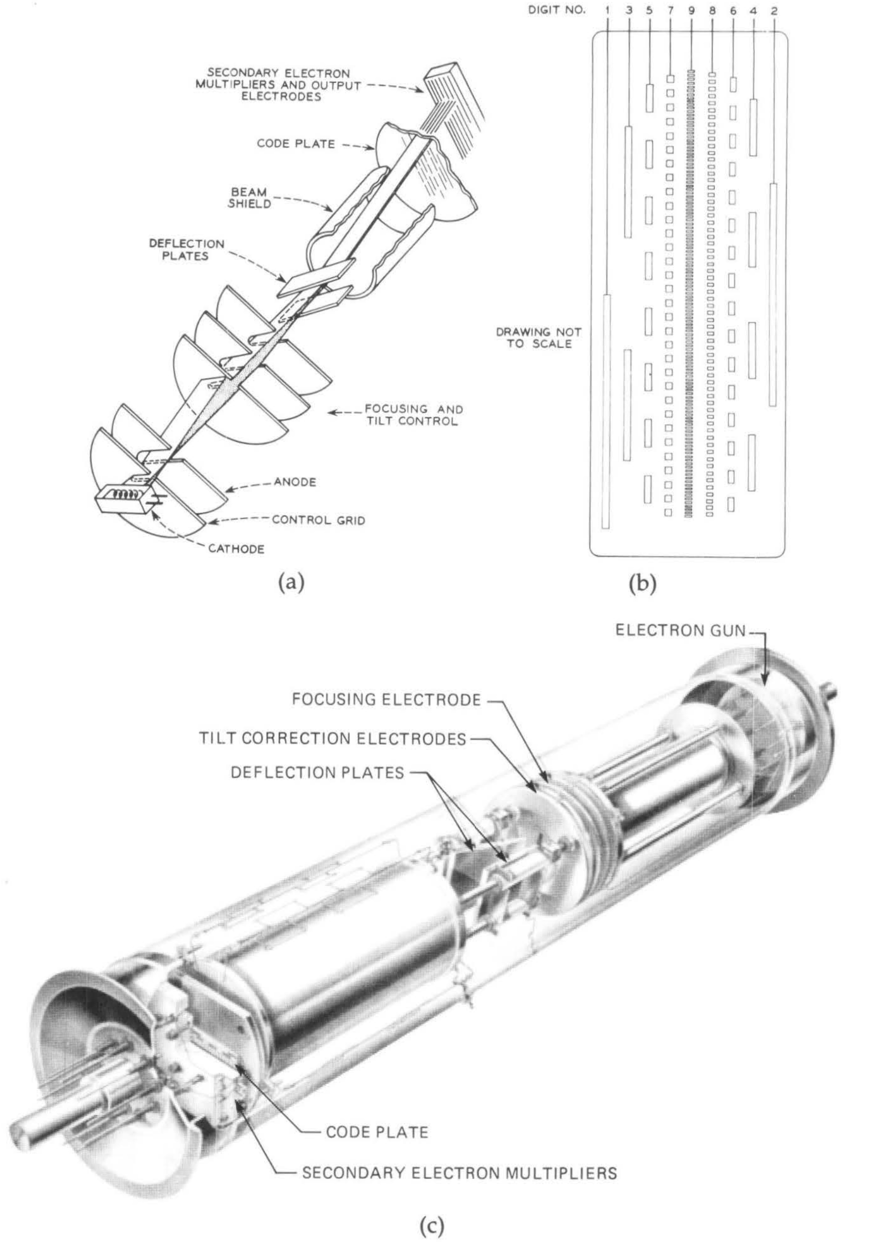


Fig. 20-1. The electron beam coding tube, 1965. (a) Schematic representation of coding tube. (b) Nine-digit code plate. (c) The experimental 12-megabit, 9-digit coding tube.

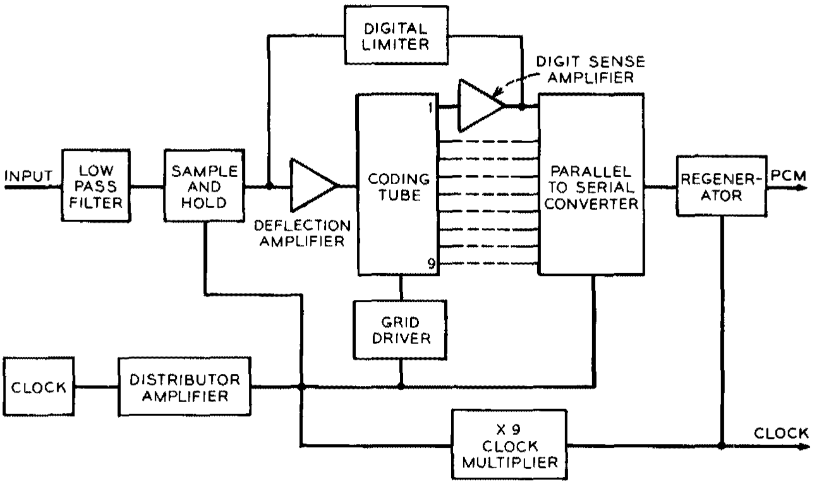


Fig. 20-2. Block diagram of tube coder for 9 digit encoding of television sampled at 12 MHz, 1965. The sample produced a 9-digit parallel code at the coding tube which was then converted to serial PCM.

12.352 MHz. The output of the 9-bit coder was, therefore, $9 \times 12.352 = 111.17$ megabits per second. In the experimental system, two of these bit streams were combined. Allowing for framing and synchronization of different input streams by pulse stuffing, this arrangement resulted in a line rate of 224 megabits per second. In the original tests, the terminals were connected back to back across the laboratory, but, by early 1966, a regenerative repeater for the 224-megabit signal had been built and tested. Ten of these regenerators were placed at one-mile intervals in a 0.270-inch-diameter coaxial line, and transmission at 224 megabits with an error rate less than 10^{-10} was demonstrated. (The small-diameter coaxial cable was used to save space in the laboratory but was also part of an investigation into alternate coaxial structures better suited to digital transmission.) This was more than sufficient to meet the total error rate of 10^{-6} judged to be needed for a 4000-mile coast-to-coast system.[5]

1.3 Evolution of Broadband Coder-Decoders

1.3.1 Mastergroup Codecs

The feasibility of digitally encoding a 600-channel analog mastergroup was also demonstrated in the high-speed demonstration system.[6] Successively improved models were designed in the following years as the device technology advanced. The last of several starts toward final development, stimulated by the expected deployment of millimeter waveguide, culminated in 1975 in a model that had all the features necessary for operation in the network, and that could have been put into production.[7] By this time, the semiconductor-device art had advanced so far that the coding process, so difficult in the early 1960s, was quite routine. Attention had turned to squeezing the coded mas-

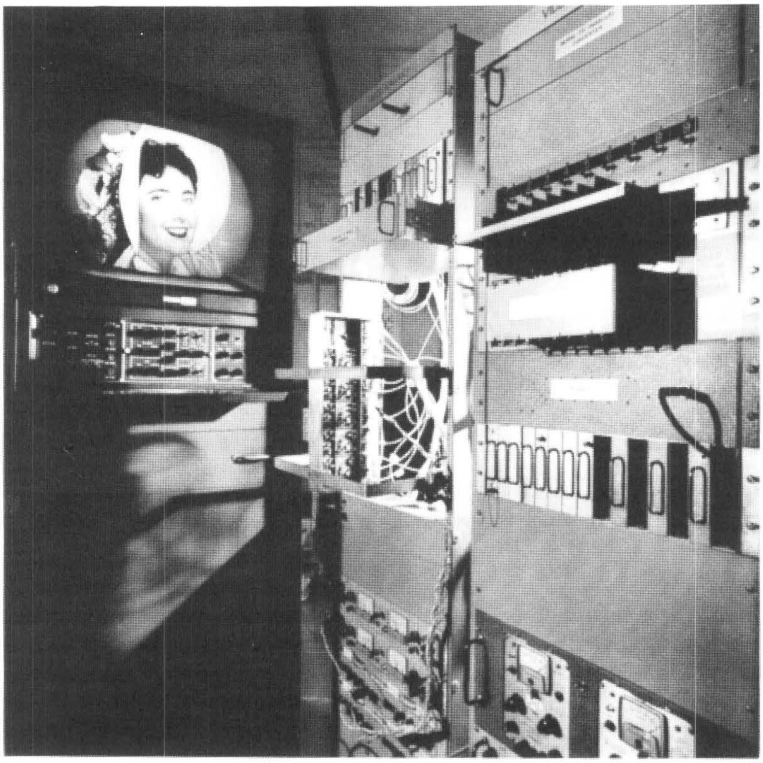


Fig. 20-3. The laboratory 224-megabit PCM system, 1966. The video coders and decoders for converting a high-quality analog color television picture to 111-megabit PCM and recovering it are the two bays at the right.

tergroup into as few bits as possible and in a format consistent with the emerging digital hierarchy. This was accomplished by shifting the frequency of the mastergroup from its normal range (0.544 to 3.084 MHz) down to 0.116 to 2.636 MHz, allowing the use of a lower sampling rate. Sampling was at 5.526 megabits, somewhat higher than twice 2.636 MHz to allow for tolerances in frequencies and components. Also, the amplitude of the mastergroup signal had an approximately Gaussian distribution with a mean of zero; low value signal samples were much more probable than high values. This made companding advantageous and a three-segment companded coder was used. The companded coding was 2.2 dB better than linear eight-bit coding, so nine-bit coding was not required. The digital output signal was at the digital-signal 3 (DS3) rate of 44.736 megabits per second, which had been chosen earlier with this coding plan in mind and was by then standard [Fig. 20-4].

This final mastergroup codec, like the earlier tube codecs, was never put into production, as the waveguide project for which it was intended was discontinued. The subsequent development of an inexpensive analog-to-digital

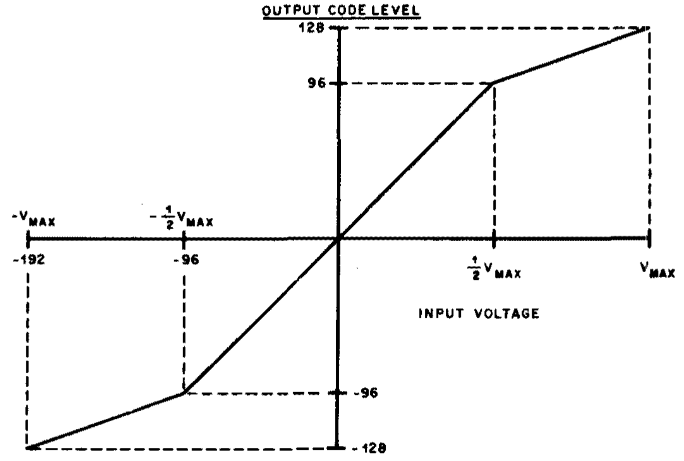


Fig. 20-4. Three segment companding characteristic for eight-digit mastergroup codec. [Andrews, *IEEE Trans. Commun.* **COM-25** (1977): 267.]

junction, the LT connector, filled the niche that the mastergroup codec was intended to occupy. The LT connector converted analog signals into digital signals compatible with the digital hierarchy, exactly as if the signals were passed through back-to-back analog and digital channel banks. This allowed signals in the 60- to 108-kHz group band of analog carrier to enter the digital hierarchy at the 1.5-megabit DS1 rate and be demultiplexed and decoded in the same way as digital signals from any other source. Thus, LT connectors had a substantial advantage over mastergroup codecs, which had to be used in pairs and so could not provide a direct connection between the two hierarchies.

1.3.2 Evolution of Video Codecs

The 111-megabit per second video codecs in the high-speed demonstration system were intended for the long-standing Bell System application of distributing broadcast-quality color video signals. This application required the extremely high quality that the codecs provided. Impairments that were introduced by the codec were almost undetectable, even to a skilled observer making comparison with studio-quality pictures and monitors. As with mastergroup codecs, technical feasibility was demonstrated, but as no long-haul digital system was deployed, there was no need in the field for television codecs until the 1980s.

The next major effort in video codecs after the demonstration system of 1964 was inspired by the plan for economical long-haul transmission of PICTUREPHONE* visual telephone service signals. Here, the need was for large numbers of channels. Transmission had to be at the lowest possible bit rate,

* Service mark of AT&T.

and the pictures were black and white and small, so some quality could be sacrificed to reduce transmission costs. This led to the development of a 6.3-megabit codec using three-bit differential PCM, which was put into service as part of the experimental PICTUREPHONE network. Differential systems, such as delta modulation, and more general differential PCM techniques were investigated, as they provided the ability to shape the spectrum of noise and could take advantage of the sample-to-sample correlation in the signal characteristic of picture transmission.[8]

Work also began on a codec requiring even a lower transmission rate, based on more elaborate signal processing techniques proposed by Bell Laboratories researchers.[9] This involved sending only the differences between successive complete video frames, instead of the conventional method of coding and transmitting each sample. Algorithms were explored and selected, methods for dealing with line errors were worked out, and a model was built when the PICTUREPHONE project was discontinued. Other manufacturers subsequently offered similar codecs and when, in 1982, PICTUREPHONE visual meeting service was instituted by the Bell System, it employed commercial codecs built along these lines. In the early 1980s, acceptable pictures were being transmitted for PICTUREPHONE visual meeting service at a 1.5-megabit rate.

II. THE T4M SYSTEM

Following the successful laboratory demonstration of high-rate terminals and lines in 1966, attention turned to the characteristics appropriate for a commercial system. Any new high-rate digital system had, of course, to be compatible with an overall plan for the digital network, but several other factors were considered as well.

2.1 The Digital Hierarchy

For several years, system engineers in Bell Laboratories had been developing a plan for a digital network that would be capable of carrying all types of signals then in service or contemplated, including voice, digital data, facsimile, television, and (important in those days) PICTUREPHONE visual telephone service. In the plan, digital signaling (DS) rates, convenient for the most frequently transmitted signals and appropriate to the transmission media, determined the hierarchy of principal digital rates. Lower-rate signals would be multiplexed together for transmission over higher-rate pulse streams. Even in the mid-1960s, the complete plan contemplated digital switching, but because of the rapid progress in digital transmission, the main emphasis was in that area.

The existing T1 and planned T2 carrier systems fit into the concept, and the experiments at higher rates were also carried out with the plan in mind. Before the rates were finally standardized around 1970, they were frequently adjusted as continuing analysis or advancing technology indicated that a change was desirable. The basic eight-digit coding of the 8-kilobit samples of a voice band signal generated the 64-kilobit signal, later referred to as the DS0 (digital signal zero) rate. Several higher rates were included in the plan [Fig. 20-5].[10]

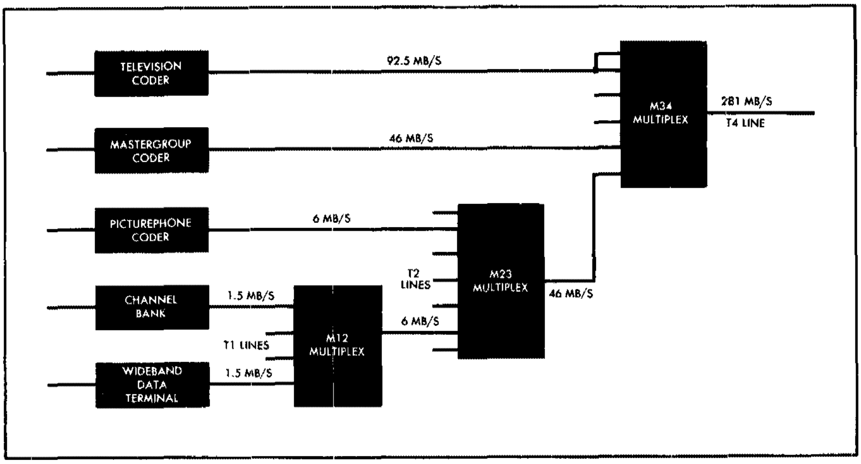


Fig. 20-5. Principal rates included in the digital hierarchy plan in 1966 (rates are rounded).

- The 1.5-megabit DS1 level was capable of transmitting 24 voice channels and was already widely deployed. The plan also included using all or portions of this rate for data.
- The 6-megabit DS2 level was capable of carrying 96 encoded channels over a T2 line. This signal could come from four DS1s combined in an M12 multiplexer or from a connecting T2 line. A PICTUREPHONE visual telephone service coder for encoding a picture of about 1-MHz bandwidth was another possible source for this level.
- The 46-megabit DS3 level was necessary for high-quality coding and transmission of a 600-channel analog mastergroup assembled by conventional frequency division multiplex (still a prospect in 1966). This level could also be generated by multiplexing seven DS2 signals in a M32 multiplex.
- The 281-megabit DS4 level was largely determined by the technical limitations of what could be transmitted over a coaxial line. This rate would be capable of carrying six DS3s or two broadcast-quality television signals coded at twice the DS3 rate, or 92 megabits. DS4 also appeared to be a convenient rate for the planned waveguide system.

These or similar rates were generally used in later Bell System digital systems. The Europeans use a 2.048-megabit rate for a 32-channel T1-type system and many other higher rates have been proposed and used. Fortunately, the diversity is not very difficult to handle in digital processing terminals and the chaos is much less than might have been expected.

2.2.1 *Multiplexing for the Digital Network*

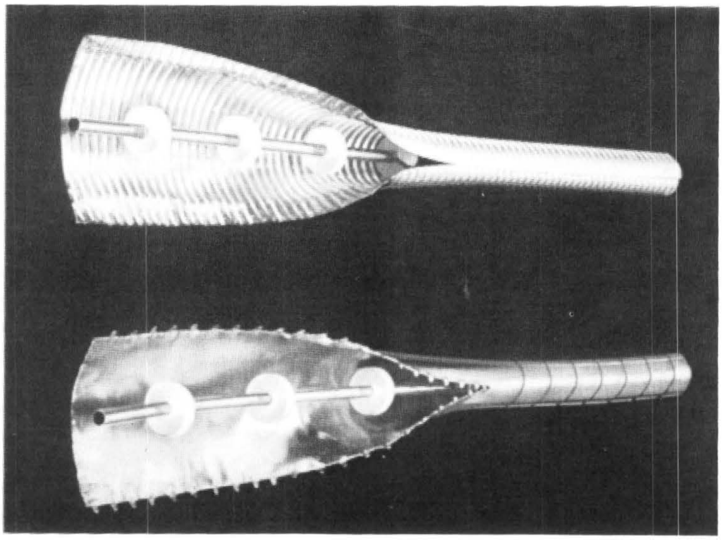
A major concern in moving to higher-speed systems was the method by which digital bit streams could be multiplexed together or exchanged between intersecting digital routes. These operations would be possible if all digital bit streams were timed from a common master clock and thus were all of the same rate (or all multiples of some common rate). While such a master-clock arrangement was later used for some applications, in the early 1960s, the method appeared potentially fragile and, in the absence of much existing digital plant, very expensive to implement. Uncontrollable and varying delays among transmission paths were also a concern. Another technique that was considered involved mutual synchronization—a number of independent clocks would adjust themselves to the average of the rates coming in to each node, thus creating a networkwide, multiloop feedback system.[11] While analysis showed that this arrangement could be made to work, it was never implemented. The technique used in the high-speed demonstration system of 1964 and in subsequent systems was pulse stuffing, similar to that used at the lower rates in constructing the 6-megabit line signal of the T2 system.[12,13] This technique allowed all clocks in the network to run entirely independently and was used in all Bell System multiplexes above 1.544-megabits per second.

2.2 Alternate Cable Structures

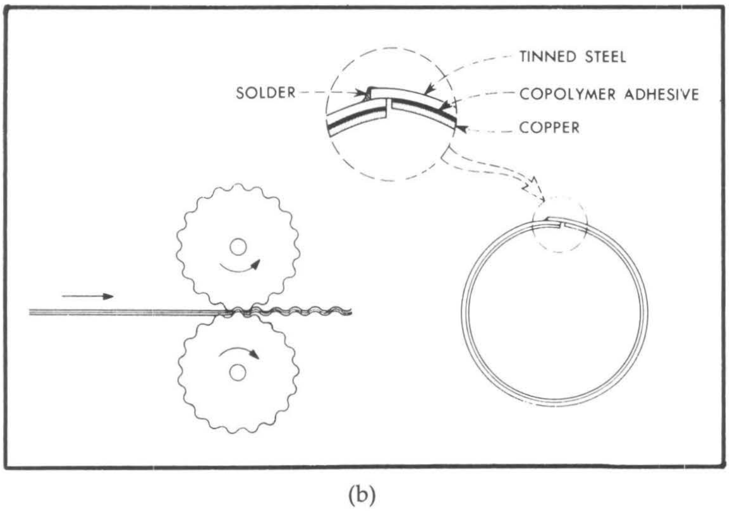
In the late 1960s, the most readily available medium for high-rate, long-haul digital transmission was coaxial cable. Digital radio was under study, but adequate frequency bands for very high-rate systems were available only at 18 GHz or higher, and these bands did not look promising for long-haul transmission. Millimeter-waveguide development was reactivated in 1969, but a practical guide was still several years away. Optical fibers were still very much in the research stage.

The study of lower-cost possibilities for coaxial-cable structures had continued ever since the first cables were produced, but the effectiveness of the 0.375-inch longitudinal-seam design, adopted in the late 1940s, was such that these studies had been in low gear for several years. In the 1960s the rising price of copper, and the awareness that very little copper was actually needed for conduction at the very high frequencies at which most of the transmission took place, reactivated the interest in a lower-cost cable. The interest was especially lively for the high-rate digital-system development, where costs were a constant concern and transmission was at rates about ten times that of contemporary analog coaxial systems. Several options were considered and experimental cables built. The 0.270-inch coaxial cable for the 224-megabit laboratory demonstration system was an experimental structure with a corrugated outer conductor, entirely of copper. The effort finally focused on a corrugated-laminated coaxial (CLOAX) cable, with a composite copper-steel outer tube, corrugated for strength and flexibility [Fig. 20-6].[14]

Several difficulties attended the cable-development effort, however. Thin copper was very hard to handle; manufacturing costs tended to offset material savings. Also, it was not practical to install a new coaxial cable that was not compatible with the existing and planned analog coaxial systems. By the time enough copper was provided to transmit the low-frequency end of the existing L4 and planned L5 systems adequately, savings tended to vanish. Finally, there was the extensive existing installed-cable network. A high-speed digital system would have to be installed on existing cables and would not be really competitive if it depended entirely on new cable installations. Work on the alternative structures was stopped and all system effort focused on the existing coaxial design.



(a)



(b)

Fig. 20-6. Corrugated-laminated coaxial (CLOAX) cable. (a) *Top*, Corrugated-laminated coaxial (CLOAX), and, *bottom*, standard serrated seam coaxial cable. (b) Fabrication of CLOAX from copper-steel laminate.

2.3 Competition with Analog

It was always recognized that, if a high-rate digital system were to be successful, its novel transmission features would not alone be enough to justify it. The existing systems did, after all, transmit all types of signals. A digital system would have to be economically competitive with analog alternatives and the competition was very tough indeed. Several factors combined to make the competitive situation difficult for a long-haul digital system. At a DS4 rate of 274-megabits a T4 digital line would provide 168 DS1s of 24 channels, or 4032 channels per coaxial pair. The analog L4 system, with 3600 channels per coaxial pair, was in service by 1967, and development was starting on the L5 design. By 1972, a successful trial of L5 demonstrated that 9000 channels per tube pair was certain, with a very good prospect for 10,800 (and ultimately 13,200). For a time, effort was directed to a 560-megabit T5 system to furnish 8000 channels, but the high rate put extremely difficult demands on the device technology, and additional economic factors were working against a long-haul digital line in any case.

Conversion, or overbuilding, of existing radio routes with single-sideband-radio (AR6A) promised to furnish channels at a cost even less than that of L5. In addition, the Long Lines Department of AT&T had so improved its methods of administration and maintenance that a significantly higher efficiency in network use was possible, making a large pool of already installed capacity available. Later, the sharp economic downturn of 1974 and 1975 reduced the growth of traffic and further deferred the need for new facilities of any kind. In this climate, the start-up cost of an analog-to-digital line conversion loomed very large. On a route nearing capacity, new L5 lines could provide 10,800 or more circuits on two coaxials and share the existing spare line and maintenance facilities. By contrast, a new digital T4 line would provide only 4032 circuits, require four tubes (two working and a spare digital pair), and a new terminal and maintenance complex. Finally, as the 1970s passed, the bright hopes for a significant amount of digital PICTUREPHONE service traffic faded. The times and technology were not ripe for a long-haul digital system.

2.4 Metropolitan Digital Coaxial—T4M

In contrast to long-haul development, T1 PCM systems in the exchange area continued to grow rapidly throughout the late 1960s and early 1970s. By 1972, more than 25 million voice-circuit miles of T1 were in service and growth was at a rate of 30 percent per year. A high-rate digital system on coaxial cable appeared to offer an attractive prospect as a metropolitan backbone system to interconnect large areas of installed and growing digital T1 systems. A major attraction of a metropolitan system was the provision of three times as many circuits per cable duct as paired-cable systems. Anything that deferred the need for new duct construction under city streets was, and has continued to be, extremely attractive to local telephone companies. Analysis of a possible high-rate, short-haul system was carried on as early as 1970. In 1972, a decision was made to concentrate on T4M (T4 Metropolitan), a system to serve as a superhighway for T1 systems in city areas, to be available in 1975.[15,16] This decision did not preclude the possibility of long-haul application.

The decision to stay with the 0.375-inch coaxial cable and a DS4 line rate of 274 megabits per second made a regenerator spacing of about one mile feasible with adequate margins. M12 multiplexers from 1.5 to 6.3 megabits already existed. Development of M23 multiplexers from 6.3 to 46 megabits and M34 multiplexers from 46 to 274 megabits had continued, but, since most of the traffic would be from T1 systems, a skip level M13 multiplexer from 1.5 to 46 megabits was also developed. Inputs at any level could be asynchronous, with pulse-stuffing synchronization. Span-terminating and maintenance terminals provided power feeds, fault detection, automatic switching to a spare line, and equipment for fault localization.

In addition to the new rate, the line signal featured an efficient two-level code. T1 carrier used a three-level bipolar plan in which binary ones were transmitted alternately as positive and negative pulses and binary zeros as the absence of pulses. The three-level plan was effective on pairs, as noted previously, but used 37 percent of the inherent information-carrying capacity of the pulse stream to obtain the desirable properties of crosstalk reduction and better timing recovery at repeaters. On coaxial cables, crosstalk was not a problem. In the T4M system, a two-level plan required more pulse processing but was more efficient, in that only two percent of the information capacity was sacrificed to obtain desirable line-signal characteristics. In another departure, quantized feedback was used to compensate for the loss of low-frequency content in the signal stream as it passed through the filters and equalizers of the regenerator. Quantized feedback was an old idea dating back to the early days of digital-transmission research. Where there was no DC signal path (zero frequency DC was used to transmit power over the coaxial center conductors to the regenerators), the absence of low-frequency response caused a tail to follow a pulse, reducing the detection margin in the following time slot. By a quite simple feedback connection, it was possible to generate an inverse tail and subtract it at the input [Fig. 20-7].[17]

The high line-signal rate and other features required a line repeater considerably larger and more complex than those used on paired cable (but far less bulky than the complement of T1 or T2 repeaters required for the same number of circuits). Precise, stable tantalum resistors and delay lines and very high-speed transistors were combined on ceramic plates in hybrid integrated circuits for critical functions. The very high frequency and consequent high line loss

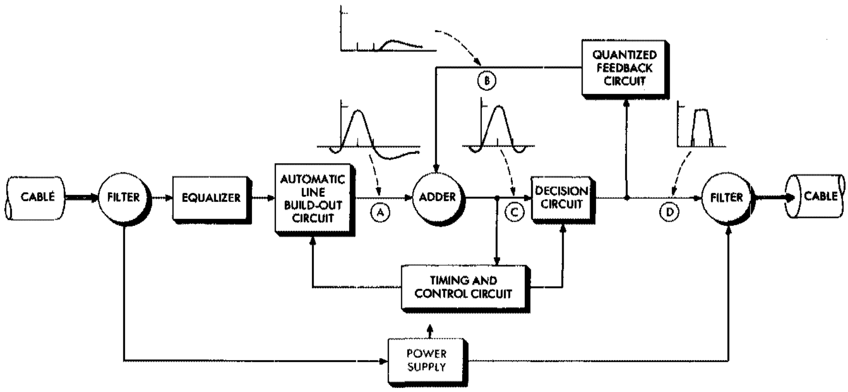


Fig. 20-7. Quantized feedback to compensate for the lack of a low-frequency transmission path in the T4M system.

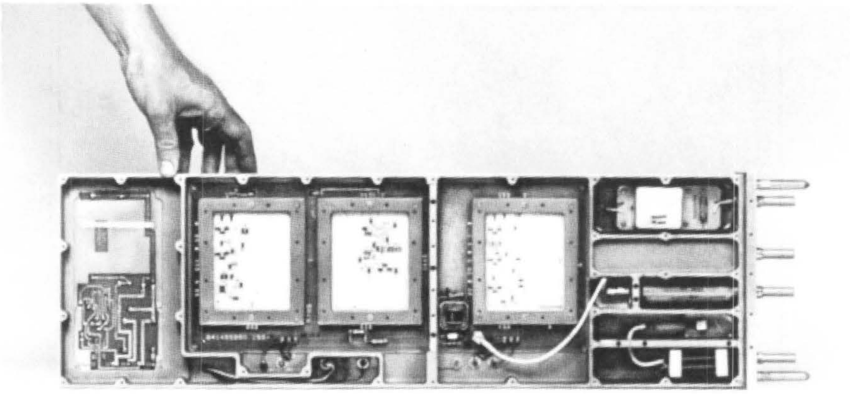


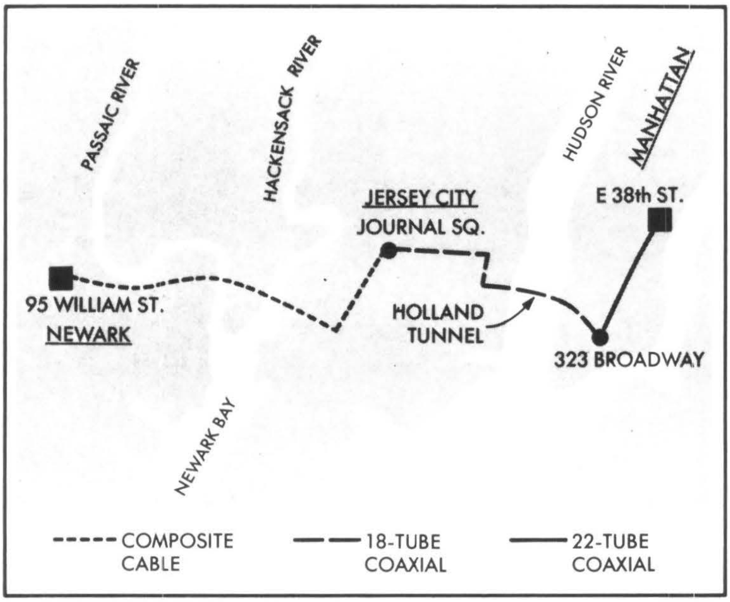
Fig. 20-8. The T4M line repeater.

(57 dB at the 137-MHz Nyquist rate) required extremely careful physical design and shielding [Fig. 20-8].

The initial installation, designed for service but also intended to serve as a field trial (the schedule did not permit the sequence of a trial followed by additional development prior to service), was between a New York Telephone Company office on 38th Street in Manhattan and a New Jersey Bell office in Newark [Fig. 20-9]. This was a high-density, rapid-growth route with duct-congestion problems—a good candidate for T4M. A working and a protection coaxial pair were equipped, along with a third pair for test purposes. Seventeen intermediate regenerators were installed on a variety of cable vintages.

The trial went well and, except for the inevitable minor difficulties, the system performed as expected. The service cutover was scheduled for April 1, 1975, but, on February 27, the New York Company was hit by one of the worst service disasters in the history of the Bell System. Fire swept through the Second Avenue office in lower Manhattan, destroying equipment worth many millions of dollars [Fig. 20-10]. About 170,000 telephones were out of service and—significantly for our story—166 T1 lines connecting New York and New Jersey were knocked out. The T4M system was rushed into service. Within hours, the first new T1 link over T4M was helping to restore service, and, within a few weeks, a substantial fraction of the lost capacity was regained over the new line. The system met the test and service was maintained with no unusual problems, thanks to the reliable equipment and careful training of the operating personnel that was under way long before the emergency need arose.

T4M was installed in small quantities in a number of other locations, but business conditions, the constant upgrading of the systems on pairs, and the failure of the PICTUREPHONE visual telephone service market to materialize caused sales and installations to languish. When interest in high-speed digital systems revived several years later, optical fibers presented a much more attractive prospect.



(a)

(b)

Fig. 20-9. The T4M trial and initial service installation. (a) The trial route. (b) Coaxial cable for T4M being installed outside the New Jersey entrance to the Holland Tunnel.

Fig. 20-10. The disastrous fire at the Second Avenue New York Telephone office, February 27, 1975.

III. MILLIMETER WAVEGUIDE

In the late 1930s, considerable work was done by G. C. Southworth and his associates to realize propagation through a circular cross section waveguide in the low-loss circular-electric TE_{01} mode. (See Chapter 7, Section 3.2.) During World War II, a great deal of additional work was done with waveguides but little on the circular-electric mode. Following the war, waveguide research on the low-loss mode was resumed in Bell Laboratories, with the hope that it would lead to the practical realization of a major new medium capable of transmitting enormous bandwidths. The frequencies necessary to realize low loss in a guide of reasonable cross section were greater than 30,000 MHz (30 GHz). The corresponding wavelengths of less than one centimeter led to the designation *millimeter waveguide* as one of the shorthand terms for the project.

Millimeter waveguide was not necessarily a digital medium. The wave transmitted could be modulated in any of the ways used on cable or radio, but the potential advantages of digital transmission were recognized very early. The work on waveguide came to be viewed as closely related to the efforts on high-speed digital transmission over other media and as the most likely solution for long-haul links.

3.1 Research in the 1950s

During the 1950s, theoretical work by S. E. Miller and his associates provided more detailed insight into the relation between waveguide loss and imperfections of waveguide geometry.[18] As it turned out, circularity and straightness had to meet stringent requirements to achieve the low loss predicted by theory. At the extremely high frequencies necessary for the propagation of the low-loss mode, many other modes can also propagate through the waveguide. It was possible to generate and launch a relatively pure circular-electric mode, but hundreds of other modes could be created by imperfections in the waveguide. The unwanted modes could propagate at a different velocity from the TE_{01} mode for some distance, and then be reconverted to the original mode, causing loss and a severe distortion problem. Miller recognized this problem in the early 1950s and suggested that an interference-resistant modulation, such as PCM, and regenerative repeaters were likely to be needed. Some consideration was given in the early days to FM and, as work progressed and a more nearly perfect waveguide was obtained, the modulation options were frequently reviewed. Some form of digital transmission always proved best.

The understanding of mode conversion was important to several aspects of the system design, including the choice of waveguide diameter. To reduce the heat loss and lower the operating frequency required in devices, a large diameter was desirable. But the number of possible modes and the mode-conversion loss became larger as the diameter was increased. A diameter of about two inches was adopted as convenient and near optimum. Designing a millimeter-waveguide transmission system with low loss thus involved trade-offs among the diameter of the waveguide, the operating frequency band, and the problems of manufacture and installation of waveguide with the required precision.

As experimental and theoretical work progressed in the 1950s, the research group at the Holmdel laboratories designed various waveguide circuit components to work at millimeter-wave frequencies. These included such items as directional couplers, wave filters, detectors, hybrid junctions, bends, and other elements essential to the realization of a complete transmission system. Fabrication of these components required meeting mechanical tolerances much closer than those in the components used in the earlier application of waveguides to communications and radar systems.

In addition to the waveguide components, a critical item was the availability of suitable signal power sources. The low waveguide loss could be realized, in theory at least, over a bandwidth of 50 gigahertz or more. It was difficult to design klystrons or magnetrons that would operate at all in the millimeter-wave region, and the devices could not be tuned over any significant fraction of the desired bandwidth. However, by the early 1950s, the invention of the backward-wave oscillator (BWO) by R. W. Kompfner provided a broadband signal source.[19] By 1956, the Bell Laboratories Electronics Research Department in Murray Hill had achieved power output from a backward-wave oscillator of about 10 mw over the range from 40 to 60 GHz. In 1957, power of a few hundred milliwatts at 100 GHz was obtained, as well as minuscule power at 200 GHz. During the same period, the research group also undertook to design

TABLE 20-1: Principal Objectives of Proposed Millimeter Waveguide System, 1958	
Capacity per Waveguide	80,000 Digital Voice Channels
Broadband Channel Rate	160 megabits per second
Frequency Range	35-75 GHz
Repeater Spacing	15 miles
Waveguide Diameter	2 inches (50.8 mm)
Waveguide Type	100 percent Helix

a traveling-wave tube of very small dimensions as a broadband amplifier for the millimeter wave range.

By the end of the 1950s, waveguide theory, components, amplifiers, and power sources were sufficiently developed by the research people that the stage appeared to be set to embark in earnest on the development of a commercially viable system. As a result, planning and engineering groups, in cooperation with the research area, prepared a prospectus for the development of a millimeter-waveguide system in 1958 [Table 20-1].

3.2 Waveguide Development (1958–1962)

3.2.1 Helix Waveguide

The early experiments with even the most carefully made 2-inch-diameter copper waveguide ran into difficulties in achieving sufficient geometrical perfection to reduce unwanted mode conversion to acceptable levels. One approach to reducing the resulting interference was to introduce mode filters that would pass the desired mode with little loss but absorb all others. An early and effective version was the helix mode filter invented by J. R. Pierce [Fig. 20-11].[20] The

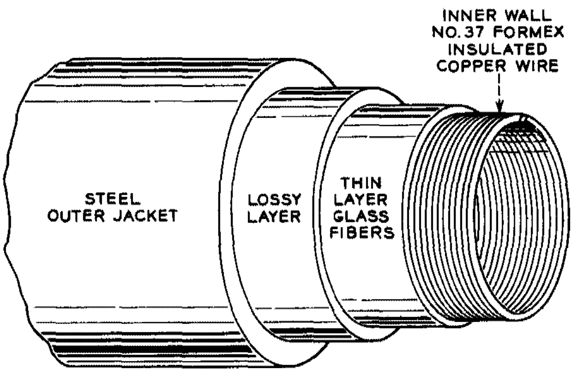


Fig. 20-11. Helix waveguide mode filter.

mode filter for a two-inch waveguide was constructed by winding a fine insulated wire in a very low pitch helix on a carefully machined two-inch-diameter mandrel. This assembly was then inserted in a very straight steel tube, and the space between the wire winding and the outer tube was impregnated with a material that was lossy for millimeter waves. After this material was cured in an oven, the mandrel was withdrawn and the ends of the filter fitted with flanges so they could be joined to other filters or waveguides. This structure had minimal effect on the TE_{01} mode, which has no longitudinal electric field at the waveguide wall. The undesired modes, however, all have substantial longitudinal electric field components at the wall, and their energy was absorbed by the absorptive material that backs the wire winding.

The mode filtration property did not eliminate the need for a high degree of straightness and circularity to keep overall TE_{01} mode losses low, but it reduced the average loss and eliminated the abrupt changes of loss with frequency that were characteristic of the plain-copper guide [Fig. 20-12]. The helix filter, in effect, constituted a new form of waveguide that favored the circular-electric mode exclusively. In the summer of 1956, two miles of helix waveguide was fabricated in the Holmdel laboratory and installed at Holmdel in a triangular loop. The installation was along a precisely aligned path in an underground conduit supported every ten feet on concrete piers [Fig. 20-13]. In this experiment, the measured attenuation varied smoothly and corresponded to about 1.5 dB per km at 60 GHz. This test established the feasibility of the helix mode-filter concept and, in the 1958 development proposal, it was to be used as the medium for the entire waveguide line.

The Holmdel helices were made in the model shop, and the installation was idealized to obtain the basic data. When system development started, facilities at Chester, New Jersey were outfitted to explore the problems of producing high-quality helix guide and to develop practical installation techniques. Helices were wound at Chester on mandrels ground to the highest achievable precision by the Battelle Memorial Institute at Columbus, Ohio. As a test of installation methods, two 1500-foot lengths of waveguide were installed at Chester in

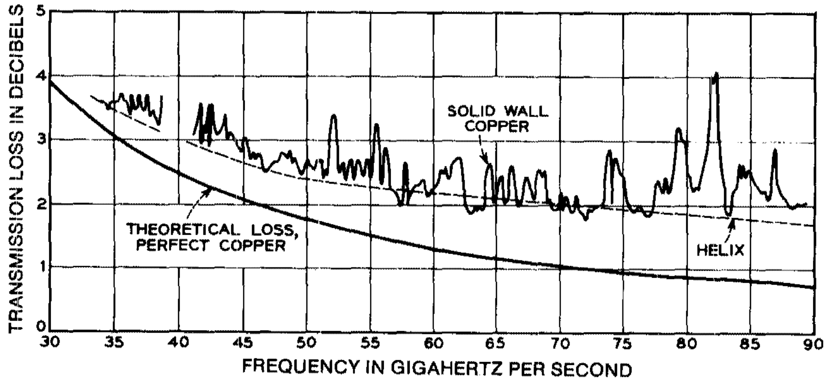


Fig. 20-12. TE_{01} mode loss versus frequency for solid copper and helix waveguide.

p. 627 is missing and will be added when available.

could be resolved by a Fourier analysis and represented as the sum of periodic functions, that is, sine or cosine waves of specific mechanical frequencies in cycles per meter and their amplitudes. A measured spectrum of waveguide mechanical distortions could then be used to compute the incidence and amplitude of mode conversion. It was concluded that deviations from straightness with wavelengths between roughly 1.4 and 4.4 feet were the significant ones for two-inch inside-diameter waveguide in the frequency band from 35 to 90 GHz.

Special instrumentation was then developed to measure the straightness of the waveguide. After many ideas were tried, a three-point curvature gauge became the favored method. A sled that fit into the waveguide was provided with carbide feet on both ends, and a deflection gauge was mounted between these feet [Fig. 20-14]. The deflection reading then was a measure of the local distortion, and, as the sled was passed through the waveguide, the variation of that distortion with distance could be obtained. From this reading, with appropriate filtering of the data, the Fourier curvature components could be deduced. Programs were written to obtain the critical parameters by processing the data in a computer. The simple curvature gauge was later expanded to measure waveguide circularity as well. Curvature measurements and Fourier analysis became the key to the understanding of manufacturing and installation processes and spawned several generations of curvature sleds, dubbed *mice*. Data from these were coupled with increasingly sophisticated computer processing techniques as the computer art exploded in the 1960s and 1970s.

3.2.3 Dielectric-Lined Waveguide

One mode in the waveguide, the transverse-magnetic (TM_{11}) mode, was particularly troublesome in deliberate bends, such as would be required along a practical route. The TM_{11} mode has a waveguide wavelength identical to the desired TE_{01} mode—they are referred to as being degenerate. In bends, the identical wavelength causes a systematic transfer of the energy from the desired to the unwanted mode. Indeed, in ordinary copper waveguide, a complete

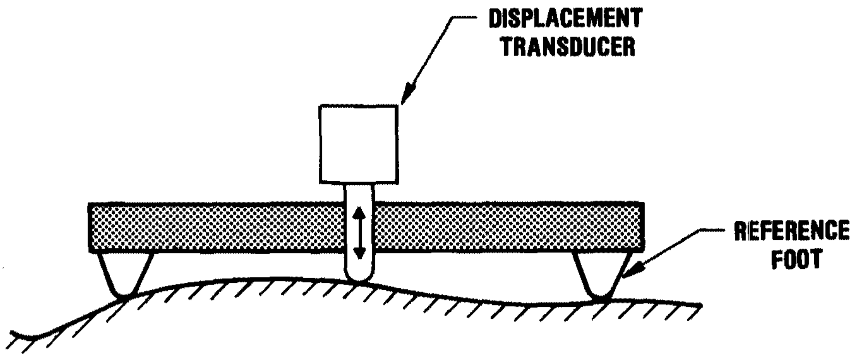


Fig. 20-14. Three-point waveguide curvature gauge.

transfer can occur. If a curve is continued, the energy will periodically transfer back and forth between the two modes with distance, suffering loss as it does so. The transfer depends only on the angle of the arc traversed and is independent of the radius of curvature. Unlike other waveguide transitions, the problem could not be reduced by gradual transitions or gentle curves. (The problem of transmitting the circular-electric mode around bends had been recognized and the coupling between the TE_{01} and TM_{11} modes analyzed as early as the late 1940s in France[22] and in Bell Laboratories.[23] In a 1952 paper, Miller had suggested breaking up the degeneracy by a special waveguide structure, as well as other possible approaches to the problem.[24])

Fortunately, not long after the invention of the helix waveguide, another type of waveguide was invented that was effective for bends, and that later became the backbone of the system designed for commercial applications. In this waveguide, a thin dielectric lining was applied to the interior of an otherwise ordinary continuous copper guide [Fig. 20-15].[25] The lining affected the two modes quite differently; the TE_{01} mode, with a very weak electric field near the wall, was almost unaffected, while the TM_{11} mode, with strong electric fields near the wall, was shifted in wavelength by the dielectric, thus breaking the degeneracy. Dielectric-lined waveguide made low-loss bends possible.

The first tests of dielectric-lined waveguide were made by coating the interior of copper waveguide with a film of paraffin. As development progressed in the early 1960s and the advantages of dielectric-lined waveguide were better understood, it was anticipated that dielectric-lined waveguide would be much cheaper to manufacture than helix waveguide. What was envisioned was a line of mostly dielectric-lined waveguide, with periodic insertion of helix waveguide to mop up other residual modes. By late 1962, waveguide development had progressed to the point where the fabrication and installation of helix waveguide had been demonstrated and the properties and excellent prospects for lower-cost, dielectric-lined waveguide were well understood. However, at the end of 1962, all waveguide system development was put on hold because of problems that became apparent in the repeater design.

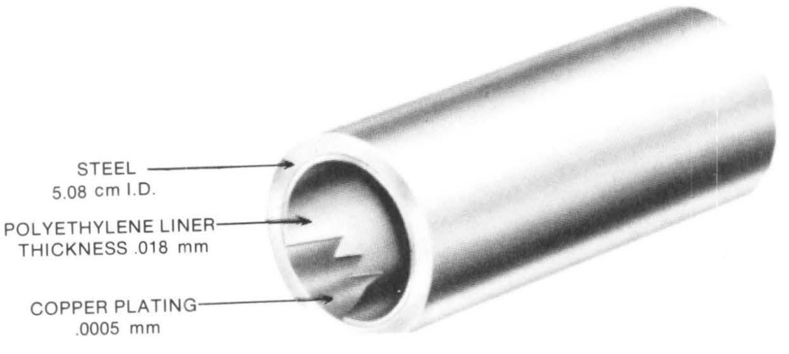


Fig. 20-15. Dielectric-lined waveguide, 1962.

3.2.4 *The Vacuum Tube Repeater*

When development started in 1958, the guided-wave research group set out to design and construct an experimental millimeter wave regenerative repeater. The repeater was of the IF type with an intermediate frequency of 11.2 GHz. Backward-wave oscillators were used as the millimeter-wave power sources and traveling-wave tubes as amplifiers at IF [Fig. 20-16]. The experimental repeater achieved low digital error rates with tube output power in the range from 10 to 100 mw, but it consumed a large amount of primary power and dissipated high levels of heat. Most significantly, the lifetimes of the vacuum tubes were much too short and prospects for extending them by a large factor were not bright. The short life would have required frequent attendance (perhaps full-time) at each repeater station to replace tubes. A large power plant and air conditioning would also be required.

The remaining problems involving the waveguide appeared solvable in a reasonable time and indeed were solved over the next several years, but a vacuum-tube repeater simply was not practical at that time. Solid-state devices were developing rapidly, but there was nothing even close to the requirements for the millimeter-wave range. In late 1962, a decision was made to change the emphasis and to resume a more research-oriented approach to waveguide work.

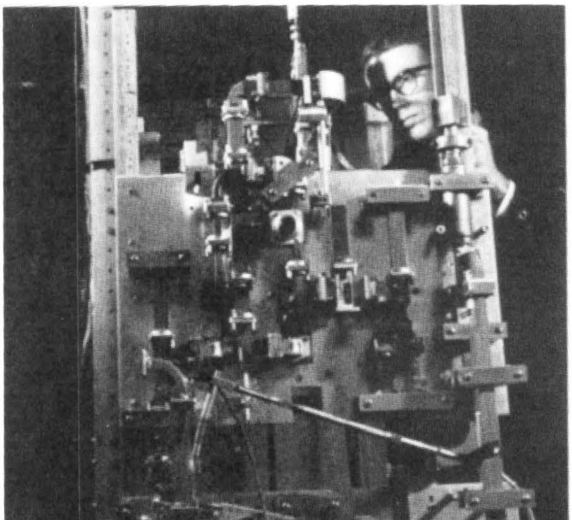


Fig. 20-16. The 1962 vacuum-tube millimeter-wave experimental repeater.

3.3 Research (1962–1969)

By 1962, the quest for a solid-state microwave source had been under way for several years. Most of the Bell Laboratories research was inspired by a 1958 proposal of W. T. Read to construct an oscillator around a specially fabricated negative resistance diode.[26] The real breakthrough, however, did not come until microwave oscillations were obtained in gallium arsenide and indium phosphide diodes by J. B. Gunn at IBM in 1963.[27] This discovery triggered intensive effort in many places to achieve microwave generation by bulk or transit-time effects in semiconductors. In Bell laboratories, effort was concentrated on limited-space charge accumulation (LSA) and impact avalanche transit time (IMPATT) diodes, models of which were realized in 1964 and 1965.[28,29] With these, the Guided Wave Research Laboratory at Holmdel undertook to develop an all-solid-state regenerative repeater for a millimeter-waveguide system.

A repeater using a form of differential phase modulation, capable of regenerating binary pulses at a 306-megabit rate, was assembled in 1966 [Fig. 20-17].[30] The circuit used both LSA diodes and an IMPATT diode with a nonlinear multiplier as millimeter-wave power sources. The IF was at 1.3 GHz in contrast to 11.2 GHz for the earlier vacuum-tube version. An error rate of 10^{-9} was achieved at a signal-to-noise ratio of 13.6 dB, only 0.6 dB above the ratio theoretically required.

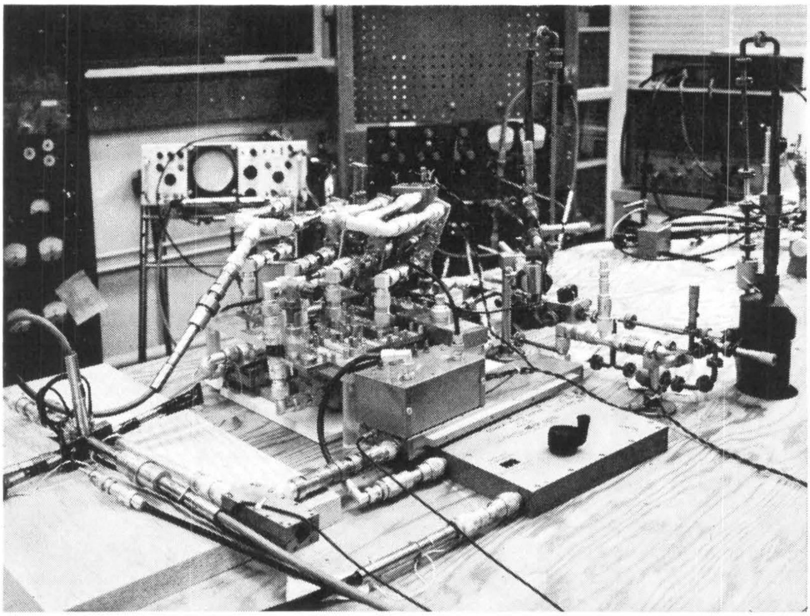


Fig. 20-17. The experimental all-solid-state millimeter wave repeater, 1966.

Meanwhile, work had continued on waveguide fabrication. Both improved helix and dielectric-lined waveguide had been built and tested. The gain of the repeater and the quality of waveguide achieved was sufficient to allow at least a 15-mile repeater spacing with two-inch diameter guide. The stage was set to resume the millimeter-wave system development.

3.4 Development (1969–1975)

Based on the successful demonstration of an all-solid-state repeater and the waveguide development, a new system prospectus was prepared in 1969 [Table 20-2]. The 1958 objectives and the performance finally achieved in 1975 are given for comparison.

By the late 1960s, the rate of growth in transmission capacity had been so large and so long sustained that forecasts anticipated a tripling or quadrupling of network capacity by 1980. It was estimated that an eventual deployment of about 10,000 miles of waveguide would be required. The developers recognized that this strange new transmission medium, which did not like curves and bends, would provide new challenges in route engineering and installation techniques in addition to the hardware design.

A special laboratory was set up in Union, New Jersey to develop the waveguide medium further, and groups were established to develop route engineering and installation methods, areas that had been addressed only lightly in the previous efforts. The systems-design and repeater-design work was at Holmdel, New Jersey and device development at Allentown, Pennsylvania. The Merrimack Valley laboratory developed the millimeter-wave filters. The Western Electric Engineering Research Center constructed a pilot plant at Forsgate, New Jersey to develop manufacturing technology and to produce waveguide in quantity for a planned field evaluation experiment.

Thus began a large team effort to make the millimeter wave system a commercial reality.[31] Concurrent work on systems in Japan, England, Germany, France, Italy, and the Soviet Union heightened the interest and competition and produced international symposia, twice in England and once in Japan, devoted solely to guided millimeter-wave systems.

TABLE 20-2: Millimeter-Wave System

	Development Objectives		Performance Achieved
	1958	1969	1975
Voice Channels per Guide	80,000	233,000	238,000
Broadband Channel Rate (mb/s)	160	282	274
Frequency Range (GHz)	35–75	40–110	38–104.5
Repeater Spacing (miles)	15	20	31–37
Waveguide Diameter (mm)	50.8	50.8	60
Helix Waveguide (percent)	100	5–10	0.6

3.4.1 Renewed Medium Development

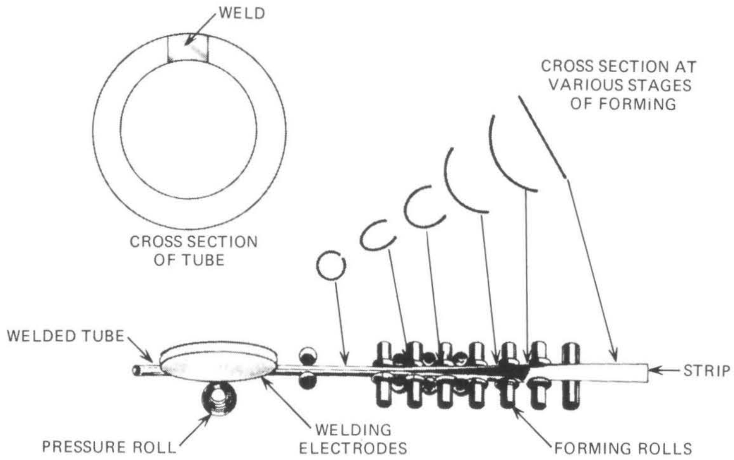
By 1968, it was clear that dielectric-lined waveguide would be much cheaper to manufacture than helix waveguide. A major effort was, therefore, launched to develop the required technology. This included manufacture of very straight steel tubing, low-loss copper plating, dielectric-liner bonding, and electrical and mechanical measurement techniques suitable for commercial operation.

Very straight steel tubing is a specialty item in the steel industry. It is usually manufactured by forming a flat strip into a tube with a welded seam that is then progressively drawn and finished to the desired dimensions. The curvatures imparted by the unevenness of the drawing process are corrected by plastic deformation in a roller straightener [Fig. 20-18]. By ordinary standards, the resulting tubing was quite straight. Unfortunately, each step of the process tended to leave a characteristic periodic signature, and periodic deformations were the prime source of mode-conversion losses. It was therefore necessary to state tolerances in terms of the mechanical Fourier spectrum rather than by the conventional test of rolling a tube on a flat table and measuring the bow. A special mouse gauge coupled to a minicomputer was designed to calculate the Fourier spectra on the spot and display them on a plotting board with a mask showing the limits at each frequency. This was probably the first application of the Fourier theorem as a routine, factory, mechanical quality-control inspection tool. The spectrum plot was a very useful trouble-shooting tool, often pointing directly to the source of a periodic deformation.

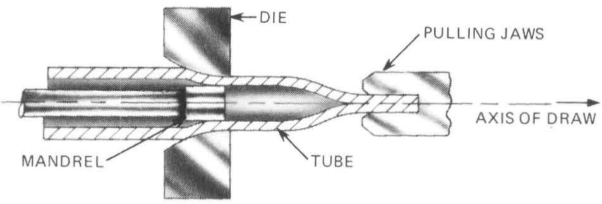
The new instruments and minicomputer coupled to a modern automatic-plotting board produced something of a culture shock in the steel plant. The ability of outsiders to spot instantly specific difficulties in proprietary process steps appeared to be black magic. Nevertheless, with the complete cooperation of the supplier, and with only relatively minor changes, mostly in the roller straightener, tubing of extraordinary perfection was produced.[32-35]

The copper electroplated on the interior of the tubing had to be uniformly deposited with a fine, dense grain to approach the conductivity of pure solid copper. Methods were developed at the Forsgate pilot plant that provided close control of plating-chemistry, temperature, and current. Application of the dielectric liner proved to be a troublesome problem. The liner had to be extruded with a high degree of physical uniformity, with uniform and stable dielectric constant, and had to adhere firmly and reliably to the copper plate for many years, even in the event of accidental flooding. After a number of different approaches, the most successful method was to inflate a high-purity irradiated polyethylene bag inside the tubing and then heat the entire assembly to 300 degrees C to melt-bond the liner to the waveguide securely [Fig. 20-19].

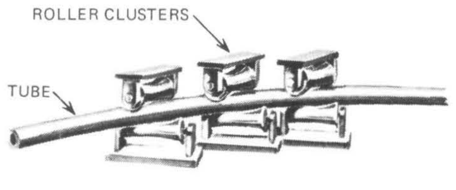
Early measurements on the finished waveguide showed an amazingly high conductivity. The loss achieved was lower at millimeter-wave frequencies than had previously been achieved with high-purity solid copper under ideal laboratory conditions. After much head scratching and analysis, the conclusion was that the superb conductivity was due to a combination of the superior plating techniques and a phase change in the copper crystal structure, which came as a by-product of heating the waveguide during the liner-bonding operation.[36] (This was a rare and notable exception to a general pattern. After



(a)



(b)



(c)

Fig. 20-18. Manufacturing process for steel tube. (a) Tube forming and welding. (b) Cold drawing. (c) Roller straightening.

long experience and refinement, commercial products often surpassed the performance of laboratory prototypes, but getting early production up to design intent was usually a tough struggle and a very important phase of development. The waveguide conductivity was perhaps the only instance in this history where first production exceeded laboratory performance.)

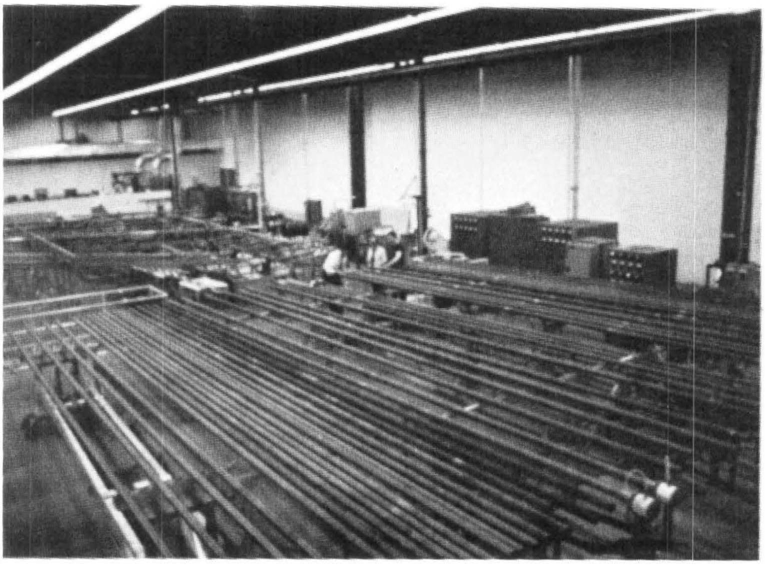


Fig. 20-19. Plating and liner bonding area of the Forsgate, New Jersey millimeter waveguide pilot plant. The guide was fabricated in 30-foot lengths.

It was expected from the earlier waveguide work that an inside diameter of 2 inches (50.8 mm) was about optimum for a band from about 40 to 110 GHz. A larger diameter would reduce copper-dissipation losses (most important at lower frequencies) but increase mode-conversion loss (controlling at higher frequencies) [Fig. 20-20]. What was sought was a waveguide in which the balance of heat loss and mode-conversion loss gave the lowest average loss. The extreme straightness achieved in the steel-fabrication process reduced the mode-conversion loss. The dielectric waveguide manufactured from the superior-quality steel tubing had mode conversions so small that helix mode-filters were required only every half mile or so, a long way from the 100-percent mode-filter design of 1959. Even though the copper heat loss had been reduced compared to earlier waveguides, mode-conversion loss had been reduced even more. To take advantage of these developments, in 1971 the designers increased the waveguide diameter from 50.8 mm (2 inches) to 60 mm, lowering the loss at the low end of the band and raising it at the high end. At that same time, the band was narrowed slightly to 66.5 GHz from 38 to 104.5 GHz. The loss ranged from about 1 dB per km at the band edge to 0.5 dB per km at the center. The change in size of the waveguide and the extremely low losses made possible an increase in the repeater spacing from the original target of 20 miles to as much as 37 miles in country of gentle relief and 31 miles in rugged mountainous terrain. (Despite the fact that the project was always referred to as millimeter waveguide, the development group, which included mechanical and civil, as well as electrical, engineers, was still in a stage of metric transition. Wavelengths were always stated in metric units.

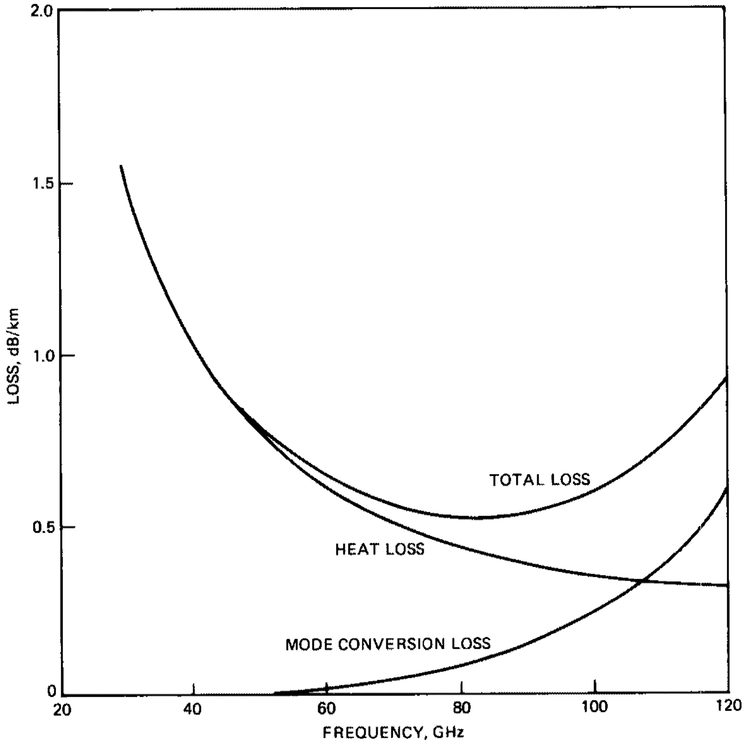
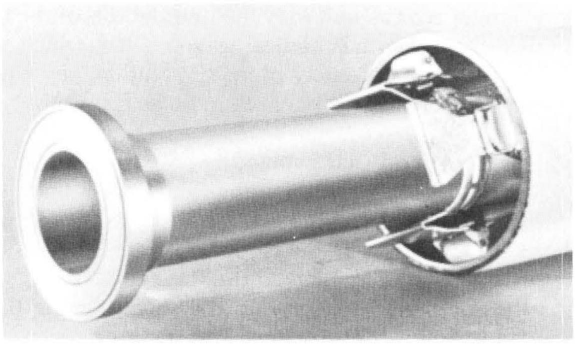


Fig. 20-20. Waveguide loss components, 1975.

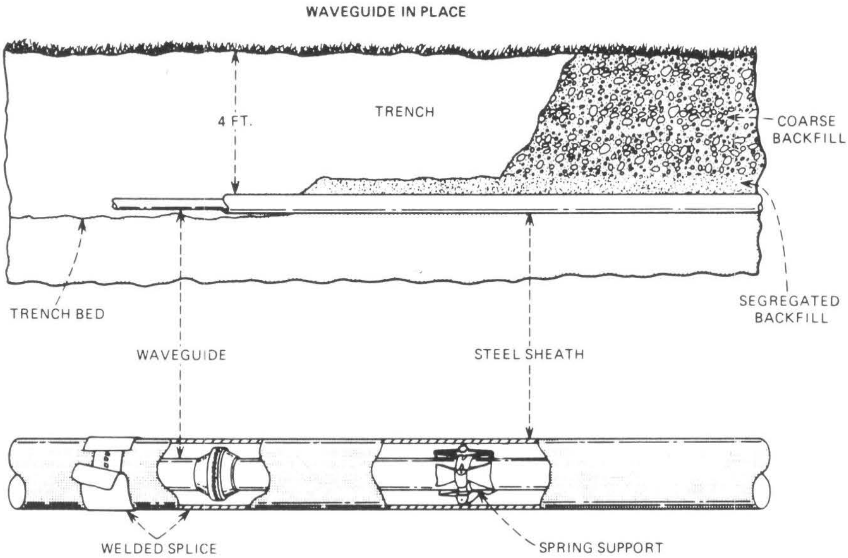
Waveguide diameter and unit losses might be metric or English, but common steel-pipe size, trench depths, and repeater spacings seemed more natural in English inches, feet, and miles. The mixed system used at the time [with lapses] seems appropriate for an account of the work.)

3.4.2 Installation, Route Engineering and Measurements

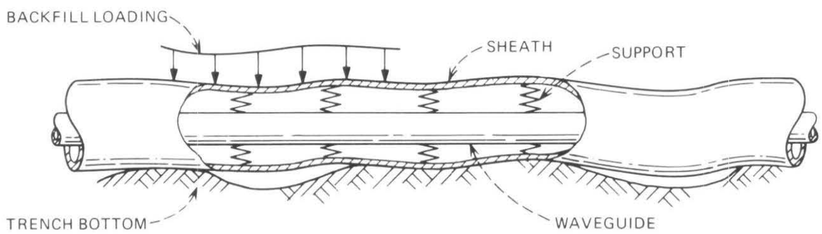
Straightening the waveguide tubing in a roller straightener did not affect that part of the mechanical spectrum and the resulting mode-conversion losses contributed by trench and terrain irregularities. In early work, an attempt was made to excavate a very smooth trench using a large bucket-wheel excavating machine with the depth of penetration controlled by a laser-beam reference. But the waveguide had to be very good initially and stay that way—and the long-term effects of direct burial were worrisome. The Chester experiments had shown the effects of burial and demonstrated the stabilizing value of a conduit. The designers decided to insert the waveguide into a steel sheath 5-1/2 inches in diameter with the guide mounted on spring-roller supports [Fig. 20-21]. The stiffness and compliance of the sheath-spring-waveguide combination acted as a mechanical filter, effective in eliminating undulations for mechanical wavelengths up to about 20 feet. The quality of the installation



(a)



(b)



(c)

Fig. 20-21. The millimeter waveguide support and sheath design. (a) Butterfly spring-roller supported waveguide in 5-1/2-inch inside-diameter steel sheath. (b) Buried sheath and waveguide. (c) The sheath and spring support formed a two-stage mechanical filter.

trench was no longer critical, and only the undulations of the terrain and route bends controlled the longer mechanical wavelengths, which were, in general, not critical. This was an extremely practical solution, since the requirements on trench evenness were less stringent than those for ordinary sanitary sewer-installation.

The installation method that evolved required placing many miles of the sheath in a four-foot-deep trench, joining individual sheath sections by conventional pipeline-welding techniques. An installation machine, which was almost a portable factory, was developed to insert the waveguide into the buried sheath. The installer first joined the waveguide flanges by welding and then pushed the completed section into the sheath, using the rollers on the filter springs to lower insertion friction. The friction was low enough so that waveguide-insertion points in the sheath could be several kilometers apart. With the low-loss guide and the mechanical filters supporting the waveguide in the sheath, the shortest radius of curvature that could be negotiated without excessive loss was 250 feet. The acquisition of right-of-way for the field evaluation test was acquired within this constraint.[37]

For the installed waveguide to meet all requirements, a variety of special mechanical and electrical test methods and equipment was developed. These items ranged from simple gauges to mice with horizontal and vertical curvature-deflection gauges and rotating heads to measure both curvature and circularity [Fig. 20-22]. The most elaborate was a long-distance mouse for installed waveguide. This assembly was an articulated train over 16 feet long. It contained a battery-powered propulsion unit that could be assisted by the application of nitrogen pressure against a piston face in case it should get stuck [Fig. 20-23]. The long-distance mouse contained an 8-kilobit telemetry transmitter-receiver operating at 80 GHz, and used the waveguide under test to transmit the control and test data. Data from the mouse gauges was fed to a transmission-measuring system with a built-in minicomputer that fully automated operation and data management. Accurate measurements were obtained from short lengths in the laboratory to the entire 8.7 miles of the field experiment.[38]

3.4.3 *The Repeaters*

Since the highest-rate regenerators could operate only up to a few hundred megabits per second, the system plan involved splitting the 66.5-GHz band into 124 subbands of about 500 MHz each, 62 for each direction of transmission. (The subbanding provides some perspective on why waveguide was so exciting a prospect; each of the 124 subbands was as wide as the entire radio band at 4 or 6 GHz.) Splitting was accomplished by passive waveguide filtering arrays, called multiplexing networks. The entire band was first divided into seven major parts by cascaded high-pass/low-pass filters, each part containing from 12 to 21 subbands. The individual subbands were then separated within each of the seven major bands by channel diplexers, analogous in function to microwave-radio channel-dropping and channel-combining networks [Fig. 20-24].

Each 500-MHz channel was initially to carry a 274-megabit, DS4-level, binary bit stream, and plans included later upgrading to four-level modulation

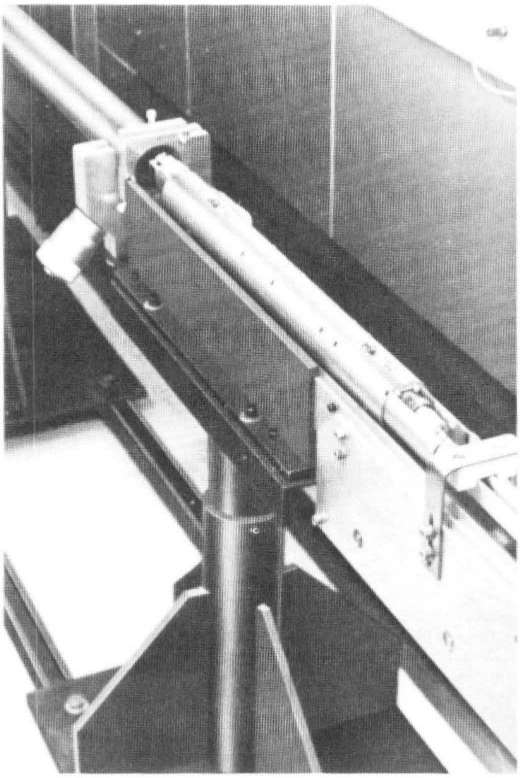


Fig. 20-22. Rotating-head mouse to measure waveguide circularity as well as vertical and horizontal curvature.

for an effective 550-megabit rate. Provision was also made for housekeeping functions such as protection switching, fault location, and order wire connections. Three channels were allocated to protection switching to cover channel failures and maintenance routines. The two-phase binary-system design, fi-

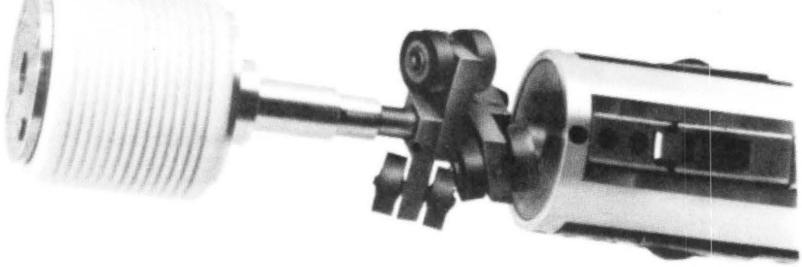


Fig. 20-23. Propeller and pneumatic piston of long-distance mouse.

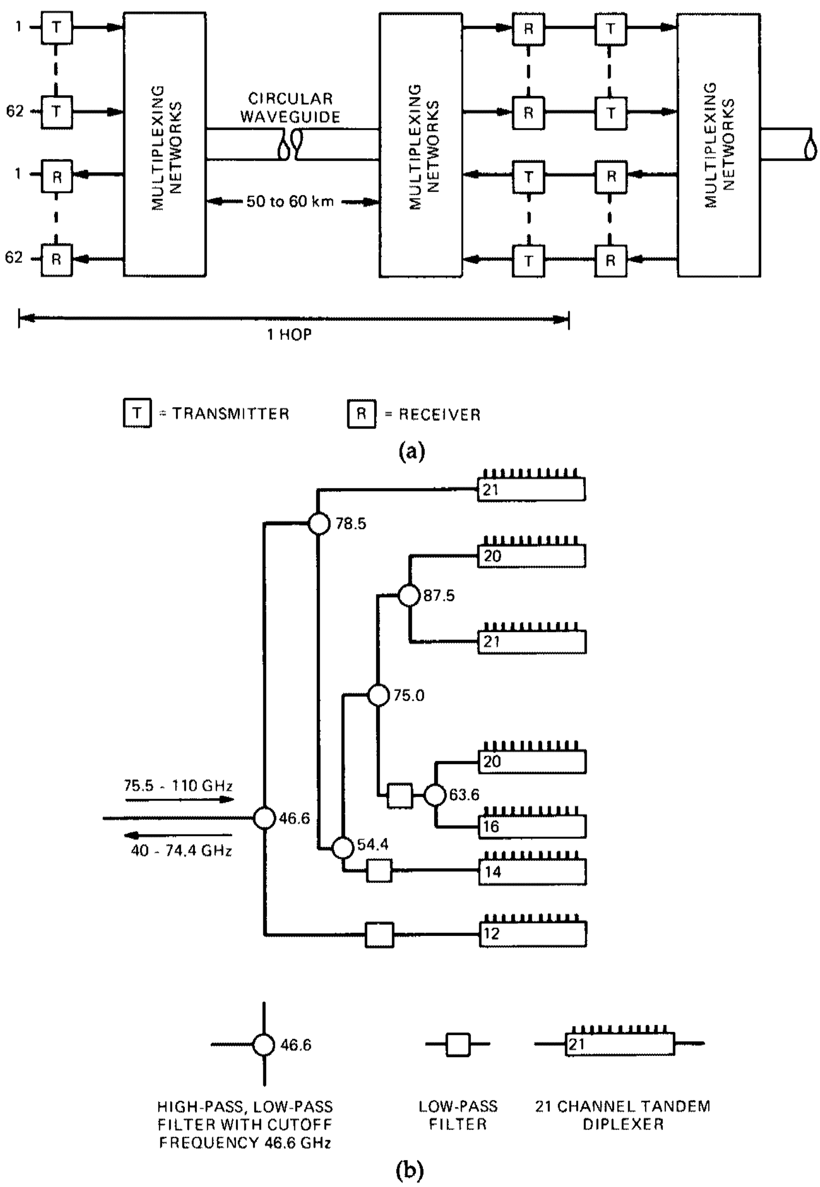


Fig. 20-24. The millimeter-waveguide band splitting and combining plan. (a) Block diagram of a repeater hop. (b) Multiplexing network plan.

nalized in 1975 and called WT4, could provide 238,000 4-kHz voice channels. The physical arrangements were such that, when new repeaters using four-phase modulation became available, the capacity could be doubled to 476,000 4-kHz channels by a plug-in substitution of repeaters in a successor WT4A system.

Many aspects of the repeater station posed new challenges. Despite the short wavelength, the waveguide multiplexers were large, since short-radius bends were forbidden, and long tapers had to be used to avoid the excitation of undesired modes when reducing waveguide cross sections for the diplexing filters. The final design was a factory-assembled frame 43 feet long, 8 feet high, and 1 foot deep, to which the repeaters were secured [Fig. 20-25]. Heat, the nemesis of the original vacuum-tube repeater, had not completely disappeared as a problem, even with all-solid-state repeaters. Water was circulated through the mounting frame to cool the repeaters.

The base-band DS4 signal was available at the regenerator of each repeater, although most of the gain was furnished at the IF of 1.3 GHz [Fig. 20-26]. Frequency control and phase-locking circuits operated at IF. This permitted the millimeter-wave IMPATT transmitting oscillator to be free running, a major circuit simplification. The two-level modulation of the continuous wave output was accomplished by 180-degree phase shifts in a path-length modulator. The effective electrical length of a reflecting stubline was changed 180-degrees by opening or shorting a diode across a printed-circuit reflection line.[39]

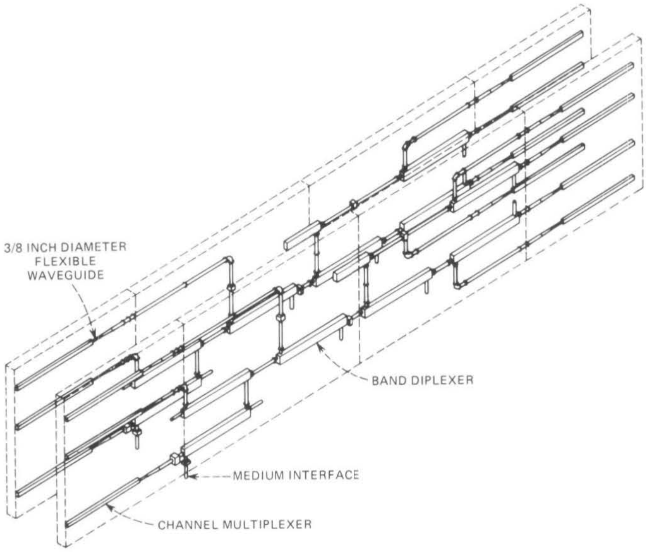
In spite of the success of the all-solid-state research repeater of 1966, production of reliable devices presented formidable problems. At millimeter wave frequencies component dimensions became very small. The active area of the IMPATT diode was only 0.001 inch square [Fig. 20-27]. Within that area, it had to handle a heat flow of 2w, a power density of 2 megawatts per square inch. To remove the heat, the IMPATT wafer was bonded with gold to a heat sink made of diamond, since diamond is both an electrical insulator and an excellent heat conductor. The entire oscillator assembly, in turn, was mounted with a short thermal path to the water-cooled mounting frame [Fig. 20-28].[40] (From time to time, economy-minded Western Electric managers had accused Bell Laboratories of delivering gold-plated designs. There is no report on record of their reaction to the gold-plated diamond heat sink.)

3.4.4 Field Evaluation Test

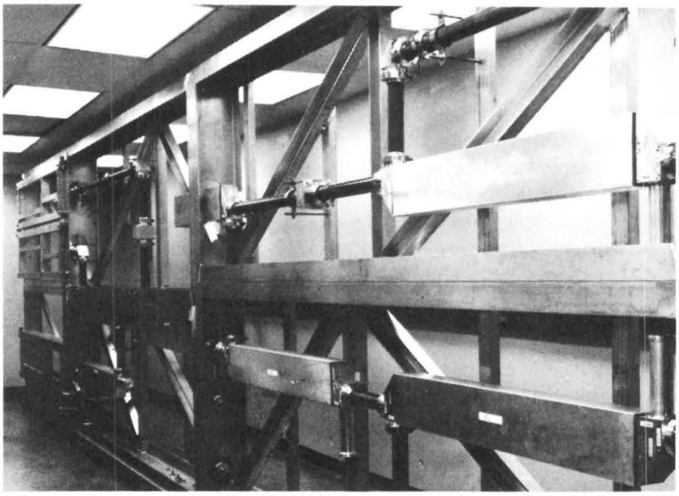
A field evaluation test was conducted in northern New Jersey starting in 1974. Its objective was to evaluate the various manufacturing, installation, transmission-performance, and service features of the system in a simulated commercial environment. A 14-km (8.7-mile) waveguide line was installed in New Jersey between the AT&T metropolitan-junction station in Netcong and a temporary station in Long Valley. The waveguide was fabricated to manufacturing specifications in the experimental pilot plant at Forsgate, New Jersey by Western Electric Engineering Research Center and Kearny Works personnel.

The terrain along the waveguide run offered a variety of construction situations, including route bends, road and stream crossings, a steep grade, rocky terrain that required blasting, and swampy areas. The sheath was installed by commercial subcontractors under the direction of AT&T Long Lines. Long Lines craft personnel installed the waveguide, using the tooling and equipment developed by Bell Laboratories and collaborating commercial suppliers. The viability of the installation and maintenance methods on a commercial basis was thoroughly demonstrated.[41,42]

First, the sheath was welded and buried in a trench on the right-of-way



(a)



(b)

Fig. 20-25. The WT4 waveguide system channel multiplexing network. (a) Schematic of channelizing networks. (b) Diplexer components in four-section aluminum frame.

that was acquired by standard real-estate acquisition practices. This demonstrated that the minimum-curvature right-of-way requirement of the waveguide route could be managed at an affordable cost. A waveguide installation-train,

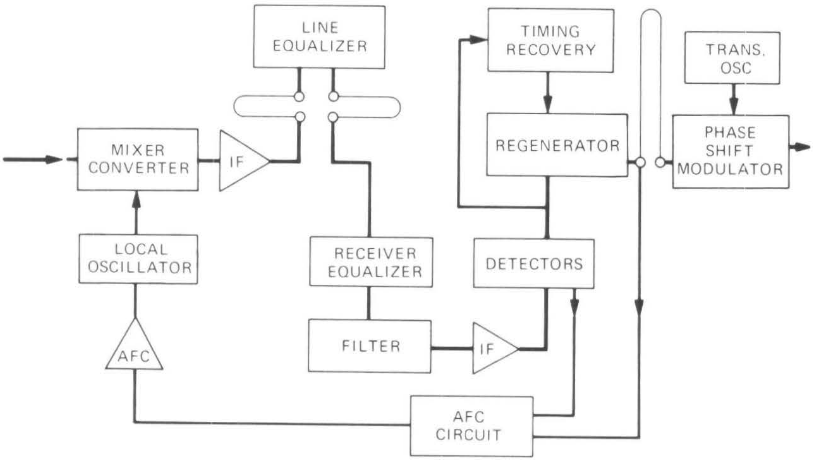


Fig. 20-26. WT4 repeater block diagram.

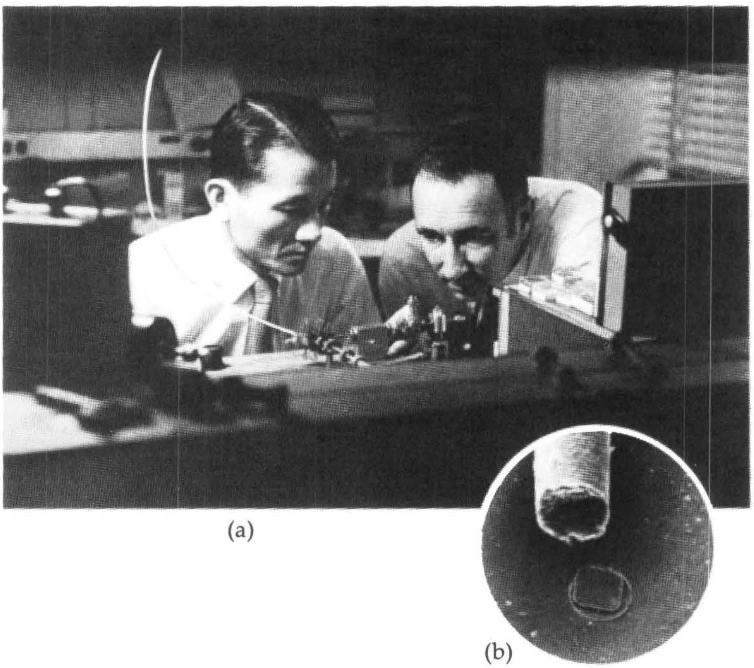


Fig. 20-27. The millimeter-wave IMPATT diode. (a) IMPATT oscillator laboratory assembly. (b) The diode compared to a human hair.

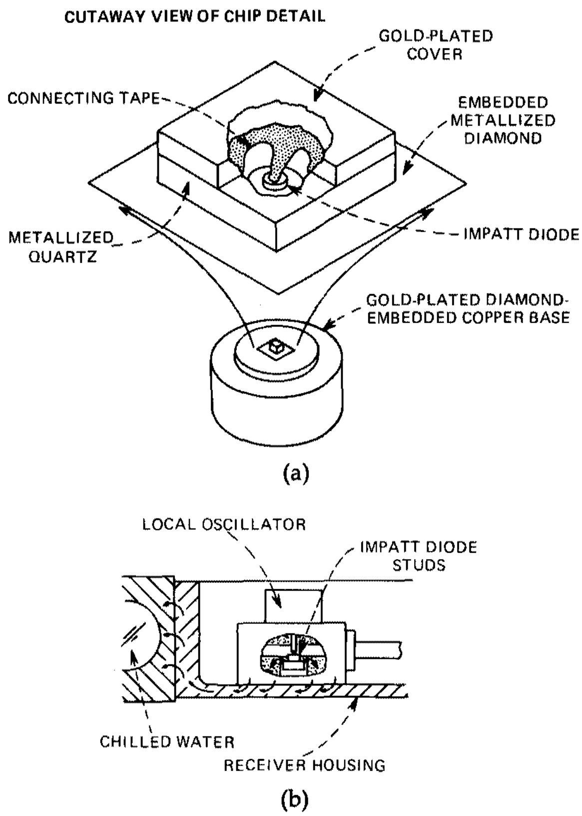
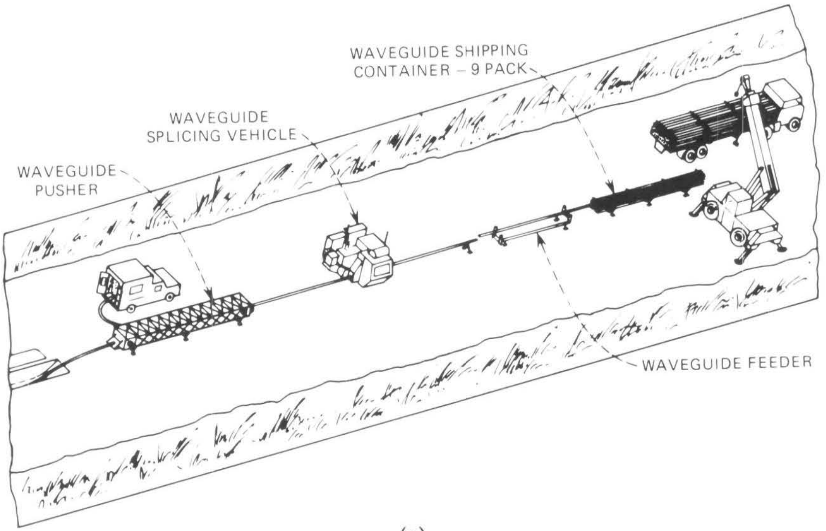


Fig. 20-28. The millimeter-wave IMPATT diode. (a) Diode structure and package. (b) Thermal path to the water-cooled frame.

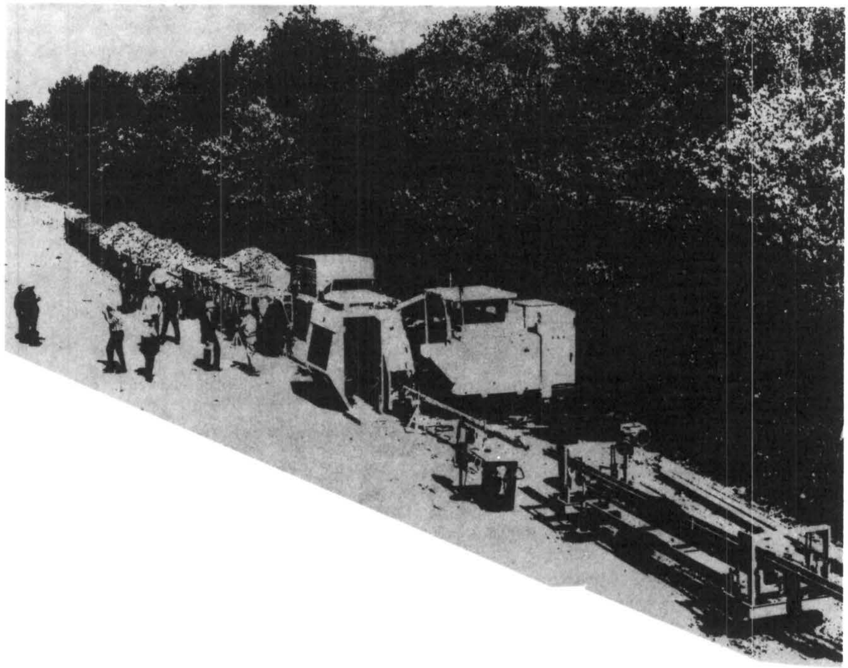
consisting of an automatic waveguide splicer-welder mounted in a cab on a Caterpillar* tractor and a pushing fixture, followed [Fig. 20-29]. It was demonstrated that 6 km (3.75 miles) of waveguide could be inserted in the sheath by pushing from a single insertion location. The entire length of waveguide was installed from only six such push points. The guide consisted of 17 mode-filter sections in which a single 9-meter section of helical mode filter was installed in approximately 800 m of dielectric-lined wave guide.

Complete band diplexer modules were installed in both terminal stations, but only 12 frequencies were fully equipped with channel-dropping networks and repeaters. Maintenance equipment included means for gas-leak localization in both the sheath and waveguide. An in-trench welder-splicer was tested at the Chester laboratory as part of the repair procedure in the event of a rupture or excessive damage to the line. The tests showed that the operation in the

* Trademark of Caterpillar Tractor Co.



(a)



(b)

Fig. 20-29. Field evaluation test waveguide installation. (a) Waveguide installation train. (b) Installation of field evaluation waveguide. *Right,* the unloading rack lined up the 30-foot sections for welding in the splicing cab (*center*). *Left,* behind the splicing cab, the hydraulically-driven chain pusher grasped a section of waveguide by the flange and pushed it into the buried steel sheath.

field was essentially the same as in the laboratory and that the system met all objectives [Fig. 20-30].

3.4.5 *Epilog*

The field evaluation test was the culmination of the long effort over the decades to establish the technical and economic viability of the high-capacity, circular-electric mode, guided-wave transmission system. It was demonstrated that the system could provide long-haul digital transmission at a cost substantially lower than competing systems in a period of rapid growth on newly constructed routes. Unfortunately for its developers, growth slackened in the middle 1970s, and, when new facilities were again needed, a new and even more interesting and versatile waveguide was available.

In 1962, looking back over 40 years of interest in and research on low-loss waveguide, Southworth had written, "Almost from the first, however, the possibility of obtaining low attenuations from the use of circular-electric waves, carrying with it, at the same time, the possibility of extremely high frequencies and accordingly vastly wider bands of frequencies, appeared as a fabulous El Dorado always beckoning us onward."[43]

Like earlier El Dorados, this one had always seemed to recede beyond the next horizon. Finally, with realization in view, Southworth's millimeter-wave

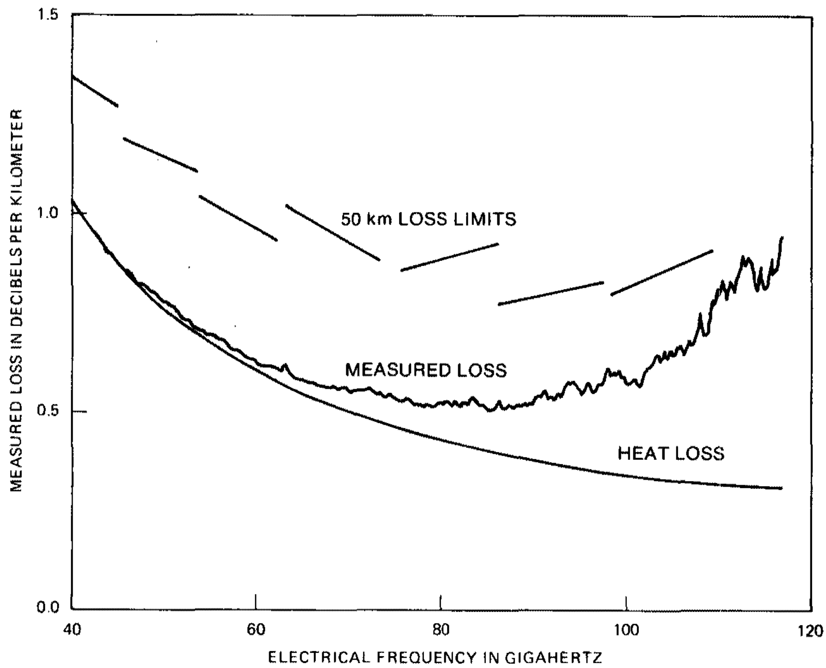


Fig. 20-30. The WT4 millimeter waveguide field evaluation test, 1975. The broken lines show the loss limits in the seven major subbands for 50-km repeater spacing.

El Dorado had vanished, only to reappear as the optical-fiber, micrometer waveguide.

As far as is known, no millimeter waveguide was installed for commercial telecommunications. However, the work was far from in vain and left a legacy for the future. The knowledge gained in the devices and techniques for high-speed digital systems was invaluable when attention turned to the light-wave medium. The insights gathered on the behavior of multimode guided-wave propagation were important elements in the successful development of low-loss optical fibers. Many of the same research people who developed the understanding of the millimeter waveguide made Bell Laboratories a leader in light-wave technology.

Millimeter waveguide also left a legacy to radio astronomy. During the 1970s the National Radio Astronomy Laboratory constructed a very large array (VLA) radio telescope in Socorro, New Mexico [Fig. 20-31]. This array consisted of three adjustable legs in the shape of a Y, each up to 21 km long. Along each leg, nine large parabolic-dish antennas were deployed to collect the celestial radio signals. In order to realize the array resolution, the antenna units had to be connected to a central detector and computer by very broad band links with precisely controlled delay. The signals were transmitted over 60-mm TE_{01} waveguide. Bell Laboratories provided some consultation to the National Radio Astronomy Observatory on the design of the transmission lines, but, by the time the waveguide was needed, the pilot plant at Forsgate had been shut down, and the waveguide for the antenna links was manufactured in Japan. When the waveguide project was disbanded, the Bell Laboratories stockpile of waveguide was donated to the VLA project to help it probe the universe.[44,45]

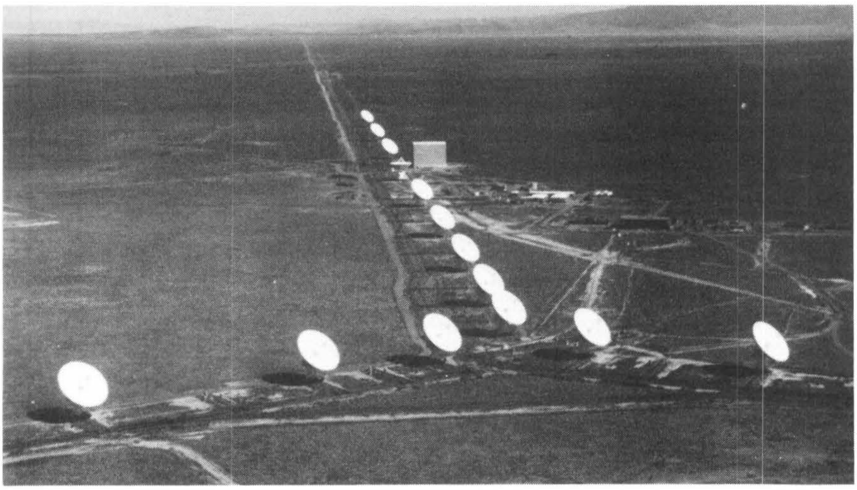


Fig. 20-31. The National Radio Astronomy Laboratory very large array, Socorro, New Mexico. Signals from the antennas along the legs of the array wire transmitted to a central processor by circular-electric mode millimeter waveguide.

IV. DIGITAL RADIO SYSTEMS

As the use of digital systems on pairs grew rapidly in the late 1960s, many studies were conducted to determine the best strategy to develop towards a nationwide digital network. The expected need for many six-megabit links for PICTUREPHONE visual telephone service was an important consideration in the planning. One response was the development of guided systems—the 6.3-megabit T2 system on low-capacitance pairs and the work on high-speed digital coaxial and waveguide systems. Another was a decision to develop digital radio, both as modifications of existing systems and in wholly new designs.

The early experimental needs of PICTUREPHONE service were met by the development, started in 1967, of digital terminals capable of sending three 6.3-megabit digitally-encoded video signals, multiplexed to a rate of 20 megabits over existing TD2 radio channels. In 1969, development started on DR18 (digital radio-18 GHz), a system that would carry the DS4 rate of 274 megabits in the 18-GHz common-carrier band, where the necessary broad channel bandwidths were still available. DR18 was intended not only for PICTUREPHONE visual telephone service but to carry large bundles of T1 carrier in metropolitan areas and to be a digital feeder for the future millimeter-waveguide system.

The need for an intermediate-rate, digital radio facility was also recognized in the early 1970s, leading in 1974 to a decision to develop an 11-GHz DS3-rate (45 megabit) digital radio system, the 3A-RDS. It was intended to serve the Bell System operating companies for intertoll trunk applications, especially for the interconnection of islands of T1 carrier, spaced 30 to 100 miles apart.

4.1 Digital Terminals for TD2 Radio

When PICTUREPHONE visual telephone service was first planned, it was clear that, at least initially, the signals would have to be carried over existing plant. By the late 1960s, the TD2 microwave radio network had reached every major population center. It was, therefore, the logical choice for early long-haul intercity transmission.

The 1-MHz video telephone service signal was transmitted in analog form over customer loops and short trunks but was coded into a 6.312-megabit digital format for longer trunks. The picture was not coded in the standard PCM format. The 1-MHz-bandwidth picture, sampled at 2 MHz and coded by an 8-bit word, would have required 16 megabits and would have been prohibitively expensive to transmit. By taking advantage of the high sample-to-sample correlation in the picture signal (i.e. it did not change very much or very rapidly), it was possible to obtain an acceptable picture with a three-digit code word and a 6.3-megabit transmission rate. The technique was to subtract each sample from the succeeding one in a difference detector. Only eight discrete levels could be coded with three digits, but by concentrating these near zero signal amplitude for accurate coding of the small difference signal, a reasonable picture could be reconstructed [Fig. 20-32].[46]

In 1967 development started on terminals to combine three of the 6.3 megabit binary picture signals into a 20.2-megabit four-level signal for transmission over TD2 [Fig. 20-33]. The four-level signal carried two bits of information in

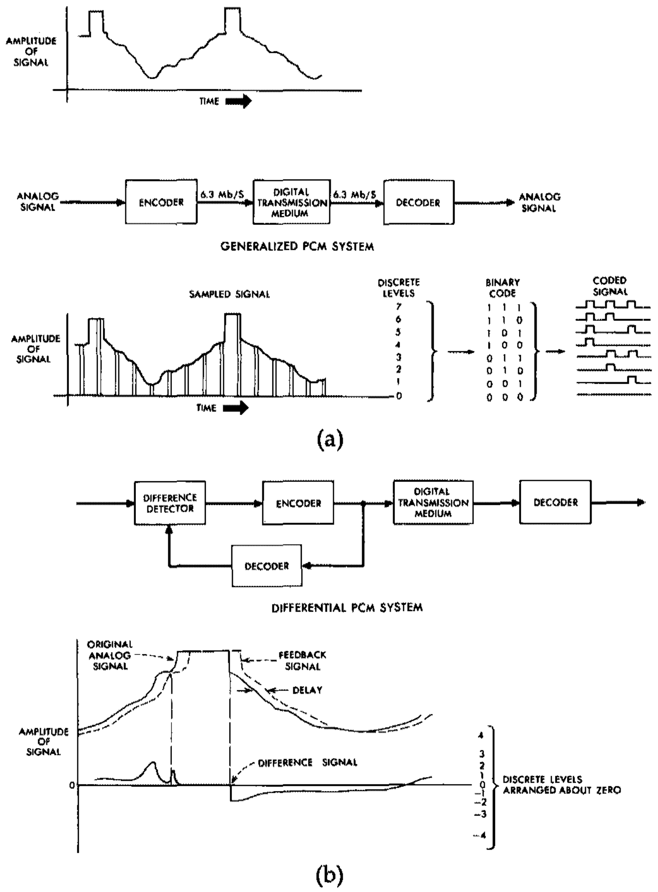


Fig. 20-32. PICTUREPHONE visual telephone service signal coding by three-digit differential PCM. (a) Uniform-step coding produced high quantization noise for low-level signals. (b) Differential PCM improved low-level, signal-to-noise by concentrating steps near zero.

every pulse, that is, the symbol rate was 10.1 megabauds. This signal was used to frequency modulate the microwave carrier by applying it to the existing FM deviator in the TD2 terminal. With careful limiting of the base-band spectrum, the four-level signal was suitable for transmission on the existing TD2 network without modifications and without interference into adjacent channels. It also did not affect the performance of the existing radio protection-switching system that was shared between the digital and analog FM-radio channels. A manually adjusted, tapped delay-line equalizer helped provide the required performance.[47,48]

The terminals, which were originally called M2R to indicate second-level multiplex for radio, were field tested on a New York-to-Pittsburgh route that

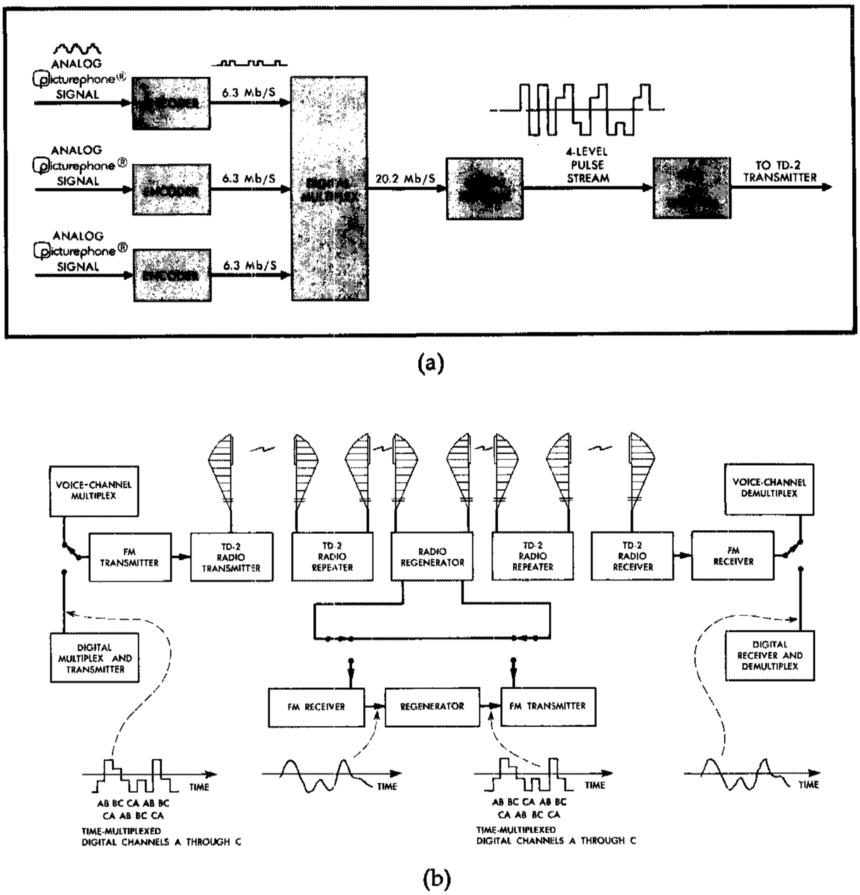


Fig. 20-33. Transmission of digital PICTUREPHONE visual telephone service signals over TD2 radio. (a) Three 6.3-megabit signals were multiplexed to a 20.2-megabit signal and then converted to a four-level 10-megabaud signal for transmission over a TD2 channel. (b) The same radio channel could carry PICTUREPHONE service or analog voice signals.

contained a midpoint regenerator at Alma, New York. The tests ran from July 1968 until August 1969 and showed that the performance was error free in 99.5 percent of the ten-second samples, and less than 0.05 percent exceeded an error rate of 10^{-7}. A limited number of M2R terminals were manufactured for an early service trial between Pittsburgh and Chicago.

Final development of an improved terminal was to include an adaptive equalizer that automatically compensated for the effects of selective radio fades, and automatic protection switching around the terminals was planned. Regular commercial PICTUREPHONE visual telephone service did not materialize as anticipated, however, and the design was not completed.

4.2 The DR18 Digital Radio System

The DR18 digital radio system was developed from 1969 to 1977, concurrent with the WT4 waveguide and T4M coaxial-cable systems, to carry multiple DS4-level digital signals (4032 voice circuits each) in 4000-mile-rated long-haul service. The repeater design was based on a pole-line radio concept studied and tested in the 1960s by the radio research area of Bell Laboratories [Fig. 20-34].[49]

Each radio route was intended to carry from one to seven DS4 signals in entrance link or feeder service for the millimeter waveguide system, with the advantages of lower start-up costs and shorter installation intervals than buried media in metropolitan or suburban areas. Later, when the waveguide system development was discontinued, radio system development continued for metropolitan high-density trunk service, with performance requirements and outage objectives modified for the shorter-link application.

The DR18 system was developed in the 18-GHz common-carrier band because all the lower-frequency bands were divided into subbands that were too narrow for the high-rate digital signal. The bandwidth used in DR18 was 220 MHz, capable of carrying a 274-megabit (DS4) signal on each polarization,

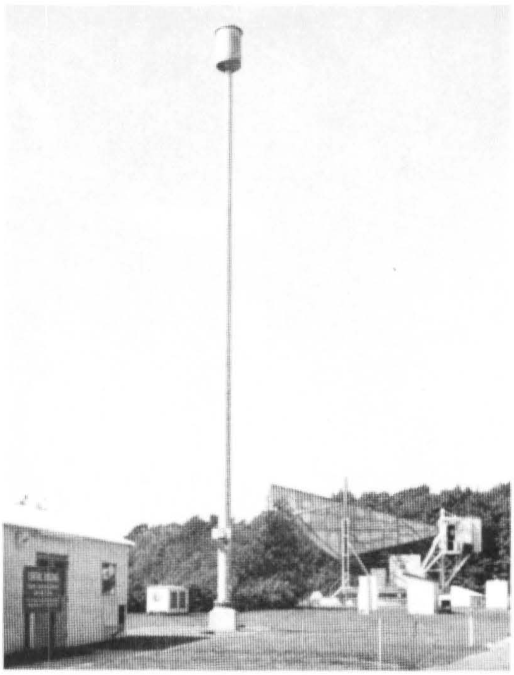
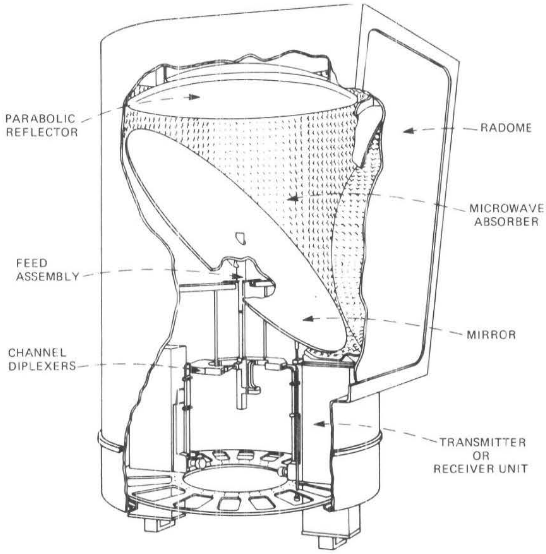
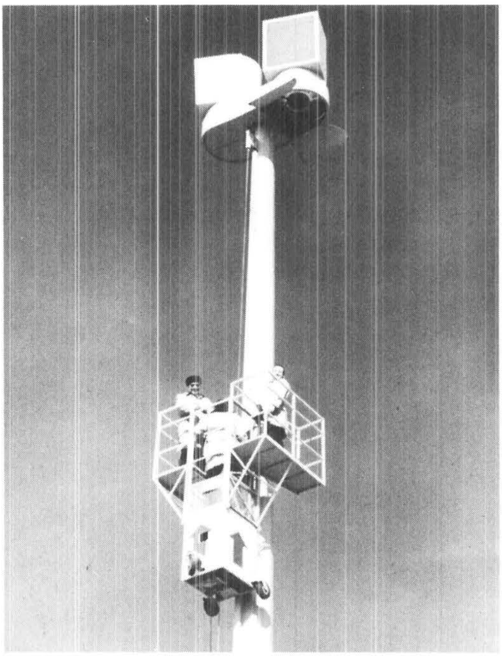


Fig. 20-34. Experimental pole-mounted 11-GHz radio repeater at Crawford Hill, New Jersey, 1969.



(a)



(b)

Fig. 20-35. 18-GHz DR18 digital radio, 1975. (a) Pole-top canister with radio repeater and antenna assembly. (b) DR18 mast with repeater-antenna canisters in place. The service lift was used for maintenance access.

using a four-level phase-shift-keyed format. Transmission in the 18-GHz band is characterized by brief but substantial rain-induced attenuation; and for transmission reliability, repeater spacings of only one to three miles were allowed. This meant that the investment per station had to be low and that large channel capacity was a necessity.

Unlike the modified TD2 channel, DR18 was an entirely digital system with regeneration at every repeater. Its repeaters could not amplify an analog signal. (Just what constituted a digital system was a much-discussed issue. To some purists, only binary transmission with regeneration at every repeater was allowed. Intervening linear amplifiers were analog deviants. To others, the loss and dispersion, and consequent gain and equalization required, even in the pure binary systems, looked very much like analog processes. Additional linear gain between regenerators did not seem to matter. To pragmatists, if digital signals were sent and recovered with adequately low error rates, the system was sufficiently digital.)

The 18-GHz radio equipment made early use of thin-film, microwave integrated circuits and featured a fixed-level transmitter-oscillator at 18 GHz, followed by a four-phase, path-length-switch modulator. Many choices in the design were based on the necessity of operating in a nearly outdoor environment. Power for the repeaters was obtained from commercial AC, with a built-in battery back-up system. The antennas and radio equipment were enclosed in mast-mounted canisters. Access to the mast-mounted repeaters was obtained via a winch-driven service lift [Fig. 20-35].[50,51]

Several high-density digital corridor routes were identified, and, in 1977, a five-hop system was placed in service by the New York Telephone Company. However, as with the other early high-speed digital systems, further installations did not materialize, and the system was subsequently discontinued. Lightwave systems, with few repeaters or none, later became better suited to serve the metropolitan high-density digital trunking needs.

4.3 The 3A-RDS Digital Radio System

With the reduced traffic growth and the indefinite postponement of PICTUREPHONE visual telephone service, DR18 and the other high-speed digital systems failed to catch on, but interest in digital radio remained very high. In the early 1970s, there was active development in several organizations, both in the United States and abroad. By 1974, two one-hop digital radio systems procured from outside sources, one working at 34 megabits and the other at 40 megabits, were in experimental service in the Bell System. Both systems operated in the 11-GHz common-carrier band and were used to interconnect areas of T1 carrier. The limitations of these systems and the continued interest in digital radio led to a decision in 1974 that Bell Laboratories would develop a new 11-GHz digital radio system, the 3A-RDS.

The new system was to be capable of transmitting a 45-megabit DS3-rate signal on both vertical and horizontal polarizations in the 40-MHz radio channels. Rapid development and early service were to be achieved by using as much as possible of the all-solid-state TN1 radio system that had been recently developed for 11-GHz analog FM transmission. Only one microwave filter would have to be changed in TN1. The 3.5-w output of its solid-state transmitter and the low noise figure of its receiver would result in an unusually high maximum repeater gain of 110 dB. The high gain would allow for deeper fades before a digital error rate of 10^{-3}, the threshold for protection switching, was exceeded. The high fade margin was important because fading from rain, although less than at 18 GHz, is much more serious at 11 GHz than at lower frequencies, and the margin could be used to obtain longer hops.[52]

4.3.1 *Digital Scrambling*

The importance of scrambling the modulation bit stream for digital radio was recognized as early as 1971. At that time, Bell Laboratories, through AT&T, proposed to the FCC the use of digital scrambling to provide a smooth emitted spectrum. Scrambling would prevent the transmission of strong spectrum lines that could cause severe co-channel interference into analog FM channels. If this was not done, the coexistence of digital and analog systems in the same band would have been very difficult. 3A-RDS became the first commercial high-capacity digital radio system to use scrambling. The input bit stream was read into a 17-stage shift register and read out in a sequence that eliminated regular patterns. Descramblers at the receiver performed the inverse function and restored the original signal. The almost-random bit stream achieved with scramblers had important advantages for circuit design as well. It allowed AC coupling of the bit stream without resorting to bipolar coding or quantized feedback to keep baseline wander under control. The design of timing and carrier-recovery circuits was also simplified.

In early 1975, the FCC opened an inquiry that later led to regulations governing digital microwave-radio systems. The rules placed restrictions on the radiated spectrum that could, in general, be met by digital scrambling. They also required that systems be capable of transmitting at least 1152 channels (48 twenty-four-channel DS1s) in the 40-MHz radio bands at 11 GHz. The system plan for 3A-RDS met the capacity requirement by transmitting two 45-megabit DS3-rate signals (28 DS1s on each) simultaneously on cross-polarized radio carriers at the same frequency. The scrambling patterns on the two carriers were different, so that, if a transmitter failed, the corresponding receiver would not lock on the weak cross-component of the other carrier and receive an intelligible signal.

4.3.2 *Modulation*

Reusing TN1 dictated the use of a digital modulator operating at the IF frequency of 70 MHz, rather than a radio frequency modulator of the type used in DR18. This, in turn, limited the choice of modulation formats. Modulation had to be compatible with the amplitude-limiting characteristics of the 11-GHz IMPATT transmitter output stage in TN1 (the amplifier used an IMPATT oscillator operating at fixed level). FM, as used in TD2 with M2R, would have been suitable, but was ruled out in favor of the more efficient and less interference-prone method of phase-shift keying (PSK) with coherent detection. Four-phase modulation (4-PSK) was selected. The amplitude modulation, inherent in any PSK scheme, was greatly reduced by an ingenious new arrangement.

At the transmitting terminal, the incoming DS3 signal was first retimed, regenerated, and scrambled. At the output of the scrambler, the serial bit stream was converted to two physically separate parallel rails, with the pulses on the two rails offset by a one-half baud* interval. The offset drastically reduced the AM of this offset 4-PSK signal. The signal could then be sent through the fixed-level IMPATT power amplifier with little degradation. The 4-PSK modulation scheme was also well suited for the transmission of the DS3 rate in the 40-MHz, 11-GHz band channels within the requirements of the new FCC emission limitations.

The concept of using digital modulators at IF working together with radio transmitters and receivers that could be used either for digital or analog signals was new at the time but later became a common method of digital radio design. Later radio transmitters, however, used more linear amplifiers and could handle more complex modulation formats.

4.3.3 Cross-Polarized Transmission

In radio transmission, the electromagnetic wave is usually launched with the electric vector vertical (vertical polarization) or horizontal (horizontal polarization). If two waves are transmitted orthogonally (e.g., with electric vectors vertical and horizontal), antennas and radio-frequency coupling components can be designed to accept one polarization and discriminate against the other. In microwave systems, antennas commonly receive both polarizations about equally well, and coupling arrangements are devised to separate them. Alternating polarization sense in radio channels at adjacent frequencies was used starting about 1960 for more efficient use of the microwave bands. (See Chapter 11, Section 1.5.) Cross-polarization isolation cannot be made perfect, however, and, in the FM systems, both the cross polarization and frequency offset were necessary to keep interference at tolerable levels.

Since digital signals are more tolerant of interference than analog signals, the 3A-RDS system was planned to operate with two cross-polarized transmitters on the same frequency. Based on the experimental data available, it was expected that such operation would be viable. The FCC, for instance, based its minimum capacity requirement of 1152 voice circuits for the 40-MHz radio channels at 11 GHz on the assumed feasibility of cross-polarized operation. Since 3A-RDS could transmit only the DS3 rate, or 672 voice circuits, on one polarization, cross-polarized operation was a necessity.

First service on a 3A-RDS system was in September 1976 on a two-hop system between Albuquerque and Tijeras, New Mexico. The early installations

* A baud is a unit of signaling speed and refers to the number of times the state or condition of a signal changes per second. It is the reciprocal of the length in seconds of the shortest element in the signaling code. A bit by contrast is the smallest unit of information in a digital signal. In a binary, two-level signal the baud rate and bit rate are identical; but, in a four-level code, each symbol can convey two bits of information and, hence, twice as many bits per second as bauds.

of 3A-RDS, which operated on a single polarization and over relatively short hops, worked very well. However, on longer hops when dual-polarized systems were installed in areas subject to severe multipath fading, the performance degraded more than expected. It was suspected that the dispersive nature of multipath transmission was partly responsible, but it was also observed that there was little correlation between fades on the opposite polarizations and that crosstalk between the cross-polarized channels was higher than expected. When the signal on one polarization was in a deep fade, it was subject to strong interference from the other, which often remained close to normal level. The effect could be mitigated in 3A-RDS by the use of frequency-diversity switching, but studies of the phenomena led to a better understanding of, and a preference for, multilevel transmission on a single polarization.[53,54,55] The economics of dual polarized systems also looked bleak when multilevel modulation schemes could be used to transmit as many bits per second over a single transmitter as dual-polarized systems transmitted over two. No new same-frequency, cross-polarized digital radio systems have been introduced in the United States since the days of 3A-RDS, although advances in technology in the 1980s have revived interest in the possibility of their use.

4.4 Further Developments

The development of digital radio by no means ended with 3A-RDS. On the contrary, in subsequent years, its progress accelerated, but the full story of those developments is beyond the scope of this history. Even while the first 3A-RDS systems were being installed, the Collins Radio Group, by then a part of Rockwell International Corp., was demonstrating 90-megabit transmission in the 40-MHz channels at 11 GHz by eight-level PSK (modulation to eight different phase states of a constant amplitude signal) [Fig. 20-36].[56] By the early 1980s, Bell Laboratories had designed and put in production a system, DR11-40 transmitting 140 megabits (three DS3s) in the 40-MHz bands at 11 GHz and another, DR6-30, transmitting 90 megabits in the 30-MHz bands at 6 GHz.[57] Both systems used quadrature-amplitude modulation (QAM). In QAM, the different signal states are realized by what amounts to a combination of phase and amplitude changes in the signal. This, in turn, was made possible by the use of multiwatt gallium arsenide field-effect transistor amplifiers (GaAsFETs—"gasfets"), which became available in the late 1970s. These amplifiers provided linear gain to amplitude-varying signals, in contrast to the earlier IMPATT amplifiers that were well suited only to constant-level signals

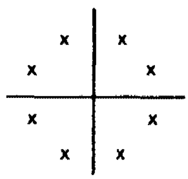


Fig. 20-36. Phase-plane representations for eight-state constant-amplitude phase shift keying.

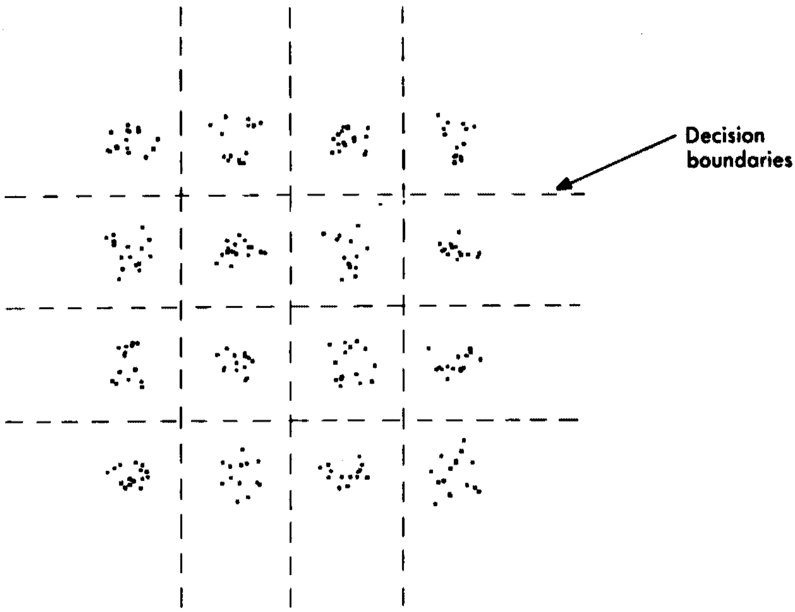


Fig. 20-37. Phase-plane representation of a 16-state QAM signal perturbed by noise as would occur in a fade.

such as FM or PSK.[58] Both DR6-30 and DR11-40 systems used 16-state QAM signal formats [Fig. 20-37]. By 1985, further developments in QAM to 64 states made possible the transmission of 135 megabits in the 30-MHz channel of the 6-GHz radio band.

V. EARLY OPTICAL-SYSTEM WORK

5.1 Setting the Stage

Only a few years after his invention of the telephone, Alexander Graham Bell experimented with communication using light. In 1880, he was granted a patent for what he named a *photophone* that used reflected sunlight over a line-of-sight path through the air. Even earlier, in 1870, the famous British physicist, John Tyndall, had demonstrated that part of the light illuminating the inside of a container of water was guided along the curved stream of the liquid flowing through a hole in the side of the vessel. That was the first recorded demonstration of light transmission along a dielectric guide by total internal reflection. However, practical light-wave communications had to wait for two essential ingredients; an intense, continuous, concentrated light source and a low-attenuation, stable transmission medium. The invention of the laser (at first called the optical maser) in 1958 at Bell Laboratories by A. L. Schawlow and C. H. Townes and the first realization of laser operation by T. H. Maiman of the Hughes Aircraft Company in 1960 were the key steps toward meeting the first requirement.[59,60] In 1966, K. C. Kao and G. A. Hockham at Standard Telephones and Cables, Ltd. in England performed experiments and showed by an analysis that there was no fundamental physical barrier to low-loss glass fiber waveguides; this was the breakthrough that pointed the way to a practical medium.[61] The first realization of fibers with loss below 20 dB per kilometer was by R. D. Maurer and others at the Corning Glass Works research and development laboratory in the early 1970s.[62,63]

The laser invention was of great significance. Earlier light sources were all incoherent, that is, they produced in effect a band of electromagnetic noise at optical frequencies, and they did so nondirectionally, over relatively large solid angles. Lasers, by contrast, were coherent and concentrated; they produced electromagnetic waves in a very narrow band approaching a single frequency; they were analogous to the vacuum-tube and transistor oscillators at low frequencies and to the klystrons, IMPATT diodes, and other oscillators at microwave and millimeter wavelengths. Like them, lasers could be efficiently coupled into a transmission line. Lasers opened the prospect for communications at optical wavelengths, with frequencies and bandwidths three to four orders of magnitude greater than any previously available [Table 20-3 and Fig. 20-38]. The prospect for systems with many millions of channels, hundreds of times greater than earlier systems, was often mentioned in discussions of lasers, but such comparisons are somewhat misleading. It was not expected that in the early stages the optical bands would be used with the same efficiency as the lower-frequency bands. Nevertheless, the prospect was for enormously expanded capacity.

5.2 Lasers and Light-Emitting Diodes

The decade of the 1960s was one of extensive research into laser structures. The helium-neon gas laser, achieved at Bell Laboratories in 1960, became the workhorse of the scientific community in studying optical phenomena. However, a gas laser is too large to be a practical source for a light-wave communications system. Much of the effort was directed to the search for a semiconductor laser. Although the first such lasers were reported in 1962, these were experimental laboratory devices that could operate only when pulsed at a very low duty factor or only if cooled to very low temperatures. In 1970, Bell Laboratories first achieved continuous operation of a semiconductor laser at room temperature.[64,65] Although much development work still lay ahead, this was an important breakthrough in indicating that tiny semiconductor lasers, literally

TABLE 20-3: Wavelengths and Frequencies of Principal Communication Bands

Band	Approximate Wavelength	Approximate Frequency	
AM Radio	300 m	1 MHz	10^6 Hz
High Frequency Radio	30 m	10 MHz	10^7 Hz
FM Radio, Television	3 m	100 MHz	10^8 Hz
Microwave Radio	8–3 cm	4–11 GHz	$4\text{–}11 \times 10^9$ Hz
Millimeter Waveguide	1 cm–3 mm	30–100 GHz	$3\text{–}10 \times 10^{10}$ Hz
Optical	1 μm	300,000 GHz	3×10^{14} Hz

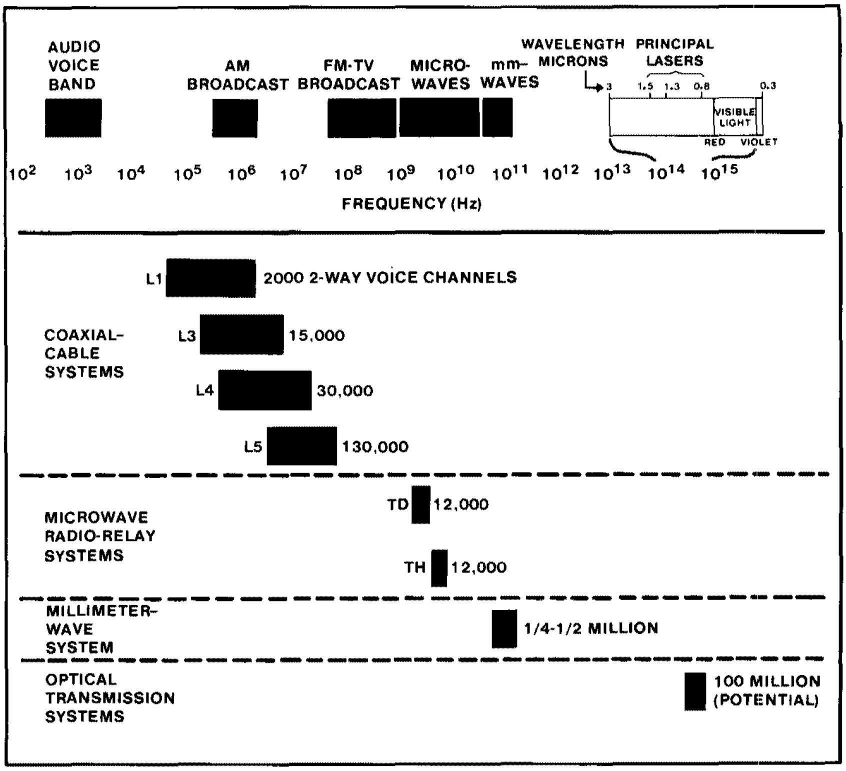


Fig. 20-38. Spectrum occupancy and system capacities.

no larger than a grain of sand, might serve as practical light-wave communications sources.

At about the same time, the research on semiconductor light sources produced specially designed light-emitting diodes (LEDs).[66] (See *Communication Sciences (1925–1980)*, Chapter 7, Section 2.4.) The light from LEDs was not coherent, but it was in the form of an intense beam from a very small source area over about a one-percent frequency band. LEDs were attractive because of their reliability and life, which was shown to be at least several thousand hours, and the convenience of operation. The diodes could be modulated simply by varying the current through them and at a rate sufficient for an information band of 10 or 20 megahertz or more. LEDs could be coupled into larger multimode fibers of a few mils in diameter and were, in fact, the light source for most of the early light-wave systems of moderate length and information capacity.

5.3 Optical Fibers

The second ingredient needed for optical communications was a good transmission medium. The earth's atmosphere is generally quite transparent to optical signals, but there are times when fog, mist, rain, or snow greatly reduce visibility. Bell Laboratories had done a great deal of pioneering experimental and theoretical work on the propagation of electromagnetic waves of various frequencies through the atmosphere. (See Chapters 2, 7 and 11.) This work, which was essential to the design of high-frequency radio, microwave, and satellite systems, was now extended to optical frequencies. The properties of atmospheric paths were quantified, with the (not surprising) result that, over paths of any considerable length, fog, mist, rain, and snow would create intolerable losses much too often for practical communications.

With the limitations of free-space transmission, research concentrated on guided-wave transmission of optical signals. Transmission in conducting hollow guides by the circular-electric mode would have required microscopic dimensions and tolerances that were completely unattainable. Early work was on a guidance system provided by a sequence of lenses in the sheltered environment of a surrounding pipe. Both glass and gaseous lenses were investigated on a theoretical and experimental basis [Fig. 20-39].[67] But by 1970, guidance by glass fibers became the overwhelmingly preferred prospect.

The fact that thin filaments of glass could guide beams of light was a well-known phenomenon, and such waveguides had been used for many years, for example, in medical instrumentation, to bend light around corners. In the medical applications, however, an entire bundle of fibers was illuminated rather than using each fiber individually. As so often in the long history of waveguides, the knowledge that dielectric wires could act as guides for electromagnetic (EM) waves was recognized in the classical period by the EM pioneers. D. Hondros and P. Debye, in Germany, published an analysis as early as 1910.[68] Dielectric guides were used as laboratory patch cords during the millimeter-wave research. However, before 1970, even the best-quality optical glass would attenuate light signals below detectable levels in a distance of 100 meters or so, rather than

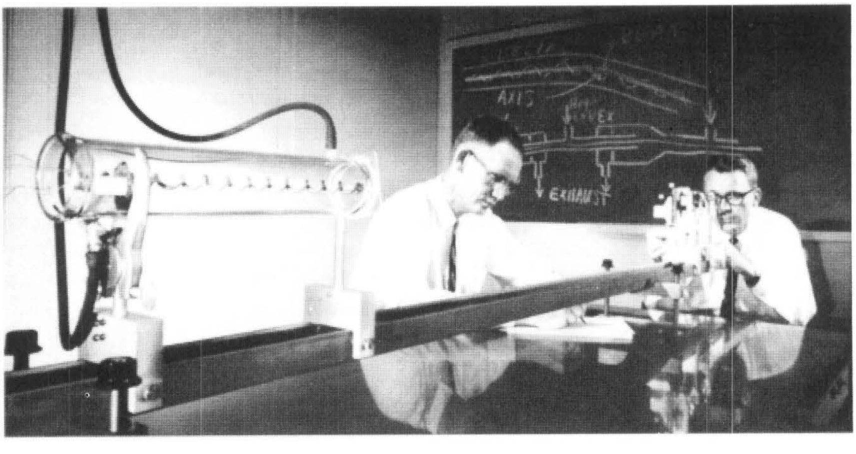


Fig. 20-39. Testing an experimental helical gas convection lens. The lens used temperature-produced changes in the refractive index of a gas to guide light.

the several kilometers needed for practical light-wave communications. In 1966, Kao and Hockham, working in England, made the momentous observation that in pure glass there was no basic physical limitation that restricted its transparency to such limits. They noted that the lowest reported absorption coefficient for glass was equivalent to a bulk loss of about 200 dB per kilometer, but they suggested that if high-silica glass were used and impurities could be controlled, losses as low as 20 dB per kilometer might be achieved.[69] Research in Britain, Japan, and the United States (primarily at Corning Glass Works and Bell Laboratories) was directed towards achieving the low-loss fibers. It was Corning that first achieved substantial progress, announcing losses under 20 dB per kilometer in 1970.[70] Rapid progress followed, and by late 1972 losses as low as 4 dB per kilometer were reported by Corning at wavelengths in the vicinity of 1 μm [Fig. 20-40].

Optical fibers consist of a highly transparent glass core surrounded by a glass cladding of slightly lower dielectric constant. As in all electromagnetic waveguides, propagation, if it is possible at all, is in modes, each of which is a solution to J. C. Maxwell's equations for the wavelength, geometry, and materials involved. In simpler terms, the guidance may be thought of as due to total internal reflection because of the discontinuity in the index of refraction at the boundary of the core and cladding. Fibers can be made to support a single mode, in which case the core can be only a few wavelengths (micrometers)

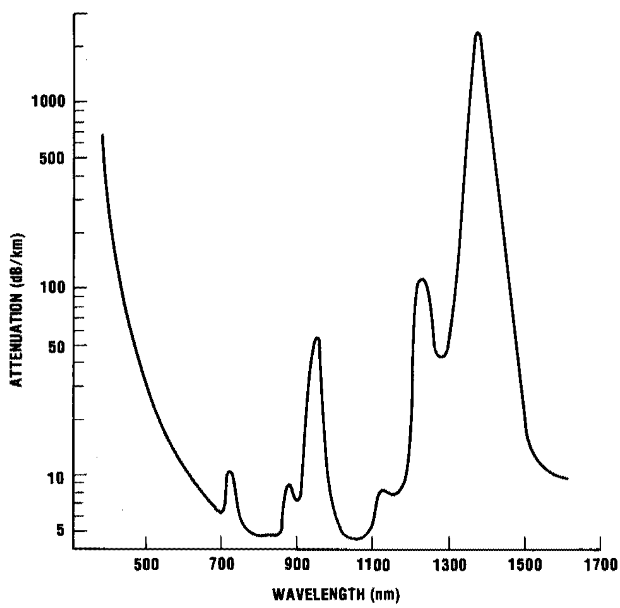


Fig. 20-40. Attenuation of a low-loss optical waveguide (from research at Corning Glass Works, 1972). The peaks of absorption are due to water impurities.

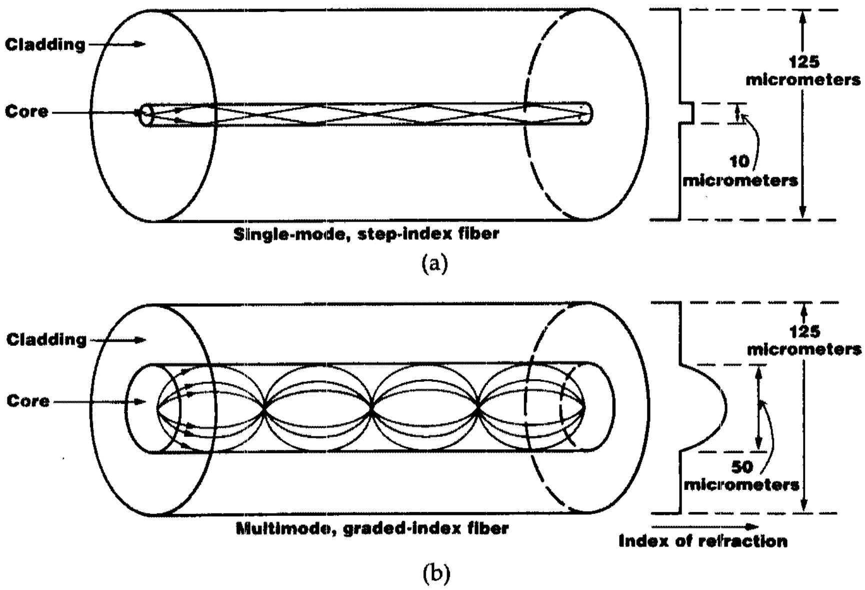


Fig. 20-41. Optical fiber lightguides. (a) A single mode fiber with a step change in the dielectric index between the core and cladding. (b) A multimode fiber with graded index.

in diameter or the fiber guide can be multimode, in which case the core will be much larger in diameter [Fig. 20-41]. Single-mode fibers have very low dispersion and can support a high digital signal rate (short pulses) over a long distance before the pulses become distorted beyond the ability to recover them. They are especially useful with laser sources that can couple a large portion of their highly concentrated monochromatic beams into the slender core. Multimode fibers are easier to produce and were often used with LED sources so that a larger portion of the noisy and less concentrated light could be coupled into and carried by the fiber. For either type of fiber, guidance can be by reflection at a single step change in the dielectric index at the boundary between the core and cladding or by a gradual bending of divergent rays back toward the fiber axis in a graded-index fiber. The graded-index design is especially useful to reduce dispersion in multimode fibers.

There were several methods of making fibers, but by the late 1970s the most common method was by first constructing a preform of the appropriate glasses for the core and cladding. The preform was typically a half inch or so in diameter and two or three feet long. One end was heated until it was plastic and then drawn into a fiber; the core and cladding maintained their proportional relationship with a high degree of perfection, governed only by the forces that control the plastic flow of the glass [Fig. 20-42].[71] A single preform could be drawn into several kilometers of fiber. (See another volume in this series, *Physical Sciences (1925–1980)*, Chapter 13, Section 2.1.2.)

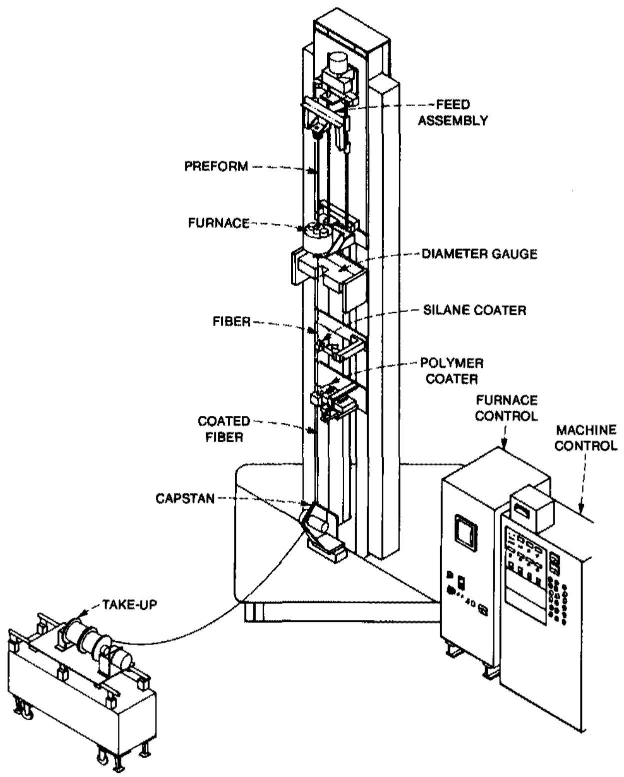
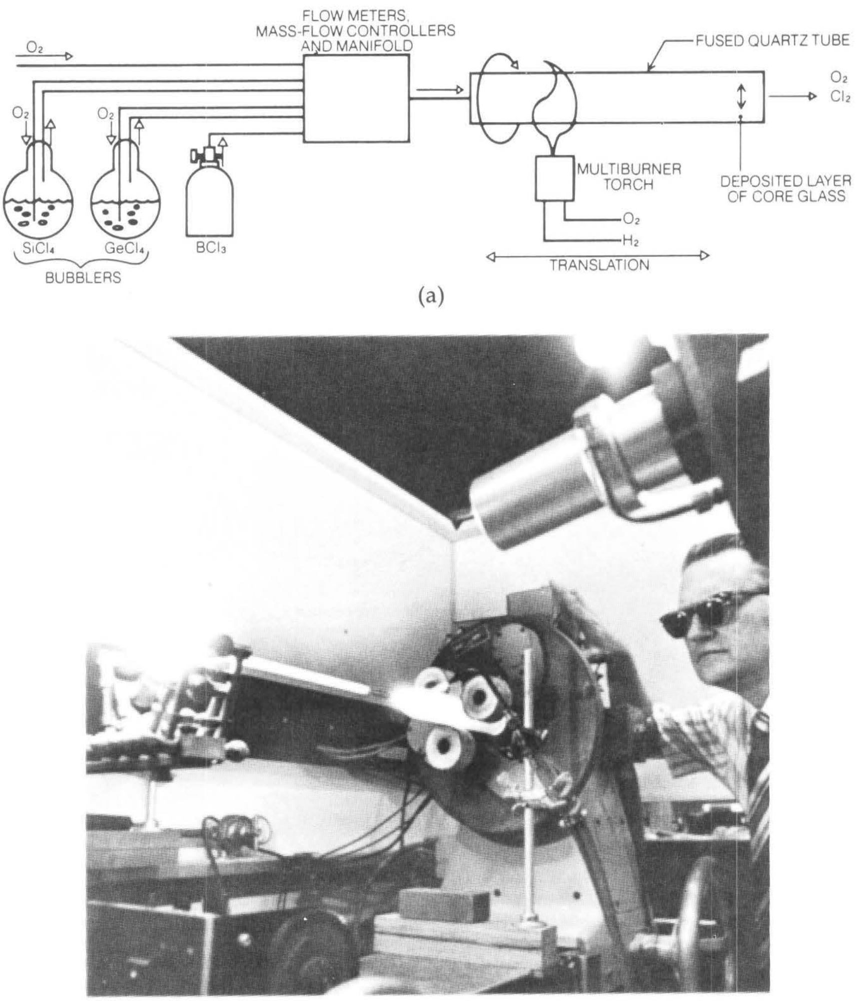


Fig. 20-42. Fiber-drawing machine, 1978.

5.4 Further Developments (1970–1974)

Following 1970, with a good light source in the form of the semiconductor laser and the realization of low loss fibers, attention began to be given to technology development. Groups in Bell Laboratories and elsewhere addressed the problems of converting the techniques used in the research laboratories to processes suitable for large-scale production of practical commercial components. At the same time, system planners began to study where this technology might be most beneficially applied in the communications network. Research, of course, continued and provided critically important input to the technology development and planning and, in turn, was stimulated by them.

The composition and purity of the glasses in the preform established the limit of what could be realized in the fibers drawn from them. Early fibers had been made from preforms built up from the core material outward, or they were extruded from concentric orifices. In 1974, a significant advance was made in Bell Laboratories in the modified chemical-vapor deposition (MCVD) process for producing the tubular preforms. In this process, layers of material



(b)

Fig. 20-43. Making an optical fiber. (a) Construction of a fiber preform by the modified chemical vapor deposition (MCVD) process. (b) A silica tube installed in a glass lathe for chemical vapor deposition.

with the required optical properties were deposited as a soot inside a heated fused-quartz tube. The tube and the deposited layers were then further heated and collapsed into the preform by final passes through a flame at a higher temperature [Fig. 20-43]. By the mid-1970s, fibers drawn from preforms produced by MCVD were exhibiting losses as low as 1 dB per kilometer.[72] The MCVD process was widely adopted by fiber producers around the world. It continued to be refined and produced fibers of astonishing transparency. By the mid-1980s, losses as low as 0.16 dB per kilometer had been achieved—

three orders of magnitude better than the best reported values in the early 1960s. One kilometer of this glass has little more loss than a few panes of window glass.

Solid-state lasers were also improved. The first continuous-wave lasers in 1970 had a lifetime of only a few hours. By 1973, this had been increased to 1000 hours. By 1976, the indicated life was 100,000 hours ("indicated," because developers could no longer observe a sufficient number of failures in the limited number of devices on test to determine the average life accurately). A year later, lasers with a probable average life of 10^6 hours were achieved. (10^6 hours is over 100 years. What was actually measured, and was more important, was failure rate. The probable life of 10^6 hours meant less than one percent would fail per year.) During the same period, highly efficient photodetectors in the form of silicon avalanche photodiodes were developed.[73]

5.5 The First Systems Work—The Atlanta Experiment

By 1974, personnel were transferred from research to the transmission systems area of Bell Laboratories, where they joined forces with engineers skilled in digital transmission systems, to investigate the characteristics and features required in commercial light-wave systems. The attributes of light-wave communications—large communications capacity, small-diameter lightweight cables, long distances between repeaters, and immunity to interference—were potentially beneficial in all portions of the telephone plant. The most sought-after prize was a digital system that would be economical in the long-haul plant. Such a system, with the widespread local-area T1 carrier systems, and the toll digital switch, which was well along in development, would provide the long-sought nationwide digital connectivity. But the early stage of the technology and the same factors that had limited the prospects of millimeter waveguide and high-capacity digital radio worked against the early realization of long light-wave systems. Systems-engineering studies indicated that light-wave systems would likely have the earliest economic prove-in in metropolitan areas, where large numbers of interoffice circuits were required and where it was desirable to eliminate repeaters in manholes between central offices. From the set of standard transmission rates previously established, the studies focused on the DS3 rate of 45 megabits per second, corresponding to 672 digital voice channels on each fiber. It was expected that about 144 fibers could be packaged in a sheath about one-half inch in diameter. This was especially attractive for metropolitan applications where space in underground ducts was at a premium [Fig. 20-44].

As in earlier systems that introduced a wholly different technology, a host of new problems had to be solved. A light-wave system would consist of many components, both optical and nonoptical. In addition to light sources, fibers, and photodetectors, means of joining fibers, both demountable connectors and permanent splices, were required. Practical cable structures had to be developed for installation in the outside plant. Electronic circuits were required to control the light sources and process the output of the photodetectors. It was necessary to provide all the interfaces to the standard electronic signals in the telephone network in which the fiber systems would be embedded.

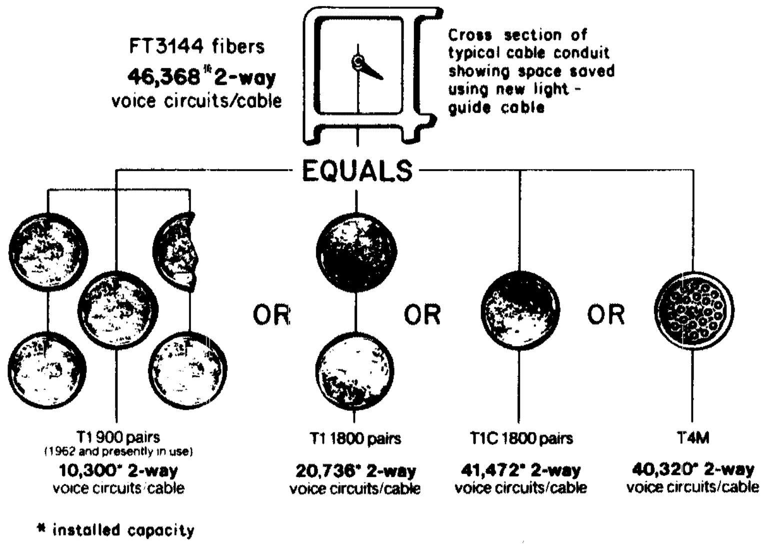


Fig. 20-44. Duct occupancy for 45-megabit FT3 lightwave system on a 144-fiber cable compared to digital systems on pairs or coaxials. A duct could hold only one of the paired cable or coaxial sheaths. [Chynoweth and Miller, *Optical Fiber Telecommunications* (1979): 9.]

In 1974, planning began for a complete light-wave system experiment at the 45-megabit rate. The purpose of this experiment was to evaluate light-wave technology in an environment approximating field conditions. In addition, it would provide a focus for the continuing exploratory development effort. The experiment was implemented in late 1975 at the joint Western Electric-Bell Laboratories facility in Atlanta, Georgia, where the fiber and cable were made and underground ducts typical of those in metropolitan areas were available [Fig. 20-45]. The system included a 144-fiber cable that was 650 meters long. A ribbon-structured cable was chosen to facilitate splicing and to provide experience in the making of large-fiber-count cables. The sheath was designed with sufficient strength to protect the fibers while the assembly was pulled into the ducts by standard techniques [Fig. 20-46]. Many factors besides glass purity controlled the loss of a complete cable. Splices and microbending as the fibers were assembled into a cable structure also contributed to loss. High, uniform tensile strength is extremely important and was difficult to achieve. The objective was to achieve at least 100 good fibers in the cable with an average loss of no more than 8 dB per kilometer. The results achieved were 138 good fibers with an average loss of 6 dB per kilometer.

The digital regenerative repeater was a key element in the trial. The receiver in the Atlanta regenerator utilized a silicon avalanche photodetector and the transmitter, a gallium aluminum arsenide laser, radiating 0.5 mw into the fiber

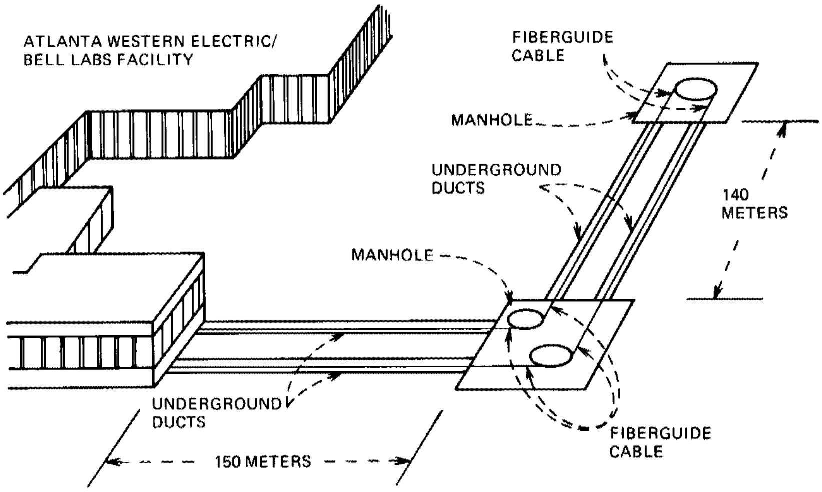


Fig. 20-45. Atlanta optical fiber trial installation.

at a wavelength of 0.82 μm. The regenerator was constructed in three separate modules, both to ease implementation and to provide maximum experimental flexibility [Fig. 20-47]. Long optical paths were produced by connecting many different fibers in the cable end to end. Original planning had contemplated inserting repeaters at 7-km spacing, with an average span loss of 51 dB, but the low fiber loss realized permitted patching together paths as long as 10.9 km before regeneration.

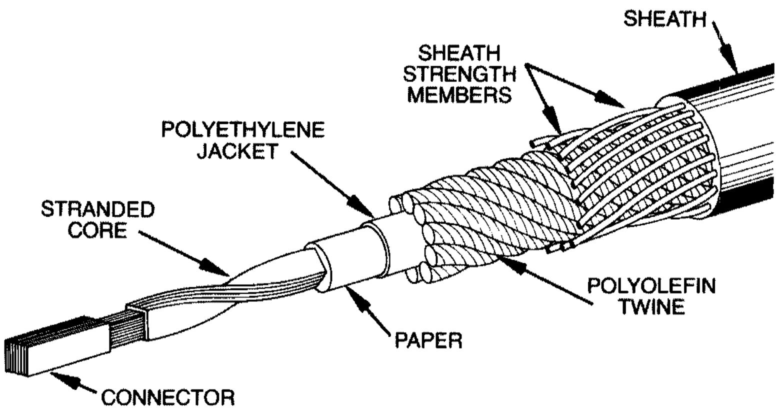


Fig. 20-46. The cable used in the Atlanta system contained 12 ribbons—each encapsulating 12 glass fibers.

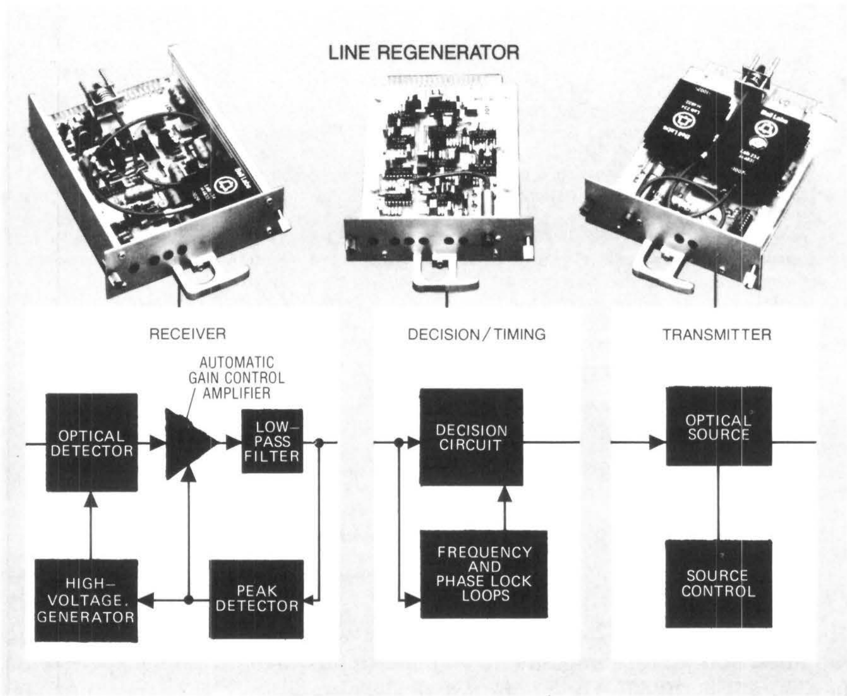


Fig. 20-47. Principal components of the Atlanta optical system trial line regenerator, 1975.

The objectives of the Atlanta system experiment were fully met. Although much specific development remained before economic manufacturable lightwave systems could be deployed, the Atlanta Experiment was a key milestone in demonstrating the practical feasibility of optical-fiber technology.[74]

5.6 Post-1975

The success in Atlanta spurred further work in both research and development and provided a much clearer picture of where progress was most needed, and progress was rapid and sustained. A few highlights up to the mid-1980s will be mentioned.

Following the Atlanta experiment, a successful trial of similar nature but with a smaller cable (24 fibers) was held in 1977 in the urban environment of downtown Chicago. The first commercial installation of the FT3 system (*F* for fiber, *T* for digital carrier, and 3 for the 45-megabit DS3 rate) was in Trumbull, Connecticut in the fall of 1979. Several other short systems followed, and fiber systems were used for both voice and video transmission at the winter Olympic Games at Lake Placid, New York in 1980 (as they were to a much greater extent at the Los Angeles summer games in 1984). By February 1983, a long-haul system at a 90-megabit rate (FT3C) was in operation along 250 miles of the Northeast Corridor from New York to Washington. Urban installations continued at an accelerating rate. By 1984, a network in downtown Miami connected 27 high-rise buildings. Transmission was at the 45-megabit DS3 rate, with light generation by LEDs at a wavelength of 1.3 μm. At the long wave length (the first systems operated at 0.8μm), and with the superior fibers by then available, no manhole repeaters were required.

After 1980, as the technology of making fibers advanced, interest shifted to single-mode fibers at long wavelengths. By 1984, AT&T Communications (successor to the Long Lines Department following the January 1, 1984 divestiture) was planning 3,300 route kilometers of optical systems, almost all at a 432-megabit rate, and in single-mode fibers at 1.3-μm wavelength. Prospects were good for at least doubling the capacity by wavelength multiplexing, that is, transmitting two high-rate signals over the same fiber at different wavelengths. Regenerators in the long-haul systems were typically at about 31-km (19-mile) spacing, but transmission at gigabit rates over more than 100 km without regeneration had been demonstrated in the laboratory on an experimental basis. (In early 1985, wavelength multiplexing of ten 2-gigabit streams over a single fiber was demonstrated. With suitable terminals, this could provide a capacity of 280,000 voice channels on a single fiber.) At the entry point to the network, optical fibers were beginning to appear in the loop plant. In the intercontinental plant, development was well along and successful deep-sea trials had been conducted on a submarine optical cable to be installed as the eighth transatlantic cable in the late 1980s.

The advance of optical-communication technology was by no means confined to Bell Laboratories; many other organizations in Europe, Japan, and America made significant contributions. But several of the key inventions and techniques were conceived in Bell Laboratories, and, by 1985, AT&T and the (by then divested) Bell System companies had installed far more fiber kilometers in systems of Bell Laboratories design than were in service for all the rival systems combined. By the mid-1980s, the fiber-optic revolution was well under way and accelerating yearly. It seemed clear that the new technology was to be a communications landmark as far-reaching in its consequences as the first telegraph and telephone transmissions over metallic conductors or the first radio transmissions.

REFERENCES

1. J. S. Mayo, "An Experimental Broadband PCM Terminal," *Bell Lab. Rec.* **42** (May 1964), pp. 152-157.
2. J. S. Mayo, "Experimental 224 Mb/s PCM Terminals," *Bell Syst. Tech. J.* **44** (November 1965), pp. 1813-1841.
3. R. W. Sears, "Electron Beam Deflection Tube For Pulse Code Modulation," *Bell Syst. Tech. J.* **27** (January 1948), pp. 44-57.
4. J. O. Edson and H. H. Henning, "Broadband Codecs for an Experimental 224 Mb/s PCM Terminal," *Bell Syst. Tech. J.* **44** (November 1965), pp. 1887-1940.
5. I. Dorros, J. M. Sipress, and F. D. Waldhauer, "An Experimental 224 Mb/s Digital Repeatered Line," *Bell Syst. Tech. J.* **45** (September 1966), pp. 993-1043.
6. W. G. Bender, "An Experiment In PCM Transmission of Multiplexed Channels," *Bell Lab. Rec.* **42** (July/August 1964), pp. 240-246.

7. H. W. Andrews, V. I. Johannes, and W. E. Woodzell, "A 44.736 Mbits/s Codec for U600 Mastergroups," *IEEE Trans. Commun.* **COM-25** (February 1977), pp. 264-271.
8. J. B. Millard and H. I. Maunsell, "The Picturephone® System: Digital Encoding of the Video Signal," *Bell Syst. Tech. J.* **50** (February 1971), pp. 459-479.
9. J. C. Candy, M. A. Franke, B. G. Haskell, and F. W. Mounts, "Transmitting Television as Clusters of Frame-to-Frame Differences," *Bell Syst. Tech. J.* **50** (July/August 1971), pp. 1889-1917.
10. D. F. Hoth, "Digital Communication," *Bell Lab. Rec.* **45** (February 1967), pp. 39-43.
11. J. P. Runyon, U.S. Patent No. 3,050,586; filed May 20, 1960; issued August 21, 1962.
12. V. I. Johannes and R. H. McCullough, "Multiplexing of Asynchronous Digital Signals Using Pulse Stuffing with Added-Bit Signaling," *IEEE Trans. Commun. Technol.* **COM-14** (October 1966), pp. 562-568.
13. F. J. Witt, "An Experimental 224 Mb/s Digital Multiplexer-Demultiplexer Using Pulse Stuffing Synchronization," *Bell Syst. Tech. J.* **44** (November 1965), pp. 1843-1885.
14. M. C. Biskeborne, W. G. Nutt, and B. Wargotz, "CLOAX . . . Next Generation Coaxial Cable," *Bell Lab. Rec.* **47** (July 1969), pp. 195-198.
15. J. M. Sipress, "T4M: New Superhighway for Metropolitan Communications," *Bell Lab. Rec.* **53** (October 1975), pp. 353-359.
16. P. E. Rubin, "The T4 Digital Transmission System—Overview," *IEEE International Conference on Communications* **3**, Seattle, WA (June 16-18, 1975), pp. 48-1–48-4.
17. F. D. Waldhauer, "Quantized Feedback in an Experimental 280 Mb/s Digital Repeater for Coaxial Transmission," *IEEE Trans. Commun.* **COM-22** (January 1974), pp. 1-5.
18. S. E. Miller, "Waveguide as a Communication Medium," *Bell Syst. Tech. J.* **33** (November 1954), pp. 1209-1265.
19. R. Kompfner, "Backward Wave Oscillator," *Bell Lab. Rec.* **31** (August 1953), pp. 281-285.
20. J. R. Pierce, U.S. Patent No. 2,848,695; filed March 15, 1954; issued August 19, 1958.
21. H. E. Rowe and W. D. Warters, "Transmission in Multimode Waveguide with Random Imperfections," *Bell Syst. Tech. J.* **41** (May 1962), pp. 1031-1170.
22. M. Jouguet, "Les Effets de la Courbure sur la Propagation des Ondes Électromagnétiques dans les Guides a Section Circulaire," *Cables & Transmission* **1** (July 1947), pp. 133-153.
23. W. J. Albersheim, "Propagation of TE_{01} Waves in Curved Wave Guides," *Bell Syst. Tech. J.* **28** (January 1949), pp. 1-32.
24. S. E. Miller, "Notes on Methods of Transmitting the Circular Electric Wave Around Bends," *Proc. IRE* **40** (September 1952), pp. 1104-1113.
25. H. G. Unger, "Circular Electric Wave Transmission in a Dielectric-Coated Waveguide," *Bell Syst. Tech. J.* **36** (September 1957), pp. 1253-1278.
26. W. T. Read, Jr., "A Proposed High-Frequency, Negative-Resistance Diode," *Bell Syst. Tech. J.* **37** (March 1958), pp. 401-446.
27. J. B. Gunn, "Microwave Oscillations of Current in III-V Semiconductors," *Solid State Commun.* **1** (September 1963), pp. 88-91.
28. J. A. Copeland, "CW Operation of LSA Oscillator Diodes—44 to 88 GHz," *Bell Syst. Tech. J.* **56** (January 1967), pp. 284-287.
29. R. L. Johnston, B. C. DeLoach, and B. G. Cohen, "A Silicon Diode Microwave Oscillator," *Bell Syst. Tech. J.* **44** (February 1965), pp. 369-372.
30. W. M. Hubbard, J. E. Goell, W. D. Warters, R. D. Standley, G. D. Mandeville, T. P. Lee, R. C. Shaw, and P. L. Clouser, "A Solid-State Regenerative Repeater for Guided Millimeter-Wave Communication System," *Bell Syst. Tech. J.* **46** (November 1967), pp. 1977-2018.
31. See special issue on WT4 Millimeter Waveguide System, *Bell Syst. Tech. J.* **56** (December 1977), pp. 1825-2196.
32. R. J. Boyd, Jr., W. E. Cohen, W. P. Doran, and R. D. Tuminaro, "Waveguide Design and Fabrication," *Bell Syst. Tech. J.* **56** (December 1977), pp. 1873-1897.

33. D. J. Thomson, "Spectrum Estimation Techniques for Characterization and Development of WT4 Waveguide-II," *Bell Syst. Tech. J.* **56** (December 1977), pp. 1983-2005.
34. P. E. Fox, S. Harris, and D. J. Thomson, "Mechanical Gauging Techniques," *Bell Syst. Tech. J.* **56** (December 1977), pp. 2007-2023.
35. M. P. Eleftherion, "Waveguide Medium Manufacturing Process Research and Development," *Bell Syst. Tech. J.* **56** (December 1977), pp. 2179-2196.
36. L. W. Hinderks and A. Maione, "Copper Conductivity at Millimeter-Wave Frequencies," *Bell Syst. Tech. J.* **59** (January 1980), pp. 43-65.
37. J. C. Anderson, R. W. Gretter, and T. J. West, "Route Engineering and Sheath Installation," *Bell Syst. Tech. J.* **56** (December 1977), pp. 1953-1982.
38. M. A. Gerdine, L. W. Hinderks, S. D. Williams, and D. T. Young, "Electrical Transmission Measurement System," *Bell Syst. Tech. J.* **56** (December 1977), pp. 2025-2034.
39. C. E. Barnes, P. Brostrup-Jensen, E. T. Harkless, R. W. Muise, and A. J. Nardi, "Regenerative Repeaters," *Bell Syst. Tech. J.* **56** (December 1977), pp. 2055-2075.
40. C. N. Dunn, O. G. Petersen, and D. C. Redline, "Semiconductor Devices for the WT4 Repeater," *Bell Syst. Tech. J.* **56** (December 1977), pp. 2119-2134.
41. S. S. Cheng, E. T. Harkless, R. W. Muise, W. E. Studdiford, and D. N. Zuckerman, "Field Evaluation Test System Performance," *Bell Syst. Tech. J.* **56** (December 1977), pp. 2147-2156.
42. J. C. Anderson, J. W. Carlin, D. J. Thomson, and T. J. West, "Field Evaluation Test—Transmission Medium Achievements," *Bell Syst. Tech. J.* **56** (December 1977), pp. 2157-2178.
43. G. C. Southworth, *Forty Years of Radio Research* (New York: Gordon and Breach, 1962).
44. S. Weinreb, R. Predmore, M. Ogai, and A. Parrish, "Waveguide System for a Very Large Antenna Array," *Microwave J.* **20** (March 1977), pp. 49-52.
45. P. J. Napier, A. R. Thompson, and R. D. Ekers, "The Very Large Array: Design and Performance of a Modern Synthesis Radio Telescope," *Proc. IEEE* **71** (November 1983), pp. 1295-1320.
46. D. W. Nast and I. Welber, "Transmission Across Town Or Across the Country," *Bell Lab. Rec.* **47** (May/June 1969), pp. 162-168.
47. C. W. Broderick, "The *Picturephone®* System: A Digital Transmission System for TD-2 Radio," *Bell Syst. Tech. J.* **50** (February 1971), pp. 481-499.
48. M. J. Thompson, "Transmitting Digital Signals On Analog Radio Links," *Bell Lab. Rec.* **49** (April 1971), pp. 102-108.
49. L. C. Tillotson, "Use of Frequencies Above 10 GHz for Common Carrier Applications," *Bell Syst. Tech. J.* **48** (July/August 1969), pp. 1563-1576.
50. A. C. Longton, "DR18—A High Speed QPSK System at 18 GHz," *IEEE International Conference on Communications*, Philadelphia, PA (June 14-16, 1976), pp. 18-14–18-17.
51. R. P. Slade, "Installation and Field Testing of DR18," *IEEE International Conference on Communications*, Philadelphia, PA (June 14-16, 1976) pp. 18-18–18-21.
52. A. J. Giger and T. L. Osborne, "3A-RDS 11 GHz Digital Radio System," *IEEE International Conference on Communications*, Philadelphia, PA (June 14-16, 1976), pp. 18-1–18-7.
53. T. O. Mottl, "Dual-Polarized Channel Outages During Multipath Fading," *Bell Syst. Tech. J.* **56** (May/June 1977), pp. 675-701.
54. P. M. Cronin, "Dual-Polarized Digital Radio Operation In A Fading Environment," *IEEE International Conference on Communications*, Seattle, WA (June 8-12, 1980), pp. 52.5.1–52.5.6.
55. A. J. Giger and W. T. Barnett, "Effects of Multipath Propagation on Digital Radio," *IEEE Trans. Commun.* **COM-29** (September 1981), pp. 1345-1352.
56. P. R. Hartmann, J. A. Crossett, and E. W. Allen, "90 MBS 11 GHz Digital Transmission System," *1977 Electro Conference Record*, New York (April 19-21, 1977), pp. 22/2-1–22/2-6.

57. C. C. Bailey, "Digital Radio Hops Over Intercity Routes," *Bell Lab. Rec.* **60** (July/August 1982), pp. 137-140.
58. J. W. Gewartowski, "Microelectronics Replace Vacuum Tubes in the Microwave Network," *Bell Lab. Rec.* **61** (February 1983), pp. 17-21.
59. A. L. Schawlow and C. H. Townes, "Infrared and Optical Masers," *Phys. Rev.* **112** (December 1958), pp. 1940-1949.
60. T. H. Maiman, "Stimulated Optical Radiation in Ruby," *Nature* **187** (August 6, 1960), pp. 493-494.
61. K. C. Kao and G. A. Hockham, "Dielectric-Fibre Surface Waveguides for Optical Frequencies," *Proc. Inst. Elec. Eng. (London)* **113** (July 1966), pp. 1151-1158.
62. F. P. Kapron, D. B. Keck, and R. D. Maurer, "Radiation Losses in Glass Optical Waveguides," *Appl. Phys. Lett.* **17** (November 15, 1970), pp. 423-425.
63. R. D. Maurer, "Glass Fibers for Optical Communications," *Proc. IEEE* **61** (April 1973), pp. 452-462.
64. I. Hayashi, M. B. Panish, P. W. Foy, and S. Sumski, "Junction Lasers Which Operate Continuously at Room Temperature," *Appl. Phys. Lett.* **17** (August 1, 1970), pp. 109-111.
65. M. B. Panish, "Heterostructure Injection Lasers," *Bell Lab. Rec.* **49** (November 1971), pp. 298-304.
66. C. A. Burrus and R. W. Dawson, "Small-Area High-Current-Density GaAs Electroluminescent Diodes and a Method of Operation for Improved Degradation Characteristics," *Appl. Phys. Lett.* **17** (August 1, 1970), pp. 97-99.
67. D. W. Berreman, "A Lens or Light Guide Using Convectively Distorted Thermal Gradients in Gases," *Bell Syst. Tech. J.* **43** (July 1964), pp. 1469-1475.
68. D. Hondros and P. Debye, "Elektromagnetische Wellen an dielektrischen Drahten," *Ann. der Phys.* **32** (1910), pp. 465-476.
69. See reference 61.
70. See reference 62.
71. D. L. Myers and F. P. Partus, "Preform Fabrication and Fiber Drawing by Western Electric Product Engineering Control Center," *Bell Syst. Tech. J.* **57** (July/August 1978), pp. 1735-1744.
72. M. D. Rigterink, "Better Glass Fibers for Optical Transmission," *Bell Lab. Rec.* **53** (September 1975), pp. 340-345.
73. H. Melchior, A. R. Hartman, D. P. Schinke, and T. E. Seidel, "Planar Epitaxial Silicon Avalanche Photodiode," *Bell Syst. Tech. J.* **57** (July/August 1978), pp. 1791-1807.
74. See special issue on Atlanta Fiber System Experiment, *Bell Syst. Tech. J.* **57** (July/August 1978), pp. 1717-1888.

PART IV
THE TECHNICAL AND BUSINESS ENVIRONMENT

A program of studies and development applicable to all systems parallelled the design of specific transmission systems. Performance objectives were established for services as they evolved. In the 1960s and later, centralization and automation of operation and maintenance became a major development activity. Economic analysis was important at all times.

In the 1920s, circuits were rated only in terms of loss. Reference circuits were maintained and comparisons made to determine the loss equivalent for the various parts of a connection. When electronic gain became generally available, the loss of long circuits was kept low enough for acceptable quality, but increased with distance in the via net loss plan, so that delayed echoes would not be objectionable. Effective loss equivalents were determined for other characteristics, such as limited bandwidth, by subjective testing, using intelligibility as a criterion. Network transmission planning was on a statistical basis, taking into account the range of loss equivalents likely in each part of an end-to-end connection.

Voice-circuit bandwidth increased successively until, by 1937, it extended from about 200 to 3500 Hz, a band consistent with the 4-kHz spacing of carrier channels standardized that year. New telephone sets, the 300 type in 1937 and the 500 type in 1950, were designed to take advantage of the lower-loss trunks and improved transmission. After 1950, standards work included the evaluation of quality factors other than intelligibility and the development of combined criteria, including loss, noise, and echo, to achieve a high level of customer satisfaction with calls. Objectives were established for new services, such as television and data, and new media, such as satellite and digital systems, as they became significant factors in the network.

Switched message voice continued to be the principal service over the network from 1925 to 1975. Direct distance dialing by customers, initiated in 1951, connected trunks in tandem without the intervention of an operator and placed tighter requirements on the transmission plant. After 1950, special services, such as private lines and networks, often required special arrangements in signaling, equalization, and maintenance. Signals other than voice increased in importance. Telegraph code was carried from a very early date. In the 1920s, audio program circuits were provided to the broadcast radio networks and, in the 1930s and 1940s, television transmission influenced the design of coaxial cable and microwave radio systems. When television became popular, television operating centers established by AT&T coordinated program switching and monitored the condition of the television networks. An attempt to establish face-to-face visual telephone service in the 1960s and 1970s, however, was premature. Data transmission over voice circuits started in the 1970s. This placed new requirements on the transmission plant, as impairments that were of no consequence to voice signals proved to be important for data. Data transmission over purely digital facilities began in 1974 as the DATAPHONE* data communications service (DDS).

Testing, adjusting, and locating and repairing faults for all new facilities became important in development as systems increased in size and complexity. The expansion of the network made the development of operation and maintenance systems a high-priority activity in the 1960s and 1970s. Telemetry and computers centralized and automated activities for both switching and transmission facilities. Centralized systems were developed to test both switched trunks and private lines. As computer-based support systems proliferated in all phases of plant operation and maintenance, the transmission support systems were coordinated with others in overall operations plans.

Two major considerations governed the economics of transmission systems: Line costs per circuit in high-capacity systems were lower than in low-capacity systems, and the cost of multiplex terminals established a minimum distance below which multichannel systems were not economical. The economy of scale for long-haul carrier systems was spectacularly demonstrated from 1925 to 1975 as the cost per circuit was reduced by a factor of at least 100. The cost of terminals generally restricted multichannel analog systems to distances of 20 miles or more. In the 1960s, the low-cost digital terminals of T-carrier brought the break-even distance down to 15 miles or less. Rapid advances in the scale of integrated circuits further reduced digital terminal costs until, in the 1980s, most new urban trunks were realized by digital multiplexed systems.

---

* Registered service mark of AT&T.

Chapter 21

Objectives, Standards, and Transmission Plans

I. INTRODUCTION

Satisfactory voice transmission over local distances was achieved in the very early years of telephony. The major technical effort for the next 50 years was addressed to the conquest of distance—the extension of intelligible communication, first across the continent and ultimately across the oceans. Two major landmarks of this stage were the completion of the first transcontinental link in 1915 and the first transatlantic connection via long-wave radio in 1927. In the face of the near miracle of practically instantaneous voice communication over such large distances, it would have seemed petty to complain that the reproduced speech did not sound natural, or that there were disturbing sounds accompanying the speech.

Initially, attenuation was the overriding obstacle to overcome to permit intelligible communication. As advancing technology provided the means for reducing loss, attenuation no longer overshadowed all other types of impairments; there were new types associated with the techniques used to decrease loss. Whereas previously the available means were strained to the limit just to transmit intelligibly over a maximum distance, the new technology permitted trading off one kind of impairment against another and overall quality against cost. These possibilities created a need to measure the effect on overall quality of very different types of impairments such as loss, noise, limited bandwidth, echo and crosstalk in comparable terms. Justification of the cost associated with some improvement in overall quality required a quantitative estimate of the resulting increase in customer satisfaction. Furthermore, as transmission quality improved, so did customer expectations, which played an important role in determining customer judgment of circuit quality. The need for quantitative information regarding all these factors dictated repeated and increasingly sophisticated use of subjective testing.

As the network grew, it became increasingly configurable. Many alternative paths, utilizing a variety of different facilities of various types, could be used to connect two particular endpoints. In addition, as the switching became increasingly automatic, operators were no longer there to screen out unacceptable circuits before a customer encountered them. To determine quality in such an environment, it became necessary to characterize statistically the performance and type of connections the users were likely to encounter. Such information was obtained by surveys of the network. These surveys determined the probable frequency of occurrence and levels of impairment of each type associated with the various facilities that made up the network. Surveys also provided data regarding the frequency of occurrence of different numbers and types of facilities in tandem. By combining the data collected from network surveys with the results of subjective testing, it became possible to evaluate the quality of transmission offered to the customer population. More importantly, by identifying areas where improvements were desirable, such studies provided the basis for a transmission plan to guide the evolution of the network. Such evolution was implemented by specifying performance design objectives for new systems, maintenance objectives to control the variations in performance that occur in service, and a switching plan that controlled the facility interconnections that were allowed to occur.

Initially designed to carry voice, the network began to be used to carry other types of signals, such as music programs, telephoto, television, and data. The various kinds of impairments affected these new signals in markedly different ways. For example, the listener to a speech signal is very tolerant of envelope delay distortion and impulse noise. The ear is not sensitive to phase differences among the components of speech or to very short bursts of noise. In contrast, some of the new signals were often highly vulnerable to these impairments. New considerations and new techniques had to be incorporated into the network transmission-planning process, in recognition of the new types of signals.

Transmission planning evolved in response to technological advances, changing customer needs and expectations, and a changing economic climate. Planning also had to take into account the presence in the network of facilities from different technological epochs. It had to anticipate changes in needs and in the means available to fulfill them. The advance of technology made it necessary to take into account new types of impairment associated with new transmission media and signal types. Fortunately, technology also provided new test equipment and computers to control the acquisition of data and process it into useful forms, and thus made possible the establishment of timely performance and maintenance objectives.

Until 1934, the responsibility for guiding this evolution resided with AT&T, jointly shared by the Operating and Engineering Department and the Development and Research Department. A good deal of discretion for establishing local-transmission limits was left to the operating companies. However, it was recognized that the orderly evolution of nationwide telephone service required uniform standards and centralized planning, and general transmission and switching plans were developed at AT&T. In 1934, the AT&T Development and Research Department was transferred to Bell Laboratories, which had been founded nine years earlier. Bell Laboratories then became responsible for carrying out transmission studies and making recommendations based on these studies. Such recommendations were considered jointly with the Operating and Engineering Department at AT&T. The results of this process led to accepted objectives that were used to help set development priorities at Bell Laboratories and guide transmission planning by the operating companies.

II. OBJECTIVES AND STANDARDS FOR SPEECH TO 1930

2.1 Status in the Mid-1920s

In 1925, the basis for transmission planning was the Transmission Equivalent Plan. This plan began evolving shortly after the turn of the century, when all telephone circuits in commercial use had quite similar characteristics, and conditions were such that loss was the principal variable of interest to the transmission-planning engineer. The performance of circuits was determined by comparing them, on a loudness basis, with a reference circuit for which the loss was variable but whose other transmission characteristics were representative of the equipment of that date [Fig. 21-1]. (See another volume in this series, *The Early Years (1875–1925)*, pages 303-308 for a description of the Bell System Reference Circuit, later called the Standard Cable Reference System.)

2.1.1 Standard Cable Reference System

The reference circuit simulated an overall telephone connection, utilizing components (telephone sets, hybrid coils, transformers, and an adjustable artificial line) typical of those in use at that time. The rating of telephone circuits by comparing them to the reference circuit involved a talker speaking alternately over the circuit being tested and the reference circuit, and a listener switching similarly at the receiving end. The reference-circuit artificial line, calibrated in miles of standard cable, was adjusted until the listener judged the volume (loudness) of the speech sounds reproduced by the two circuits to be equal. The number of miles of artificial line in the reference circuit was then used as the transmission equivalent of the circuit under test.

Permissible transmission equivalents, determined for overall connections in this manner, were allocated to parts of connections, such as telephone sets, customer loops, and trunks. The allocations were developed in such a manner that transmission engineers could combine the ratings of elements to estimate the ratings for overall connections of interest. This provided a basis for allocating loss to the elements of the network in a manner consistent with realizable quality goals.[1]

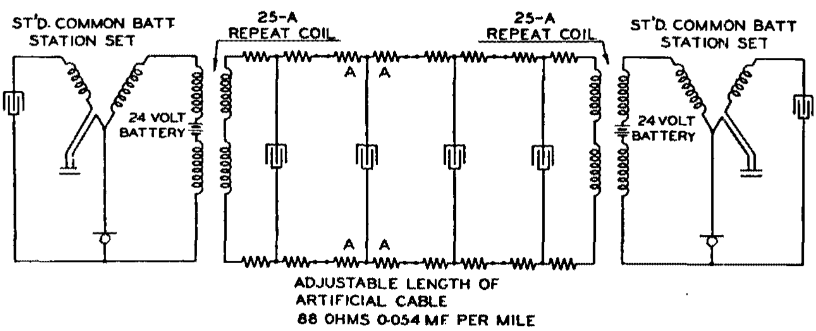


Fig. 21-1. The standard cable reference system, 1924.

2.1.2 Master Reference System—Transmission Units and Decibels

As telephone technology advanced in the 1920s, it soon became evident that the Standard Cable Reference System had major shortcomings. For example, as components such as telephone sets were improved, it was necessary to update the standard reference. A reference that was technology dependent and changed with time was obviously undesirable, and, in 1925, the Standard Cable Reference System was replaced by the Master Reference System for Telephone Transmission. This system replaced the miles of standard cable with the transmission unit (TU), which was later to become the decibel.[2,3]

The transmitter and receiver used in the new reference system were designed specifically to have very low distortion. The pronounced resonances or other features in commercial units could be simulated with networks inserted in the transmitting and receiving elements. Amplifiers compensated for the lower efficiency of the low-distortion elements [Fig. 21-2]. This arrangement led to a reference system in which transmission characteristics were more stable and reproducible. It also allowed modeling of the improved performance of the newer commercial telephones that were becoming available. Transmission engineering continued to be based primarily on loss, and the reference system was used to obtain planning information in the same manner as its predecessor. In 1926, the Master Reference System for Telephone Transmission was adopted

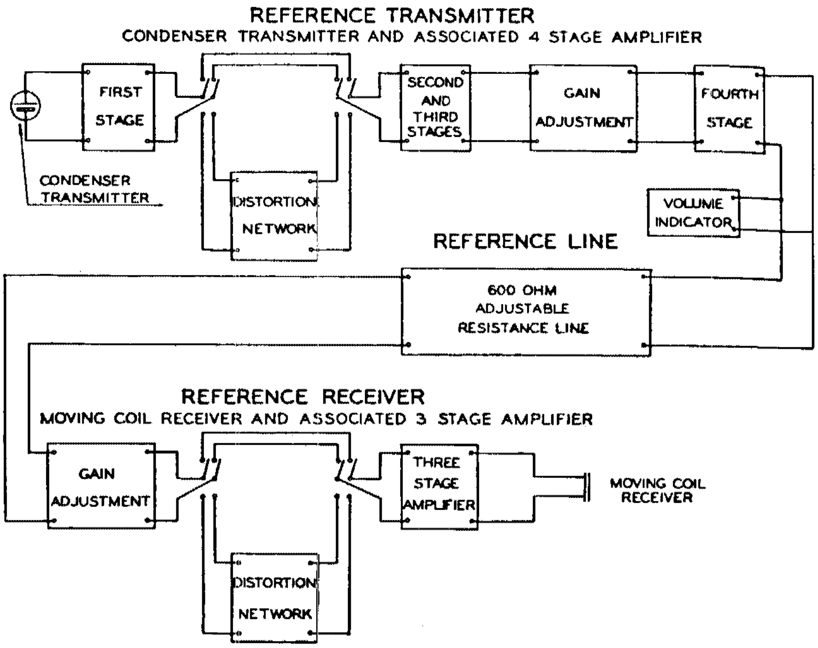


Fig. 21-2. Schematic diagram of the Master Reference System for Telephone Transmission, 1925.

as an international standard [Fig. 21-3]. The international reference systems in use for the following 50 years evolved from that original choice.[4]

At about the same time, it was noted that the loss of a mile of standard cable, with a resistance of 88 Ω per loop-mile and capacitance of 0.54 μf per mile, measured at 800 Hz, was equal to 10.56 $\log_{10} P_1/P_2$, where P_1 and P_2 are the input and output powers, respectively. In 1924, in a further step in making its standards less equipment specific, the Bell System adopted a new unit of attenuation, the "Transmission Unit," or TU, in which the ratio was exactly 10 $\log P_1/P_2$. This unit, renamed the decibel (a "bel" is $\log P_1/P_2$), was sufficiently close to the mile of standard cable so that little adjustment in thinking was required.[5]

2.1.3 Effective Loss

The Master Reference System for Telephone Transmission was a significant advance over the earlier standard, but, as the telephone system evolved further, it became apparent that a rating system that focused exclusively on loudness as a criterion of quality was inadequate. Loss-reduction techniques for circuits, such as loaded cable and voice-frequency amplifiers, provided a means for improving loudness. These techniques not only had the effect of reducing this

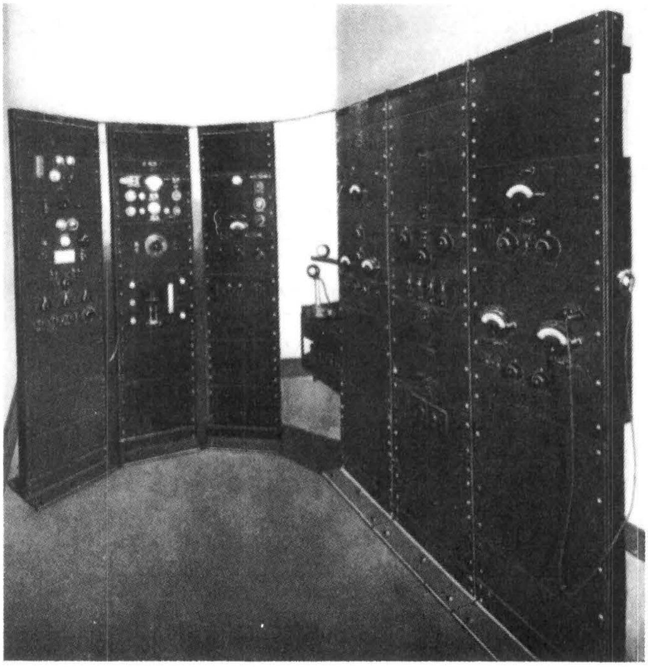


Fig. 21-3. Master Reference System for Telephone Transmission with associated calibration apparatus. This system was adopted as an international standard in 1926.

impairment but also permitted trading other characteristics, such as transmitter and receiver efficiency, against other types of impairments. As a result, telephone sets with flatter frequency characteristics, less distortion, and anti-sidetone circuits were being designed. (Sidetone is sound energy [speech and room noise] transferred from the transmitter to the receiver in the speaker's telephone.) Evaluation of the improved telephone sets in terms of the existing reference system tended to emphasize loudness differences at the expense of the other factors. A rating plan was needed that properly recognized the improvements, and, at the same time, retained applicability for the older types of telephone set. The effective-loss plan was devised to meet this need.

Ratings under the new plan were in terms of decibels of effective loss to distinguish them from loudness or volume losses of the old plan. Effective loss represented a figure of merit for evaluating the effectiveness of the transmission over telephone circuits and, as such, was a measure of the ability of telephone listeners to understand as well as to hear transmitted telephone speech. Effective loss data were expressed in terms of the effective loss in decibels introduced into the reference circuit to match an impairment (perhaps in reduced bandwidth) of the circuit under test. For example, an effective loss of 1 dB was assessed against a circuit under test if, for matched quality (that is, understanding), the loss of the reference circuit trunk was increased by 1 dB at all frequencies without other change. Any other change in a circuit that had the same subjective effect as this distortionless change in trunk loss was also defined as an effective-loss change of 1 dB. Pure loss was thus used as a unit to measure impairments that were physically quite distinct from loss in terms of their effect, such as noise, restricted bandwidth, or echo. Since the effect of an additional impairment depends on the level of impairment to which it is added, the nominal reference was chosen to have characteristics typical of a circuit in the middle of the distribution of existing circuits.

The objective measure of circuit performance adopted for the effective loss plan was repetition rate—the number of repetitions per unit time requested by telephone users while using the circuits in actual service. This measure was selected on the basis that it provided a direct quantitative measure of the success or difficulty with which telephone users carried on conversations.[6]

2.2 Transmission Design

This was all very well in concept. With a master reference system and the effective-loss equivalents, circuits could be designed to an acceptable effective loss. However, the general problem of the telephone-transmission engineer was not to design complete telephone connections one by one, but rather to specify rules for elements such as telephones, loops, and trunks in a way that would lead to satisfactory transmission in a high percentage of cases when these elements were combined to form a variety of overall connections. Each element had to be designed to perform satisfactorily with any of a large number of other elements—quite a complex undertaking. The planning engineer needed a large amount of data to accomplish this task. Appropriate information for the effective-loss plan consisted of families of curves and tables covering the range of effective losses ascribable to transmitting loops, receiving loops, trunks, terminal junctions, central offices, intermediate junctions, circuit noise, and room noise.

Transmission planning was in terms of limits. For example, the maximum allowed effective loss for a local central office area might be set at 15 dB. This value would apply, for example, to a connection consisting of a high effective loss transmitting loop, that is, one with some combination of long length, poor sidetone, high line noise, and the like, and a receiving loop of similar limiting nature. With this approach, transmission on the average would be much better than the limit, and, in general, limiting cases would still be acceptable.

Preparation of the information required involved a two-step approach. First, basic data for preparing effective-loss ratings were obtained from repetition counts made during a series of transmission observations on calls between telephone employees in the regular course of their business. These calls were made over special facilities that permitted the variation of the circuit characteristics over a wide range, and the loss ratings of the various conditions were determined from the observed repetition rates. During the tests, different types of instruments were used, and changes were made in the sidetone characteristics of the telephone sets and in the attenuation and attenuation distortion of the lines. Each type of impairment was covered over the whole range found in the telephone plant of that time and, to some extent, that expected in future plant. The infinite number of cases that resulted from combinations of different amounts of each type of impairment could, of course, only be sampled.

The second step involved specifying a reference circuit, similar in characteristics to some of the connections used in the repetition-rate tests, and developing measurement and computation methods for determining effective-loss ratings of other circuit conditions encountered in the plant, too numerous for direct measurement by repetition testing. The performance of a complete circuit could be described with sufficient accuracy for most engineering purposes as a function of volume loss, sidetone, distortion, and noise. The magnitude of all of these separate characteristics could be derived from computations based on physical measurements.[7]

2.2.1 Working Reference System

The reference system developed was called the Working Reference System [Fig. 21-4]. It consisted of representative limiting customer loops and telephone sets and a variable-loss distortionless trunk. The trunk simulation incorporated a 600-Ω variable attenuator calibrated in decibels, and a low-pass filter with a cutoff of 3000 Hz. The system introduced line noise of 100 noise units about 23 dBrnC0 at the receiver. (See Chapter 10, Section 2.2.1.) Use of the method described above enabled development of the appropriate data for plant design in terms of the connection elements (loops, offices, trunks, etc.). Of particular interest were the distortion-transmission impairment (DTI) and noise-transmission impairment (NTI) data. The DTI information was shown as a curve on a graph with an ordinate of DTI (in decibels of effective loss) and an abscissa of low-pass filter cutoff frequency. (Distortion was a much-used [and abused] term in transmission work. In the early days, it often was used to mean variations in the amplitude response as a function of frequency, that is, gain, or amplitude

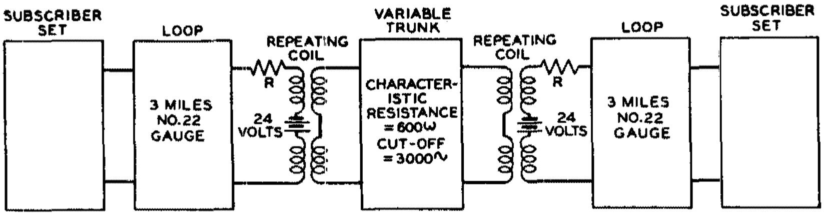


Fig. 21-4. Working Reference System for the specification of effective losses, 1933. The resistance, *R*, was 25 ohms as an allowance for office wiring and other equipment. Typical line noise and room noise were added at the receiver.

distortion. In the case discussed, the distortion was that due to the band-edge cutoff.) With this curve, the telephone engineer could determine the incremental effective loss due to a change in bandwidth and thus could take this factor into account in system design and network planning. The DTI information set the stage for the bandwidth objectives of cable-loading systems and carrier systems with wider voice-frequency bandwidths. The NTI information provided similar information on circuit noise and was an important factor in later studies concerning noise objectives.

2.2.2 *Room Noise*

The effective-loss concept also included provision for taking room noise into account. This recognized the important effect background noise had on the perceived quality of a telephone connection. Such noise can reach the listening ear of the customer via the telephone sidetone path or the acoustic leakage path between the earcap and the ear, thus directly affecting perception of the received telephone speech signal, or it can be perceived by the other ear as a distraction. Obviously, room noise could not be controlled by the telephone engineer, but information concerning its magnitude and effect was necessary to guide telephone-set design, to plan allowable circuit losses, and for service in particularly noisy locations.

To obtain quantitative data, Bell System engineers developed a sound meter incorporating frequency weighting, as determined from earlier laboratory tests, to measure effective loudness and time weighting to reflect the reactions of human hearing.[8] With this meter, extensive information on room noise magnitudes was reported in 1930.[9,10] This information enabled classification of the noise levels of different types of locations. The levels used in the Working Reference System in 1933, for example, corresponded to relatively quiet offices or relatively noisy residences.

2.3 Network Transmission Planning

2.3.1 *The 1930 Plan*

It had long been recognized that the network had to provide adequate service to customers with a wide range of talking volumes over a variety of facilities. Systematic network and transmission planning dates back at least to 1906, but early planning was limited by the restricted range of what could be done. By 1930, new telephone instruments, readily available gain in vacuum-tube amplifiers, and the advent of carrier systems made many more options available. A new network plan and the associated transmission objectives evolved in parallel with the effective-lossplan. A 1930 plan proposed a four-level switching hierarchy and noted that the network would probably be converted to the new plan in about five years [Fig. 21-5].[11] At the time, there were about 6400 central offices in the United States and eastern Canada, of which about 2500 were intercity toll centers. The plan envisioned establishing about 150 of these offices as primary outlets and eight as regional centers. Since a switching plan determines how the plant is used to achieve a desired connection, it determines the frequency with which various numbers of facilities are connected in tandem and thus the overall quality of transmission provided.

In 1932, the transmission plan to be associated with the new switching plan was further detailed. It was to be based on the use of loaded-cable circuits and lines incorporating voice-frequency repeaters; carrier was not yet a major factor and did not impose any new limitations. The extensive use of loading and repeaters permitted reduction in circuit losses to a point where other factors, such as echoes, singing, crosstalk, and noise, rather than loss, were controlling.[12]

2.3.2 Minimum Loss, Echoes, and Via Net Loss

The ability to insert gain in a connection meant that it was no longer adequate to control loss by holding it below some maximum value. It was possible to have too little loss or even net gain. Unavoidable impedance mismatches along the transmission path reflected the signal back toward the talker, who heard the reflection in his/her receiver as a talker echo. Shuttling echoes could also result in the arrival of multiple versions of the speech, displaced in time, at the listener's receiver as listener echo. Such echoes were increasingly disturbing as they became louder and were displaced further in time from the main signal.

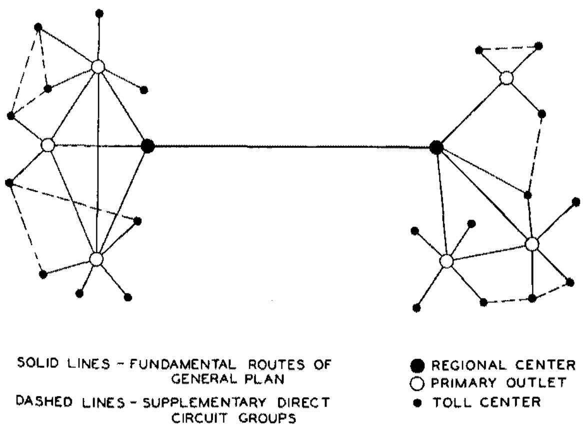


Fig. 21-5. The general toll switching plan, 1930. The diagram shows only the upper three levels of switching offices.

To a talker, a nearly instantaneous echo is indistinguishable from the sound of his or her own speech. However, an echo arriving when the talker is enunciating a subsequent portion of speech is disturbing. Indeed, if the volume and displacement of the echo are sufficiently large, the echo will interfere with the speech process, making it difficult or impossible to talk over the connection.

Since an echo must make a round-trip between its source and the point of impedance discontinuity, it undergoes double the attenuation between these two points. In the days before effective repeaters when losses were relatively high, echo was not much of a problem. However, the use of voice-frequency repeaters not only reduced loss but inserted impedance discontinuities that had not previously existed. In addition, the extensive use of loaded cable reduced propagation velocity and increased echo delay by a large factor compared to nonloaded open-wire lines. It therefore became necessary not only to maintain loss below some maximum value but also above some minimum. Further, the longer the circuit, the greater the displacement of the echo from the original signal and the greater its disturbing effect. It was necessary to require that the minimum loss increase with circuit length.

Other considerations also required that loss be above some minimum value. If there happened to be gain between two impedance irregularities, the circuit could become unstable and oscillate (sing). Even a circuit that did not sing but approached an unstable condition began to sound hollow and could sing momentarily if some transient condition temporarily reduced loss. Loss had to be kept at a level that maintained an adequate margin against these undesirable effects. Finally, amplification not only increased the level of the signal but also that of the noise. Though this did not degrade the signal-to-noise ratio when the listener was receiving speech, it did increase the annoying effect of the noise in silent intervals.

The trade-off between the beneficial effects of loss in controlling these degradations and its negative effect with regard to volume was a primary reason for adopting effective loss as a common unit of quality. The 1932 plan used the so-called "via net loss" plan for controlling loss on the toll network. The plan associated a loss, which increased linearly with length, to toll facilities. This length-dependent loss was supplemented by an additional constant loss, inserted by pads located at the two ends of a toll connection. The latter ensured adequate loss in short connections where singing margin and crosstalk were the controlling factors. The via net loss, varying linearly with length, provided adequate loss for echo control and margin for the statistical variations that occurred when a number of facilities were connected in tandem. Beyond a certain length, the loss required to control echo became excessive because of the consequent low volumes. For such long circuits, loss was not increased and echo was controlled by inserting an echo suppressor that opened the return path (i.e., in the listener-to-talker direction) when a talker was speaking. However, early echo suppressors did not work ideally and introduced impairments of their own in addition to increasing circuit costs. Their use on loaded cable was initially limited to circuits longer than about 500 miles where the via net loss plan would otherwise have excessively reduced volumes. As loaded cable was replaced by more modern high-velocity facilities, it became possible to increase the distance.

During the 1930s, further calculations, measurements, and subjective testing were carried out to refine the concept of effective loss. Additional room-noise data were collected using improved sound-level meters.[13,14] The Bell System sound-meter design was adopted as a national standard for use in broadcast radio, recording, and other audio applications, as well as in telephone engineering. The techniques and approaches that had evolved to become part of the 1932 plan provided basic concepts and tools that guided toll-transmission planning and the evolution of the transmission plant throughout the period of this history.

III. FURTHER EVOLUTION OF STANDARDS FOR SPEECH TRANSMISSION

3.1 Carrier Systems and Channel Spacing

In the latter part of the 1930s, as more modern carrier systems began to be introduced into the network and as it became evident that these would be widely interconnected, it became necessary to establish a standard channel spacing. At an international conference on communications in 1938, the Bell System's proposed 4000-Hz spacing was accepted and has since become a worldwide standard. The Bell System's interest was in international standardization, since it was already widely interconnected overseas by high-frequency (HF) radio. The plea for 4000-Hz spacing (there was a group favoring 3000-Hz) was supported by demonstrations of transmission over 2700-, 3800-, and 5500-Hz bands. Within the 4000-Hz spacing, a low-frequency cutoff much below 500 Hz did little to improve intelligibility, but extending the band to 250 Hz considerably improved naturalness. There was also little to be gained by a cutoff much above 3500 Hz, whereas a high frequency limit of 2700 Hz caused a noticeable decrease in intelligibility.[15] (Discussion of such issues were held from an early date in a sequence of international committees, generally known by the initials of their full names in French. By 1938, this was the CCIF and later the CCITT, a branch of the International Telecommunication Union, itself a part of the United Nations.)

The 4000-Hz spacing was the result of balancing quality and cost, taking into account the following considerations: (1) The wider the channel transmission band, the better the quality [Fig. 21-6]. (Syllabic articulation, the measure of quality, was the ability to understand unconnected words and syllables. It was closely correlated with the repetition-request measures used earlier.) (2) For any given effective channel bandwidth, the larger the spacing between channels, the lower the cost of the filters in the multiplexes that stacked the channels into the carrier band. This followed from the fact that the larger spacing allowed wider guard bands between channels and therefore less steep cutoff slopes in the channelizing filters. (3) For a fixed total facility bandwidth, the wider the channel spacing, the fewer channels realized and therefore the greater the cost per channel. Alternatively, if the carrier bandwidth were increased for a given number of channels, more repeaters of a more difficult design would be required to handle the increased loss and bandwidth.

The quality of long toll circuits had been improving steadily since the establishment of the first transcontinental circuits in 1915 [Fig. 21-7]. Improve-

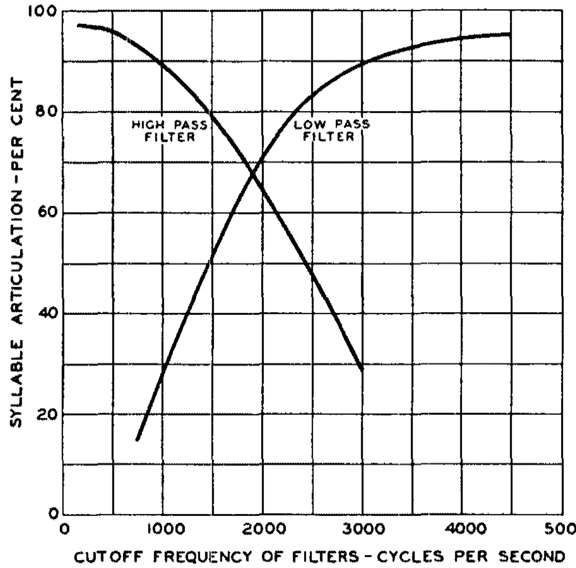


Fig. 21-6. Effect of cutoff frequency on speech quality, 1937. Syllabic articulation was a measure of the ability to understand unconnected words and syllables.

ments were due to successively better repeaters, the elimination of loading on open wire, improved loading and repeaters on voice-frequency cable, and, finally, the introduction of broadband carrier on pairs and coaxial cables. The end-to-end quality still depended on the characteristics of the telephone sets

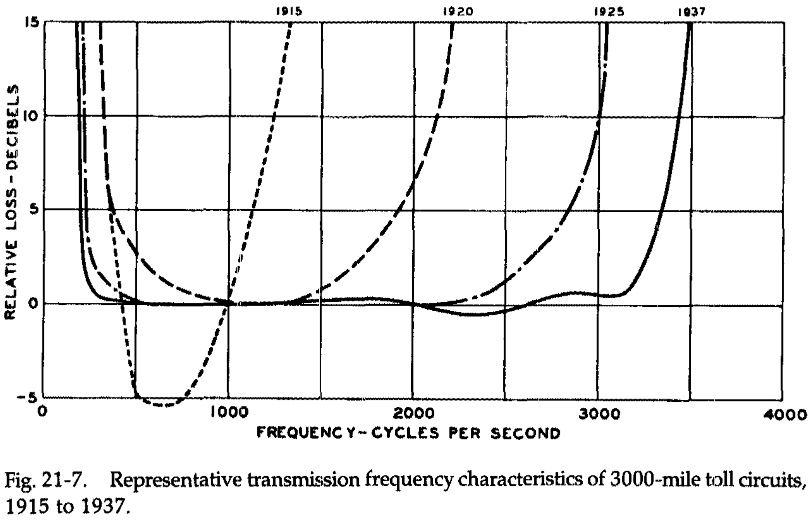


Fig. 21-7. Representative transmission frequency characteristics of 3000-mile toll circuits, 1915 to 1937.

and of the loops and connecting trunks to the toll links, but these, too, were improving in quality during the same period. By the end of the 1930s, the objectives for five toll systems in tandem, a number expected to be only occasionally exceeded, had been set: (1) The net-loss variation at 1 kHz would not exceed ±3 dB. (2) Transmission was to be no more than 10 dB down at 200 and 3300 Hz. (3) The attenuation distortion over the band from 400 to 3000 Hz was not to exceed ±2 dB relative to the 1-kHz loss. (4) The noise at the receiving toll office (the −9-dB transmission-level point) would not exceed 29 dBa (dBa was an early noise measure; 29 dBa equaled 44 dBrnC0). The carrier systems of that time generally met these objectives [Fig. 21-8].

The channel-spacing standard for analog systems has remained at 4 kHz, and the speech sampling rate of 8000 per second for digital systems was selected to provide comparable performance. An exception to the standard was made in the design of the analog submarine-cable systems, the first of which was installed in 1956. Because of the high cost of these systems, nonstandard multiplex terminals and relatively expensive sharp-cutoff filters were used. With these, a 2900-Hz band was achieved between carriers spaced at 3000 Hz, providing an increase of 33 percent in channel capacity. The increased multiplex cost and the administrative complexities of the nonstandard multiplex were judged to be acceptable because of the high value of the extra channels they made possible. (The first submarine cables were originally equipped with 4-kHz-spaced channels and later fitted with the 3-kHz high-efficiency channel banks. There were skeptics about the wisdom of this exception to the standard channel bandwidth. Back-to-back 3-kHz/4-kHz channel banks were required near the cable landings for overland transmission and distribution by standard facilities. As high-speed data and other special signals began to be more commonly transmitted, it was feared the transoceanic links would prove to be a bottleneck. The cables could, however, be refitted with 4-kHz channels, if required).

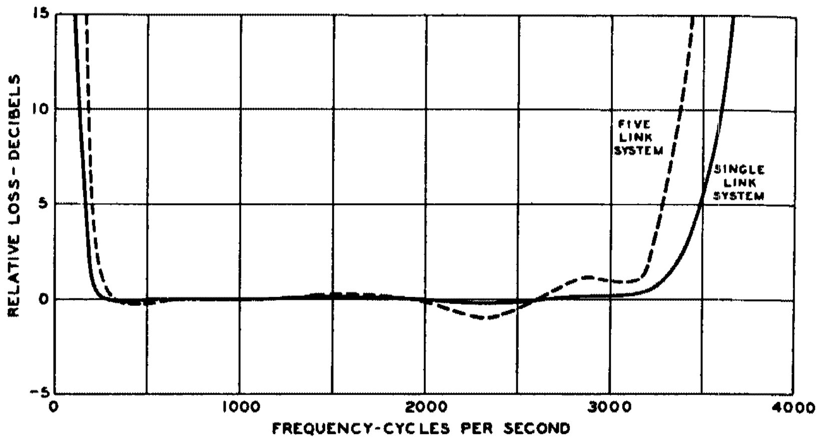


Fig. 21-8. Transmission frequency characteristics of broadband systems, 1937.

The years of World War II resulted in a hiatus in transmission planning, as it did in all other civil telephone-system development activities. In the plant, some quality objectives were temporarily sacrificed to satisfy urgent telephone service needs with the severely restricted resources. For example, split *emergency-band* channels with half the bandwidth required for high quality were used on some facilities to double the number of circuits that could be carried.

3.2 Improved Telephones

Parallel with the improvements in interoffice transmission, major improvements were made in the telephone station instruments. The candlestick deskstand, prevalent in the 1920s and 1930s, was highly resonant, emphasizing frequencies around 1000-Hz near the center of the speech band. The new low-loss trunks reduced the need for high midband (1000-Hz) efficiency and led to a much improved telephone, the 300 type, introduced in 1938.[16] In the 300-type set, the sharp 1000-kHz resonance, characteristic of the older sets, was eliminated and the overall response much improved [Fig. 21-9]. After World War II, the explosive growth driven by the pent-up demand resulted in a rapid upgrading in the average transmission quality. A high percentage of the equipment installed was of the most recent vintage. Most of the trunk growth was

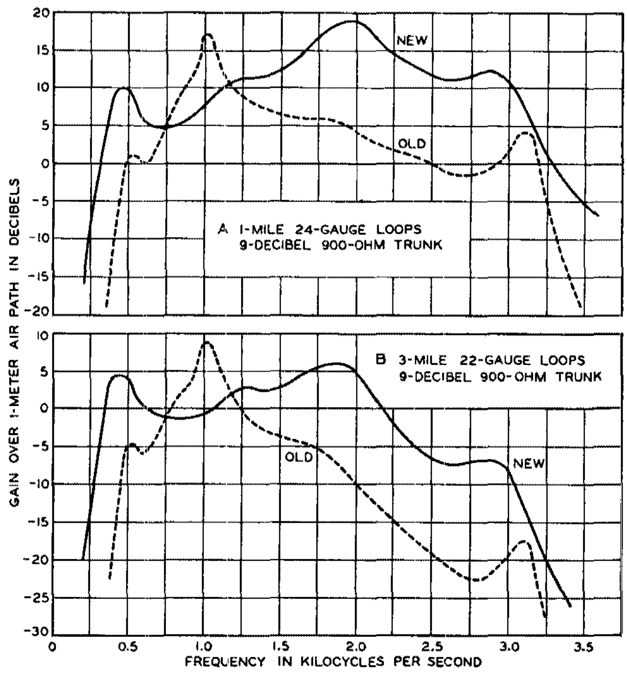


Fig. 21-9. Overall frequency response of typical telephone connections with the old candlestick deskstand telephone and the new 300 type telephone set, 1938.

on modern carrier systems, and, by 1950, 70 percent of the telephones in service were of the 300 type.

During the 1930s and the post-war years, research by H. Fletcher and his associates at Bell Laboratories contributed fundamental insights into the acoustics of telephony and the nature of speech and hearing. This knowledge was applied in the design of the 300 type and later telephone sets.[17,18,19] In the early 1950s, the 500 set, with transmission characteristics similar to the 300 type but with the added feature of automatic compensation for the loss of long loops, became available [Fig. 21-10].[20] This capability reduced the variation in loss for different stations and made it possible to simplify the design of local loops. Where previously loop design had required the selection of wire gauges, taking into account transmission and powering constraints, the design could now be

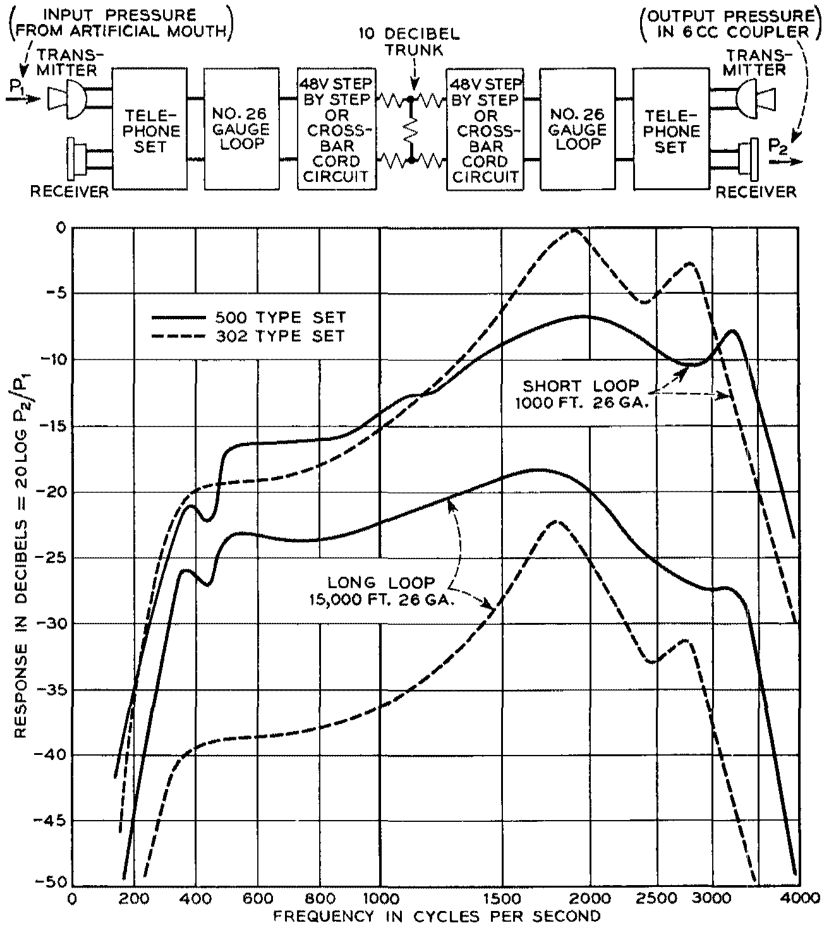


Fig. 21-10. The transmission response of the 500 type telephone and the older 302 type on short and long loops, 1951.

based entirely on just the resistance between the telephone set and the central-office battery.[21] Over 25 years of experience verified the viability of the resistance design method, which remained in use into the 1970s.

3.3 Postwar Network Planning

By 1950, machine switching had become almost universal in local areas, and its use for toll connections was growing rapidly. This led first to operator toll dialing and then to an intensive program to implement nationwide customer dialing of long-distance calls (dubbed DDD, for direct distance dialing). These developments, in turn, required a nationwide dialing plan with significant implications for transmission planning. The plan was originally based on a six-level hierarchy with a possible maximum of ten trunks in tandem between end offices. This was simplified in the plan finally adopted to a five-level hierarchy with a maximum of nine tandem trunks [Fig. 21-11].[22,23]

Fortunately, at the same time that the new dialing plan required lower and more uniform loss in trunks, the extensive use of modern carrier systems permitted an updating of the via net loss plan introduced in the 1930s. By 1948, carrier-derived toll circuit mileage exceeded that on voice-frequency facilities, and, by 1960, voice-frequency facilities made up only 15 percent of plant capacity, concentrated almost entirely in short links. The shorter delay of the high-velocity carrier systems permitted echo control with lower losses. Via net

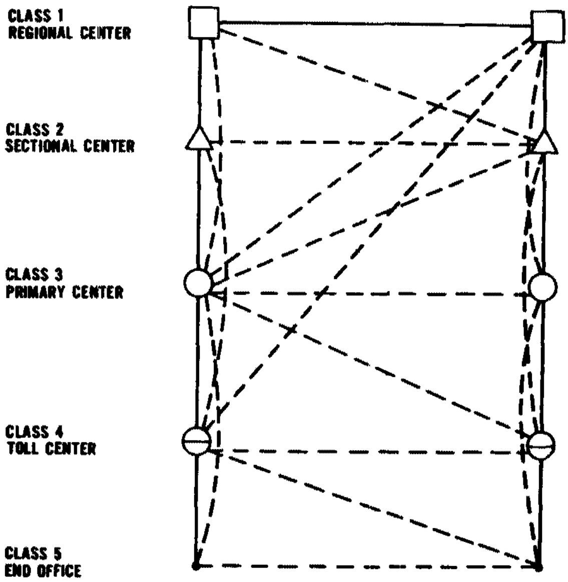


Fig. 21-11. Five-level toll switching plan in use from the 1950s. A variety of routings was possible with a maximum of nine trunks in tandem. [Andrews and Hatch, *IEEE Trans. Commun. Technol.* **COM-19** (1971): 303.]

loss varied from 0.5 dB for short trunks to only 3 dB for trunks above 1500 miles, and the distance beyond which echo suppressors were required was increased to 1850 miles.

As the number and variety of trunks connected in tandem were increased in the 1950s by automatic switching, their variability as well as their average loss became a matter of concern. Furthermore, as customers began to dial their own calls, operator supervision over the quality of a connection prior to handing it over to the customer was no longer possible. These factors put a premium on designing and maintaining the transmission plant to tighter limits. Plant surveys indicated that the standard deviation of trunk losses was about double what was considered desirable in this new environment. Improvements in system stability, careful loss adjustment at the time of system installation, routine-maintenance procedures, and routine measurements making increased use of automatic testing were introduced. By the early 1960s, average loss and variation close to the objectives had been obtained through these means.[24]

3.3.1 *New Rating Criteria*

The postwar sequence of developments and rapid evolution of the plant also led to a reexamination of the basic criteria for assessing transmission quality. As always, customer expectations rose about as fast as the system performance improved, and mere intelligibility became an inadequate measure for establishing circuit quality in subjective testing. Intelligibility was taken for granted, and, instead of just a count of the number of requested repetitions, more general judgments were elicited from subjects. In 1958, laboratory subjective tests were made to relate received speech volume over circuits with characteristics typical of that time to the satisfaction of telephone listeners. The listener ratings were expressed as excellent, good, fair, poor, and unsatisfactory.[25] The distribution of received volumes was computed from statistical data on loss from central office to central office and was also obtained from a survey of near-end central office volume conducted in 1950 and 1951.[26] A volume grade-of-service could then be obtained from the distribution of subjective ratings at a given volume and the distribution of volumes [Fig. 21-12]. Design objectives could be set such that, for example, only a negligible number of calls would be rated poor; no more than five percent rated fair; and the balance, good or excellent.

As more stringent objectives brought received volume into the desired range, the emphasis shifted to improving noise performance. An approach similar to that for received volumes was used to determine the effect of noise and its impact on circuit quality.[27] A complete characterization of the perception of noise and its interfering effect is extremely complex. Fletcher and his colleagues had shown that the perception of sound depends not only on the volume level but on the frequency as well [Fig. 21-13].[28] (These are the famous Fletcher-Munson equal-loudness curves, beloved of audio engineers and incorporated in the loudness control of many high-fidelity amplifiers.) The telephone instrument also shapes the speech and noise spectrum. Telephone channel-noise measurements and evaluation had long recognized these factors, and weighting curves were applied to allow for the combined effect of the telephone instrument and human perception. These were updated from time to time as improved

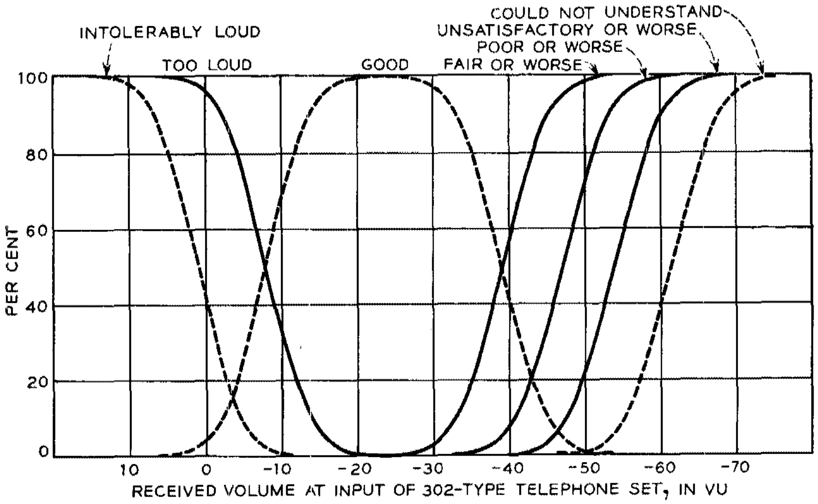


Fig. 21-12. Opinion distributions used to determine volume grade-of-service, 1958.

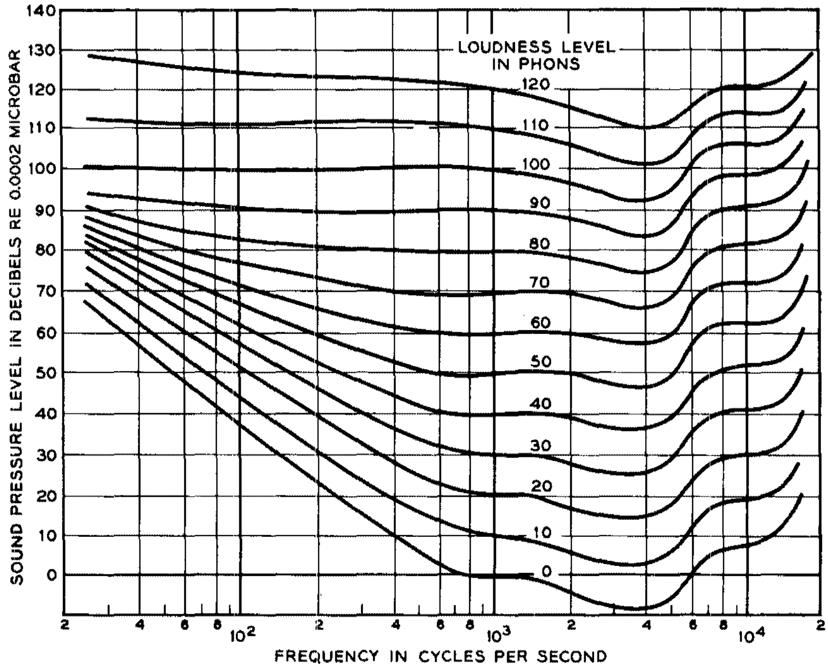


Fig. 21-13. Fletcher-Munson equal-loudness contours for pure tones. A phon is a unit of loudness.

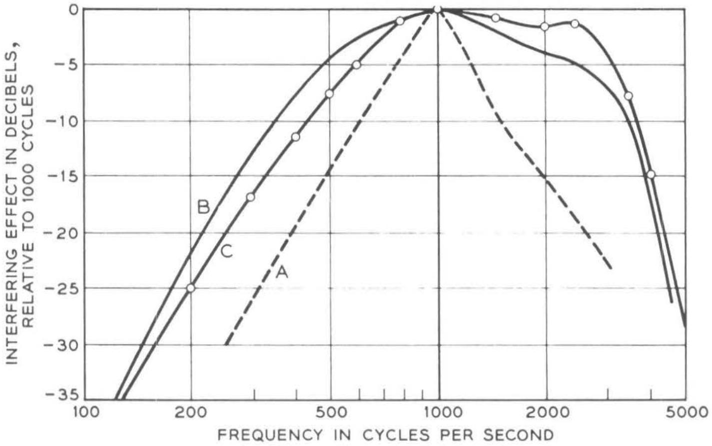


Fig. 21-14. Noise weighting versus frequency. Curve *A* is 144-line weighting, related to the response of the candlestick deskstand, Curve *B* is F1A-line weighting for the 302 type telephone set, and Curve *C* is C-message weighting appropriate for the 500 type set.

telephone sets were introduced. The range of volumes in the improved plant was small enough that a single representative curve could be used for the perception factor [Fig. 21-14].

A new noise meter, the 3A noise-measuring set, was developed including a C-message-weighting network that represented the combined frequency characteristics of human sound perception for the average received volume and the 500 type telephone [Fig. 21-15].[29] The new measuring set was used in extensive field surveys in the early 1960s to gather noise data and to correlate them with subjective tests made in the laboratory.[30] The subjective tests for

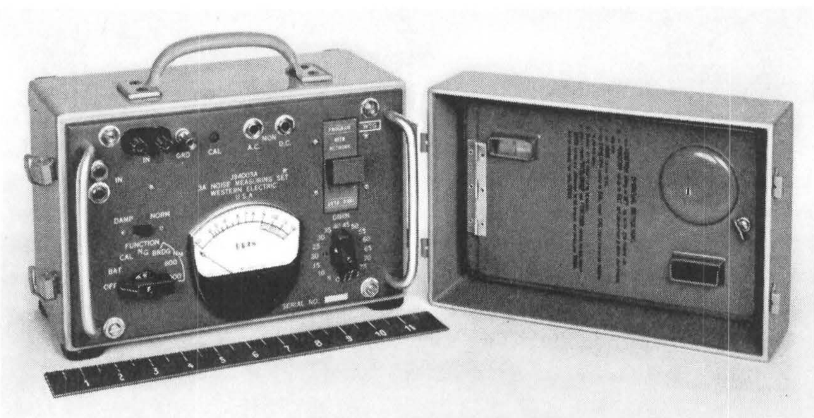


Fig. 21-15. The 3A noise measuring set.

noise were made in a manner similar to that of the laboratory volume evaluations. Volunteer subjects using a 500 type telephone set listened to controlled-volume speech with various levels of circuit noise and were asked to rate each condition according to the five-category scale: excellent, good, fair, poor and unsatisfactory [Fig. 21-16]. Results were summarized in terms of opinion distributions [Fig. 21-17]. The combination of survey data and subjective results was then used to derive a noise grade of service, analogous to the volume grade of service discussed previously.[31]

As a result of these studies, new performance objectives were recommended for noise. These were originally expressed separately for customer loops, for short, medium, and long toll connections, and for intercontinental connections. By 1970, these performance objectives were used in the formulation of new long-range design and performance objectives for short- and long-haul transmission facilities.[32] Mean performance objectives for short-haul facilities were specified as 28 dBrnC0 at 60 route miles and for long-haul facilities as 34 dBrnC0 at 1000 route miles. For other long-haul facilities, the objective was in proportion to the length, that is, 3 dB more noise for doubling the distance. The performance objective for a 4000-mile transcontinental coaxial carrier system was, therefore, 40 dBrnC0. Design objectives for new long-haul microwave-radio systems were set approximately 1 dB higher.

Both objectives were intended to be worst-circuit limits. The average noise was projected to be better and about the same in both coaxial and radio systems, and later field surveys confirmed that indeed it was. The new objectives required

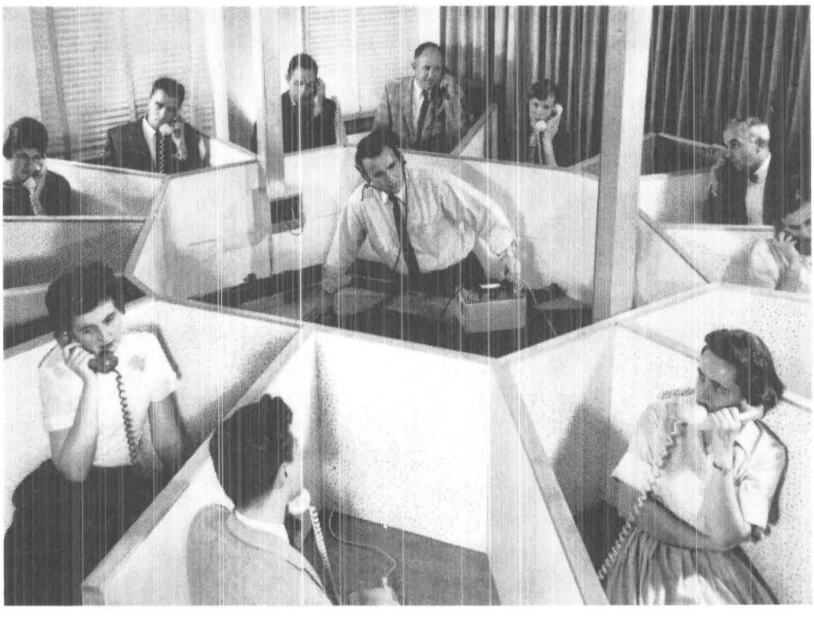


Fig. 21-16. Observers at multiple listening positions during subjective tests to rate the effect of circuit noise at controlled speech signal volume, 1964.

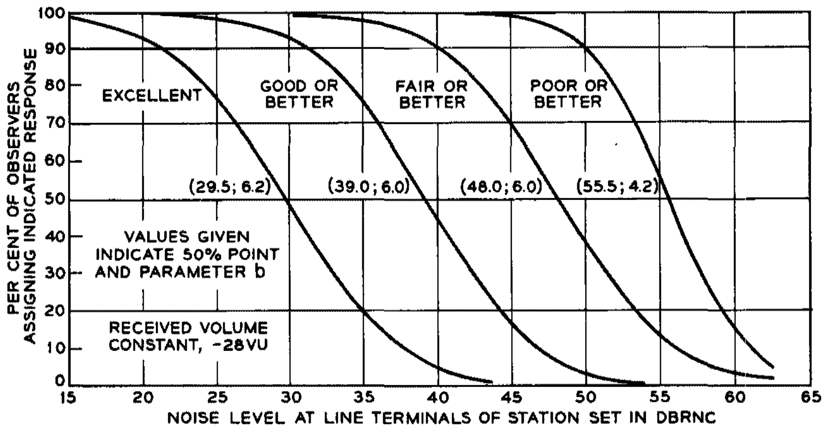


Fig. 21-17. Noise opinion curves used to determine noise grade-of-service.

noise performance about 3 to 4 dB better than that of systems designed to the earlier objectives. The noise objectives for intercontinental submarine-cable systems were set about 3 dB lower than the transcontinental objective because they were likely to be used in extremely long connections with long overland extensions on one or both ends. At the same time, the need to control noise in subscriber lines was recognized. The long-term objective for subscriber line noise was set at 20 dBrnC at the telephone-set line terminal.

Design objectives, if met, determined the initial performance of a system, but long-term quality depended on maintenance as well. Studies of noise-maintenance limits for trunks were also undertaken during the early 1960s and included in *Bell System Practices* by the middle 1960s. Routine measurements based on these limits were used to compute the noise component of a trunk-maintenance index, which was initiated in 1967. This index had only minor modifications in the years that followed. The noise-maintenance limits established for customer loops were also included in *Bell System Practices* in 1967. (No industry is measured more thoroughly and continuously than the telephone business. The telephone plant runs on indexes; almost every aspect of performance is the subject of comprehensive measurements embodied in an index. Although their value was occasionally scoffed at, they gained new respect when a review of service disasters in the early 1970s clearly showed they were foretold by declining indexes.)

3.3.2 *Combined Measures, Loss, Noise, and Echo*

In the mid-1960s and early 1970s, the approach used for establishing volume and noise objectives was further refined and supplemented by additional surveys and subjective testing. Because the variability of received speech volumes at the near-end central office was decreasing and because of the difficulty in obtaining consistent results from different observers, the emphasis shifted from measuring volume to measuring loss. Attenuation could be measured more accurately than speech signal volume and was under direct control of the transmission-planning engineer. In addition, subjective tests were made in the laboratory to determine the combined effect of impairments. Initially, the simultaneous effect of a range of loss and noise was measured; later, talker echo was included in a three-component combined effect. Some of these results were obtained using a special test facility, which allowed insertion of controlled amounts of different transmission impairments during the normal business calls of Bell Laboratories employee volunteers. On the basis of these results, loss-noise grade of service largely replaced the earlier noise and volume grade of service measurements. Results were summarized in contours showing the percentage of listeners rating a connection good or better for combinations of loss and noise over a wide range [Fig. 21-18].[33]

Field surveys to update the characterization of the telephone plant were carried out in parallel with the new subjective testing.[34] Results from these surveys were then used to construct computer models of plant performance. Perturbation of the various performance parameters permitted evaluation of the impact of changes in performance or maintenance guidelines on grade of service. Computer modeling was of special interest because of the rapid growth of digital systems with somewhat different impairments and distributions, compared to the older analog network. Using the new approaches and data, a recommended transmission plan for the evolving analog-digital network was formulated in 1976. For the connections used in this study, a new optimal loss was determined to balance the degradation due to talker echo and the degradation caused by connection loss. The study showed that, although the via net

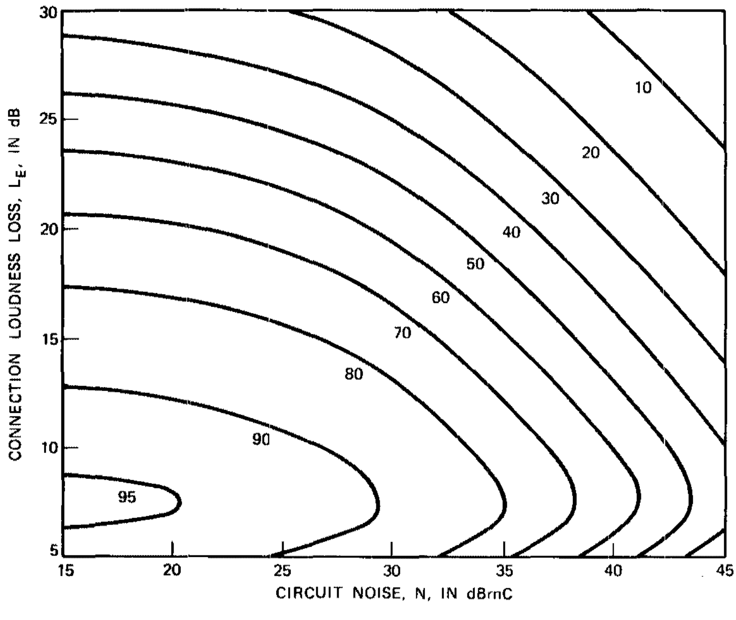


Fig. 21-18. Subjective-opinion contours of percent good or better for loss and noise.

loss plan provided more than optimum loss for short connections, the difference was not great enough to have any appreciable effect on overall transmission grade of service, and concluded that the via net loss plan need not be changed.

3.4 A Half-Century of Improvement in Transmission Quality

Transmission planning over the years evolved to take into account advances in technology, customer needs and expectations, and economics. There was much change from 1925 to 1975. Is it possible in simple terms to measure the corresponding improvement in transmission performance provided to the customer?

Qualitatively, the effect of technological advances, such as more modern telephone instruments and carrier systems, can be described, but it is difficult to express quantitatively the resulting improvement in transmission performance. The reason for this is that the criteria of performance (for example, received speech volume, loss, repetition rate, and quality judgments) have also changed and evolved over the years. However, the improvement in transmission performance provided to customers can be estimated using the effective-loss concept. Although it is obsolete and of little value for guiding new development, effective loss does provide a single measure, however crude, for the different types of impairments. With these reservations in mind, it may be used to provide a rough means to quantify the long term improvements.

This was done to estimate the total improvement for the period from 1925 to 1965 [Fig. 21-19].[35] During this period, average effective loss for local con-

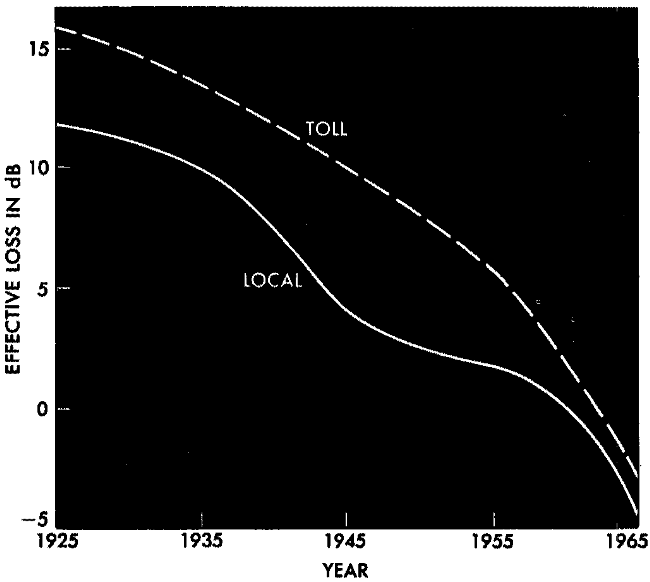


Fig. 21-19. The estimated improvement in transmission quality achieved between 1925 and 1965 in terms of effective loss. The curves should be interpreted as reflecting long-term changes only.

nections was reduced by about 17 dB, while the corresponding reduction for toll connections was about 19 dB. Advances in telephone-set design contributed substantially. For example, the improvement from 1925 to 1950 was occasioned in large part by the change from early deskstand telephones with their highly peaked gain-frequency characteristics to the 300 type. Subsequent improvement was due in part to the introduction of the 500 type telephone set and the use of the resistance-design method for customer loops. A substantial part of the toll improvement after 1950, however, is due to low-loss, low-noise, and uniform carrier systems.

Transmission performance continued to improve from 1965 to 1975 and beyond. Some of the changes did not appreciably alter average performance from that achieved by 1965, but resulted in improved performance for many situations in which the transmission performance had been below average. The rapid deployment of digital T carrier in the exchange area and for short-haul toll improved the quality of shorter-distance calls and helped to raise the overall average. Finally, control of impairments primarily important to data, such as impulse noise and envelope delay distortion, also resulted in improved voice transmission as a secondary benefit. But these improvements are not readily expressed in effective-loss terms.

IV. NEW SIGNALS AND NEW TRANSMISSION MEDIA

Analog voice signals were the principal signals transmitted during the entire first century of telephony, but DC telegraph signals were also carried from the earliest days. When carrier systems were introduced, telegraph signals were transmitted even more efficiently by using modulation schemes that gave the signals characteristics similar to voice. Still pictures (telephotographs) and rudimentary television signals were transmitted in the 1920s. Developments outside the Bell System to provide broadcast television progressed rapidly through the 1930s, and, by the end of World War II, transmission for television networking became a major Bell System concern. In the 1960s, the growth of data transmission and the introduction of new types of facilities, such as satellite and digital systems, brought with them new considerations that had to be recognized in the establishment of objectives and standards and incorporated into transmission planning.

4.1 Objectives for Television Transmission

The first system for television transmission, using both open wire pairs and radio facilities, was demonstrated by Bell Laboratories in 1927.[36] In the 1927 experiment, a neon-tube display provided a picture of only 50-line definition and transmission required a band from 10 Hz to about 20 kHz. This was no greater than the band used for the three-channel Type C carrier, but the ratio of the lowest to the highest frequency was very large. Within that band, transmission had to be flat in gain to within ±2 dB and in delay to within ±16 microseconds. The analysis of the equalization required and its achievement were a significant part of the technical advance represented by the demonstration. Work on the analysis of picture signals continued, and, in 1934, P. Mertz and F. Gray published a classic paper on the properties of an electrical signal formed from a scanned image.[37] This work placed the derivation of requirements for television transmission on a sound analytic basis.

The early interest of the Bell System in television was in point-to-point transmission and in its use as a possible adjunct to telephony. For this application, a picture of moderate definition would perhaps have been satisfactory. By about 1934, however, television technology had progressed sufficiently to indicate that broadcasting of much higher-definition pictures was likely in the future. As a result, the emphasis of the work in Bell Laboratories was changed to the problems of providing an intercity network of much broader-band channels. The new objective also fit well with the broadband capabilities of the coaxial-cable system, which was then in an early stage of development.

A very large amount of information must be sent rapidly to represent a moving picture. As a result, a television channel had to have a bandwidth on analog systems about 1000 times greater than that required to transmit voice. Furthermore, for a satisfactory picture, the gain and delay characteristics over this large bandwidth had to be flat to within very tight limits. The first successful transmission over the New York–Philadelphia experimental coaxial line in 1937 was of a 240-line picture at 24 frames per second and required transmission of a signal containing frequencies from DC to 806 kHz (translated to the range from 120 to 950 kHz for transmission over the coaxial line). Within that band, equalization of gain to within ± 1 dB and of delay to within ± 0.3 microsecond was required.[38] Interference requirements specified random noise at least 40 dB below the picture signal and single-frequency interference 55 dB below maximum signal amplitude.

In the late 1930s, further improvements in picture quality led first to a standard of 441 lines and 30 frames per second, which required about a 2.7-MHz band and, after World War II, to the current American standard of 525 lines and 30 frames per second, which requires about a 4-MHz band. The interference requirements established for the early coaxial tests survived with relatively minor modifications into the era of almost universal television distribution. Transmission of a picture to these standards was accomplished over a round-trip loop of the New York–Boston TDX microwave-radio system during its first public demonstration in 1947. Comparison of picture quality before and after transmission over this loop showed almost no perceptible impairment in the television image.[39]

In the years just before and following World War II, extensive subjective testing was carried out on every type of impairment to television images to provide a basis for setting transmission objectives and requirements.[40-45] In 1950, two methods of evaluating impairments in television images were described.[46] In the first, observers voted a preference between two pictures with different impairments; in the second, they rated the images in terms of a seven-comment category scale, ranging from *the impairment is not perceptible* to *the picture is not usable.* (The latter method was widely used and continued in use into the 1970s and beyond.) The Bell Laboratories tests also provided estimates of the relative importance of sharpness, contrast, and brightness in influencing viewers' judgments of picture quality [Fig. 21-20].

Many of the transmission deviations resulted in a distortion or displacement of the television image similar to the effect of reflections of the broadcast signal

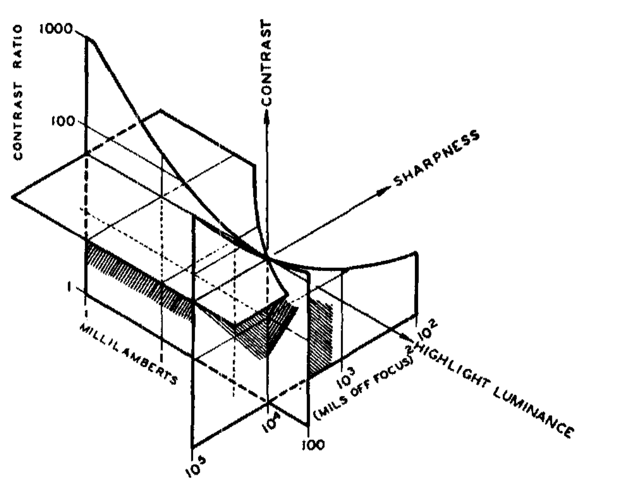


Fig. 21-20. Three-dimensional representation of television picture quality parameters, 1950. [Mertz, Fowler, and Christopher, *Proc. IRE* 38 (1950): 1280.]

known as echo. It was found from the subjective testing that the impairment from a number of different echoes, however complex, could be expressed as the equivalent of a single well-displaced echo of appropriate amplitude. In 1953, a scheme was devised for summing the individual echoes, suitably weighted, into a single equivalent echo, whose ratio to the desired signal gave an echo-rating measure of performance [Fig. 21-21]. (It is interesting to note that an impairment judged to be barely perceptible by ten percent of the observers was found to be definitely objectionable by the most critical five percent. Such contrasts were frequently noted and deplored by the development and design groups, who had to live with the requirements established, in their view, to satisfy a hypercritical fringe.) Echo rating provided a single scale by means of which the overall performance of a system could be judged. By this means, a total-performance requirement could be allocated to the different types of irregularities and trade-offs made among them as greater or lesser difficulty was experienced in meeting the initial objectives.[47]

4.1.1 Color Television

The problems of color-television transmission were also considered in the early postwar years, and the L3 coaxial and TD2 radio systems, on which development started in 1945 and 1947, respectively, were both designed with this signal in mind. Around 1953, the National Television System Committee (NTSC) adopted the system that took advantage of the absence of signal energy between the spectrum lines of the black-and-white luminance signal to transmit color information by means of the amplitude and phase of a subcarrier at 3.58 MHz. The subcarrier was made invisible on black-and-white sets by making the frequency of the color subcarrier an exact odd multiple of half the line frequency (actually, 3.579545 MHz), synchronized to the line rate. The system

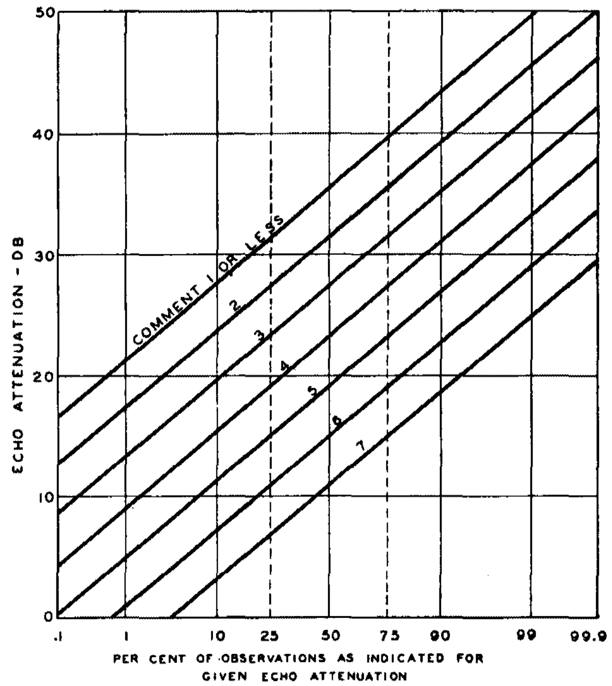


Fig. 21-21. Subjective characterization of 2-microsecond delayed echoes on five television images, 1950. Comment categories ranged from (1) not perceptible and (2) just perceptible to (6) definitely objectionable and (7) not usable. The data was for engineering use. [Mertz, Fowler, and Christopher, *Proc. IRE* 38 (1950): 1276.]

was thus compatible: color signals were received in black and white on monochrome sets and in color on color sets, and black-and-white signals were received in monochrome on both [Fig. 21-22].

The signal band required for color was not appreciably different from that for monochrome, but the tolerance for noise across the band was somewhat different, and special care had to be taken with transmission at the 3.58 MHz frequency of the color subcarrier. Requirements were derived and means for measuring and controlling the amount of differential phase and gain developed. These were nonlinear effects that disturbed the relationship of the color carrier to the other picture signal components, depending on the signal amplitude.[48,49]

While many refinements and simplifications were made, basic requirements for television transmission, including color television, were well established by the 1960s, and all later transmission systems intended to carry television signals were designed to meet them. Television networking was a powerful force in the establishment of the early microwave-radio network, but the national broadcast television network quickly grew to maturity, and the need for new facilities leveled off.

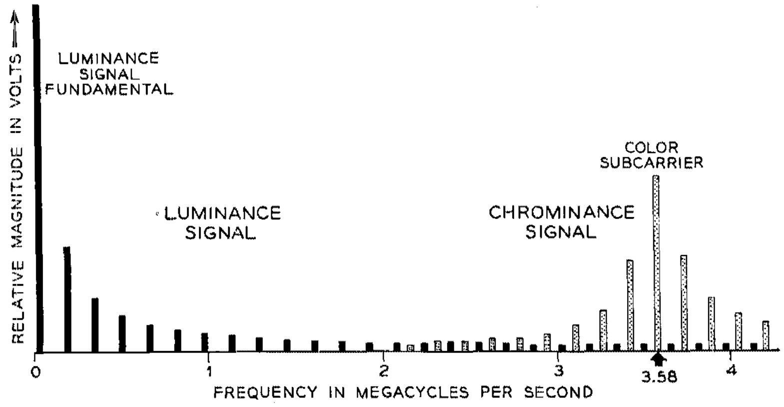


Fig. 21-22. Frequency composition of a typical NTSC color television signal. Only every tenth component is shown.

4.2 Objectives for Data Transmission

Telegraph signals were an early form of digital signal, used to transmit coded messages in binary form. Carrier telegraph systems were developed at about the same time as the first carrier systems for speech. In one form, a 4-kHz voice channel was divided into a dozen or more narrow slots of 170 Hz each. Over these, telegraph signaling rates of 100 bits per second or more were achieved using simple AM on-off keying to generate the binary signals. (Frequency-shift keying was used in some HF radio applications with lower signaling rates.) Since this voice-frequency carrier telegraph signal was limited to the voice bandwidth, it could be transmitted over almost any kind of facility that could carry a voice signal, including the multivoice-channel carrier systems. Obviously, these low-speed digital channels could be used, either singly or in combination, to transmit numerical data; but, in the early 1950s, the semiautomatic ground-environment (SAGE) system, an early-warning network for the Air Force, required transmitting data over an entire voice channel at the highest practical rate. The Bell System responded by tailoring dedicated facilities to this specific application. However, with the advent and rapid spread of computers, it could be foreseen that the transmission of such signals would become common and increasingly important in the future.

The vulnerability of data signals to various transmission impairments tended to be quite different from that of voice signals. In particular, where impulse noise (short "hits" or "bats" of a few milliseconds duration, caused mostly by relay operation in electromechanical switching offices) or envelope delay distortion (delay that is different at different frequencies) had little subjective effect on voice quality, these impairments could significantly degrade error performance on data signals. Initially, the users tended to be only large businesses or government agencies, and the special requirements of data services could be met by selecting and specially maintaining circuits for private-line data use. Also, since data signals were such a small fraction of the total traffic, higher-level signals that used a disproportionate amount of the load-carrying capacity of a line could be allowed, thus improving the signal-to-noise ratio for the data services. However, as the percentage of data traffic increased and as the transmission of data over the switched network in addition to private lines was considered, a higher power level for data became less practicable. Accordingly, in the second half of the 1960s, the average power limit for data was set at the same value as the average voice signal, that is, 16 dB below 1 mw at the zero-level reference point (−16 dBm0). This requirement was used as the basis for the signal-power limit for interconnecting customer-provided data equipment. Transmission quality objectives were thereafter set with both voice and data needs in mind.

The determination of data-transmission quality was inherently more straightforward than determining quality for voice transmission, which necessarily required the use of subjective testing. With data, in concept at least, it was merely necessary to count the errors in a given time period. However, the different impact of digital errors, different modes of dealing with them, and the large variety of data sets with different operating characteristics brought many new complications.

4.2.1 Impulse Noise—Test Sets

Early experience demonstrated the vulnerability of SAGE data to impulse noise. Since SAGE was a private-line service, the problem was met by selecting facilities with the lowest impulse noise. The selection required a capability for measuring impulse noise, and for this purpose a peak-recording device, called the impact meter, was used. Peak readings of individual impulses were recorded with a manual reset of the meter required after each reading. The readings over a one-hour period were ranked in descending order, and the power in decibels referred to 1 mw (dBm) of the 70th-highest pulse was used as a single number to characterize the impulse-noise performance of the circuit under test. Needless to say, this was a slow and cumbersome process. Later, the testing time was reduced using a half-hour period and the 35th-highest peak. Objectives for the SAGE system were set to be consistent with an overall error rate of less than one error in 10^5 bits. This requirement was subsequently adopted as a Bell System objective and similar, but more elaborate, requirements based on it were still in use in 1985.

The experience with SAGE created an interest in Bell Laboratories concerning the limits to the rate at which digital signals could be transmitted over the voice network. A number of studies in the late 1950s led to the conclusion that rates in the range of 1000 to 1600 bits per second should be possible. Modulator-demodulator (modem) terminals that converted the baseband digital signals to a form compatible with the approximately 250- to 3000-Hz telephone-channel passband were built and tested. The first modems operated at 600 and 1200 bits per second and used frequency-shift keying (FSK) because that form of modulation was expected to be less susceptible to impulse noise. FSK also provided constant power on the line, and thus held the compandors of the Type N systems at a constant operating level. Field tests with these modems in the late 1950s proved disappointing, however, and a general service offering was deferred pending further studies.

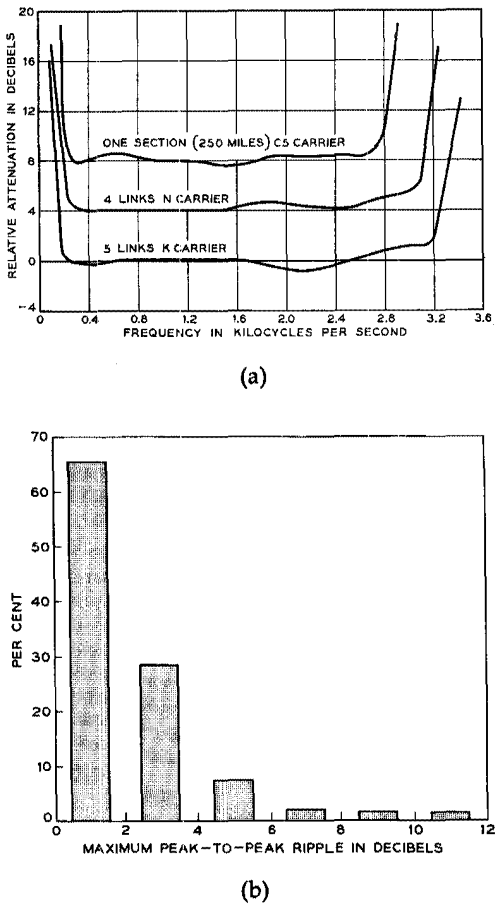
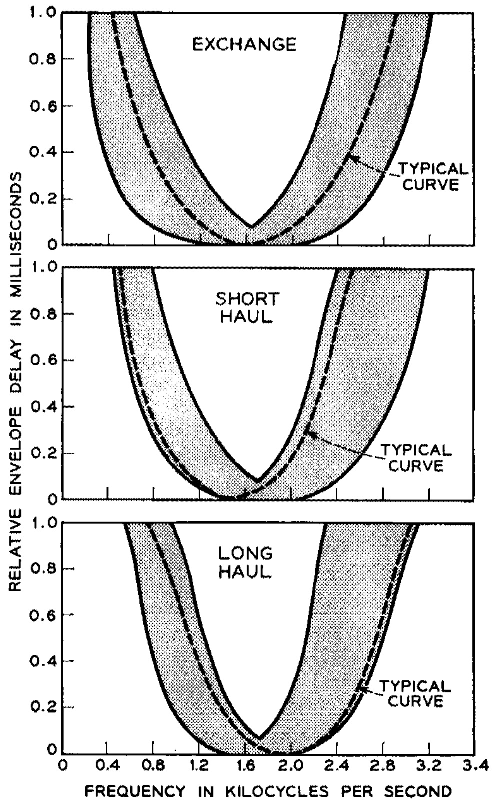


Fig. 21-23. Results of the 1959 survey to determine network data capability. (a) Representative attenuation characteristics of carrier systems. (b) Percentage of circuits with maximum peak-to-peak amplitude ripple.

In 1959, a nationwide survey of the network was carried out to characterize it statistically in terms of its ability to transmit high-speed data [Fig. 21-23].[50] An error rate of less than 10^{-5} at 1200 bits per second was observed on about 65 percent of the calls. The study led to a recommendation that certain types of older carrier systems and panel-type central offices (which were especially noisy) be avoided by the use of dedicated lines and/or special access codes. Analysis of the survey results led to the recognition that, in addition to the expected limitation imposed by the nominal voice-channel bandwidth, variations in loss (attenuation distortion) and delay (envelope delay distortion) within the band also imposed major limitations on the maximum rate feasible. As a consequence, private-line data circuits were individually equalized to reduce



(c)

Fig. 21-23. (c) Envelope delay distortion. The solid curves are the limits found for 90 percent of the circuits; the dashed curves are for typical circuits.

transmission distortion. Channels equalized to increasingly stiffer requirements were offered over a range of rates so that customers could select circuits that matched their needs. Charging different rates for channels with different limits on transmission performance was a shift from the time-honored approach of offering a single grade of service at a single price for voice-band services.

Transmission over the switched network was also improved by associating a compromise equalizer with data sets. The equalizer compensated for the average delay and attenuation distortion that might be encountered over a typical connection. Such an average approach could not, of course, achieve the same quality as the equalization that was tailored to the characteristics of a particular circuit. DATAPHONE* data communications service, as the switched service was called, was first offered around 1960. The terminal modems

* Registered service mark of AT&T.

(the 200 series data sets) operated at speeds up to 2000 bits per second on the switched network and 2400 bits per second on private line.

Continuing field surveys were essential to define the characteristics of the constantly evolving network and to help set realistic objectives for the service. But they were time-consuming and expensive, and they did not permit the parameters to be varied as a guide to future development. As a supplement to them, transmission-line simulators were also developed around 1960 in which the important impairments could be varied. These simulators were used to determine the tolerance of the various data modems and were helpful in determining where network improvements would be most beneficial [Fig. 21-24].[51,52]

Impulse-noise measurements were also improved with the introduction in 1962 of the 6A impulse counter as the standard measuring instrument for impulse noise. This test set permitted more rapid and complete characterization of impulse noise by measuring the number of impulses that exceeded various thresholds in a given period. Starting in 1962, objectives were stated in a manner consistent with this technique [Fig. 21-25]. Similarly, in 1963, studies were begun to facilitate voiceband envelope delay and attenuation distortion measurements. As a result of these studies, a new testing approach was formulated in 1964. A pulse stream with a very high peak-to-average power (10 dB) was applied to a channel, and the peak-to-average ratio was measured at the re-

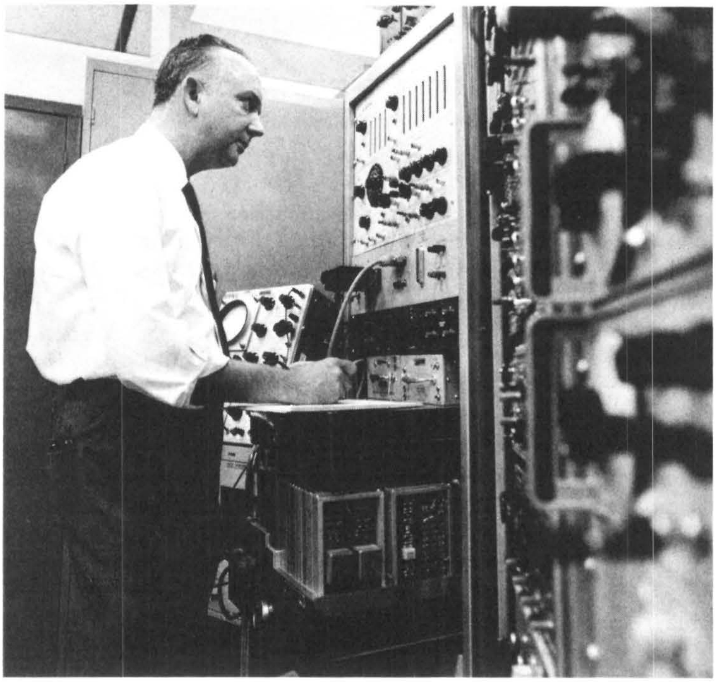


Fig. 21-24. The transmission line simulator in use to test a data set, 1971.

Fig. 21-25. The 6A impulse counter, 1962.

ceiving end. Delay and attenuation distortion reduced the "peakiness" of the signal, and the degree of reduction provided a single measure of the combined effect of the distortions, as they would be expected to affect a data signal. An improved version of a test set embodying this principle, the peak-to-average-ratio (P/AR) meter became available in 1974 [Fig. 21-26].[53]

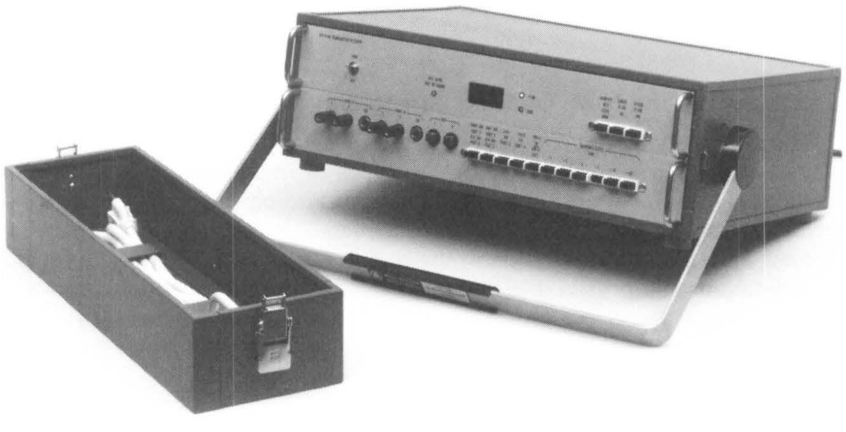


Fig. 21-26. The peak-to-average-ratio (P/AR) meter, 1974.

In the second half of the 1960s, additional data-transmission surveys were carried out using the new test capabilities. Results were incorporated in a computer program which simulated the transmission performance of data-modem terminals over modeled connections set up according to the routing rules, but which were otherwise randomly constructed. The program compiled a statistical picture of the resulting end-to-end performance. The simulation allowed the determination of the effect of changes in a single parameter or combinations of parameters. Important results of these studies were a relaxation of impulse-noise requirements and improved data transmission by a simple change in the administrative rules for laying out switched trunk networks.

In 1968, the 203-type voiceband data modem using vestigial sideband modulation and adaptive equalization was introduced with options to run at 2400, 4800, 7200, and 9600 bits per second. The two highest rates were to be used only on specially equalized private lines, but the set appeared to run satisfactorily on the switched network up to 4800 bits per second. Subsequently, it was discovered that this set was particularly sensitive to nonlinear distortion, that is, the change in the channel gain or loss as a function of signal amplitude. Nonlinear amplitude distortion within a voiceband on the network turned out to be a very complex phenomenon, and new techniques for determining its value had to be developed. Ultimately, a technique using four simultaneous tones was found to be satisfactory. By the early 1970s, a complete set of data-transmission objectives for end-to-end connections was available, but two more years of analysis and surveys were required before allocations to the various portions of the plant and the inclusion of transient as well as static characteristics were possible.[54,55]

By the mid-1970s, it was apparent to everyone that the game consisted of a constant sequence of improvements and more stringent objectives for data transmission, only to have the limits tested by more demanding data services. Digital transmission systems and special networks, such as the Digital Data Service (DDS), were clearly better suited to data transmission, but the voiceband circuits were available everywhere and were easily accessed. It seemed likely that voiceband data and its problems and progress would be a matter of interest for years to come.

4.3 Objectives and New Transmission Media

The introduction of new telephone sets, repeatered lines, and carrier systems played a major role in determining the types of impairment that had to be considered and the establishment of standards and objectives to ensure satisfactory service to voice-transmission customers. The introduction of satellite systems and digital transmission in the mid-1960s brought with them further considerations and possibilities that needed to be recognized and incorporated into planning.

4.3.1 Satellites—Delay and Echo Control

The propagation times associated with satellite circuits are much greater than those experienced on terrestrial facilities. A system utilizing a satellite in a geostationary orbit involves transmission to and from a repeater approximately

22,000 miles above the earth. An up-and-down trip takes about 300 ms and varies only slightly with the surface distance between the endpoints. A round-trip between the earth-based terminals takes twice that long, so that an echo from the far end arrives back at the talker's receiver displaced by more than half a second from the original signal. This interval is much longer than any involved in facilities along the earth's surface, and it was thus very important to determine whether the delay of a synchronous satellite system was acceptable, taking into account both the direct effect of delay on a two-way conversation and the exacerbating effect of a 600-ms delay of any echo.

When it became evident in the early 1960s that communication satellites were becoming a reality, subjective tests were made on calls by Bell Laboratories personnel by introducing additional delay into ordinary circuits, corresponding to both single and multiple hops. Results indicated that, if no echo were present, round-trip delays of up to 1200 ms went almost unnoticed. With the addition of echo and echo suppressors, however, there was some increase in customer dissatisfaction with a 600-ms, one-hop, round-trip delay and a great deal more with 1200 ms.[56]

In 1964 and 1965, tests were run in conjunction with European administrations, with similar results. In these tests, both terrestrial circuits with added delay and calls through the first synchronous satellites were used to obtain regular customer responses to various amounts of delay. (The calls were not monitored and privacy was maintained. The customers were called back and interviewed, based on the record that a call had been made.) On the basis of the results, it was concluded that only echo suppressors designed for the long delay should be used and connections should be limited to single-hop satellite links [Fig. 21-27].[57,58]

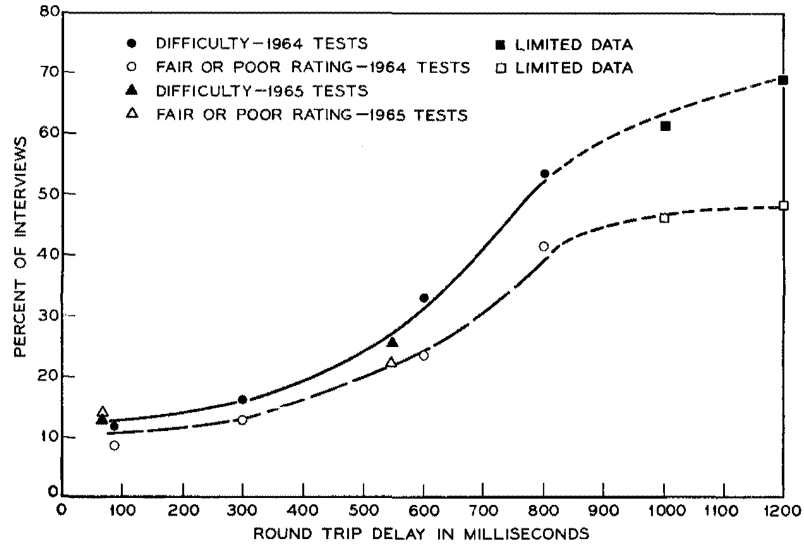


Fig. 21-27. The effect of transmission delay on the rating of transmission quality, 1964 and 1965.

These were tricky issues, since subjective tests are always susceptible to a range of interpretations, and partisans of different camps were suspected of seeing what they chose to see in the results. As always, reactions were conditioned by customer expectations. Some years later, when satellite circuits were first introduced in the domestic network, a high level of dissatisfaction was encountered (as indicated by frequent customer rejection of the circuits). Circuits that were considered perfectly acceptable on a call to London were not accepted on a call to Chicago by customers who had become accustomed to terrestrial circuits. Improved echo suppressors helped somewhat, and, as a short-term measure, satellite circuits were restricted to one direction, with a terrestrial system used for the other, thus cutting the round-trip echo delay in half.

Echoes arise in the telephone plant because the long four-wire trunks are connected to the terminating station via a local two-wire link [Fig. 21-28]. There are many possible sources of reflection in a two-wire line (which propagates in both directions), but the most important is usually the telephone receiver itself, which is a relatively poor match to the line. The balancing network in the hybrid junction is designed to match the two-wire line impedance as nearly as possible, but a close match at all voice frequencies and for all conditions is not possible in a circuit of reasonable complexity. As a result, a portion of the incoming signal is transmitted back to the talker over the return side of the four-wire line, with the disturbing effects already noted.

Echoes and echo control were subjects of interest in Bell Laboratories since the introduction of slow-propagation loaded-cable circuits in the 1920s. An early solution was an echo suppressor that opened the return path in the four-wire line when incoming speech was detected.[59,60] With the widespread use of high-velocity carrier circuits, echo suppressors were needed only on the longest trunks in use, and, for such suppresors, reasonably satisfactory designs were available. As a result, in the 1940s and 1950s, development effort on echo

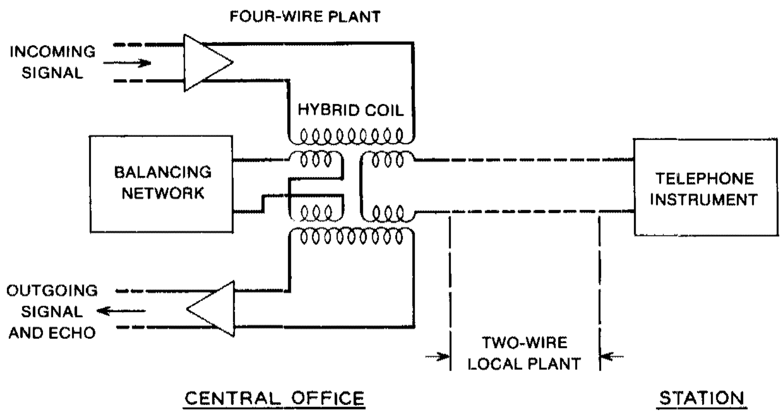


Fig. 21-28. The four-wire to two-wire junction and two-wire extension to the local station. The terminating telephone instrument was a major source of echo.

control was at a low level. In the early 1960s, the impending use of synchronous satellites and their long delay rekindled interest in better echo control.[61,62,63] Echo suppressors were effective in blocking echoes, but unfortunately they also interfered with the normal interchanges and interruptions of normal conversation. The most satisfactory long range solution appeared to be to duplicate a portion of the incoming signal and to subtract it from the reflected echo in an echo canceler [Fig. 21-29]. The cancelers had to be adaptive. In effect, they measured the echo and processed the incoming signal in amplitude and delay until the subtraction gave zero residue.

Research on adaptive echo cancellation was carried on at Bell Laboratories in the mid-1960s, after the first communications satellites were launched.[64,65] The circuits investigated were reasonably effective in canceling echoes, but they were complex. In their initial versions, they would have required bays of equipment and been far too expensive for practical use. The work was further advanced by the Communications Satellite Corporation (Comsat) and by Japanese laboratories in the later 1960s and 1970s. Work at Bell Laboratories was resumed in the early 1970s when AT&T reentered the satellite business as a partner in the domestic Comstar system, launched in 1976. The first effective echo canceler economical enough for general use was designed at Bell Laboratories in 1979 as a very-large-scale integrated circuit with 35,000 devices on a single silicon chip.[66] With echo cancelers, some problems in data transmission remained; but for voice connections the circuits with cancelers were about as well accepted as terrestrial links.

4.3.2 Digital Transmission

When the first digital transmission system was developed in the late 1950s, it was hardly necessary to consider new transmission standards or objectives. No new types of signals were being considered, and quality objectives for the existing types were already well established in the existing analog network. The task was rather to set the basic parameters for the new type of coding and modulation to meet as nearly as possible the same performance. With this in mind, several fundamental decisions were made during the development of

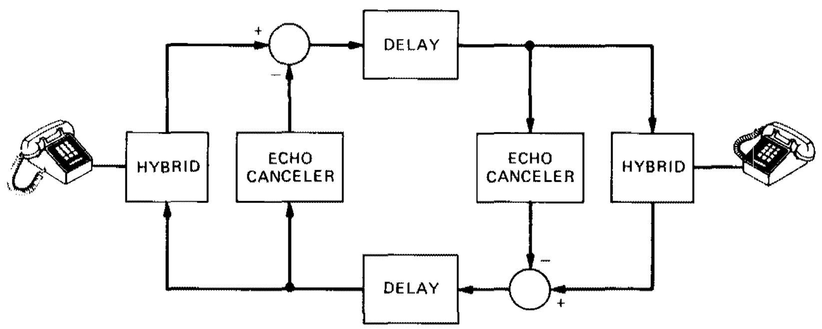


Fig. 21-29. The use of echo cancelers to control echo. A portion of the incoming signal is processed and subtracted from the reflected signal to reduce or eliminate the echo.

the T1 system that essentially established the transmission performance. These included the sampling rate, the companding law, the number of bits per sample, and the dynamic range.[67] (See Chapter 18, Sections I and II.)

Selection of an 8-kHz sampling rate, for example, was consistent with the 4-kHz spacing of earlier analog systems. (If a signal time function is sampled at regular intervals and at a rate at least twice the highest significant signal frequency, the samples contain all the information of the original signal.) Systems-engineering studies and subjective tests were conducted to arrive at the choice of the number of digits per sample and an appropriate logarithmic companding law to provide approximately uniform signal-to-quantizing-noise over a large volume range. Speech samples were coded in seven bits in the first system with an eighth bit of each code word used for signaling. The dynamic range (in effect, the maximum signal to be handled) was chosen to be just

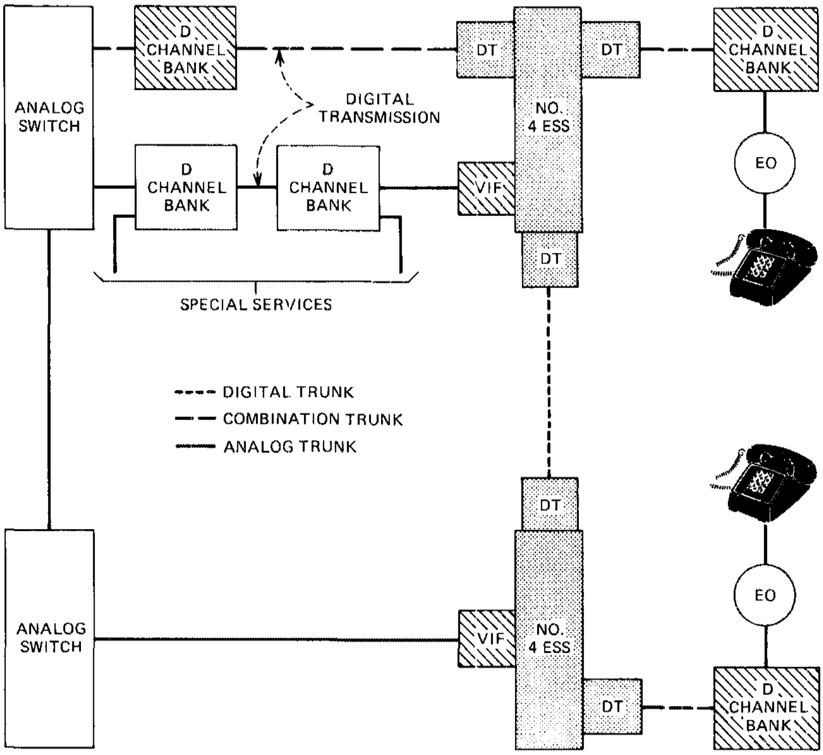


Fig. 21-30. Simplified schematic of analog, digital, and interface facilities as parts of the joint analog and digital switched network plan, early 1970s. The voice interface frame (VIF) terminates analog trunks and the digroup terminal (DT) terminates digital trunks on the switch. Open boxes are analog facilities; heavily cross-hatched items perform analog-to-digital conversions; lightly shaded items handle signals in digital format.

sufficient to transmit the peaks of a +3 dBm sine wave at the zero-transmission-level reference point without clipping. This combination of characteristics provided a signal-to-distortion ratio of approximately 30 dB for sine-wave signals ranging from −40 to +3 dBm. Idle circuit noise had a mean of less than 20 dBrnC0, with only a few channels exceeding 24 dBrnC0. Subjective tests showed that the quality of speech over a single T1 link was generally indistinguishable from that over a physical circuit with the same loss, bandwidth, and noise, over a speech volume range from −40 to −10 volume units (VUs).

T1 carrier was designed for exchange-area applications, and the performance was somewhat below what was needed for long toll links. The rapid acceptance of the T1 system by the operating companies stimulated work on the D2 channel bank, designed as a toll-grade PCM terminal. Improved performance was achieved by using the eighth bit of each word to encode the signal in five frames out of six and for signaling only in the sixth frame. Companding, which depended on the nonlinear characteristics of diodes in the earlier D1 terminal design, was improved and stabilized with a 15-linear-segment approximation to a μ=255 law. Again, 3 dBm0 was selected as the largest-amplitude sine wave that could be transmitted without clipping. The characteristics were adequate to ensure high-quality telephone performance with several encoders and decoders in tandem. In the early 1970s, a series of improvements of the same general nature in a new exchange bank, the D3 channel bank, made it usable for toll links as well. (See Chapter 19, Sections 1.2 and 1.3.)

A major step in the digitalization of the network occurred about 1970, when a decision was made to develop the toll 4ESS* switch using time division technology. The 4ESS switch was designed to switch voice channels in digital form, derived from either analog-to-digital (A/D) converters located at the switch or from T carrier transmission facilities directly, without the need for A/D conversion. The transmission objectives were essentially the same as those for the D2 and D3 channel banks.[68]

Starting about 1970, during the development of the new switch, the Network Planning Division developed a switched digital toll-network plan to guide the evolution from a network based primarily on analog technology to one based on digital technology. The evolving network could be expected to include a mixture of analog and digital facilities for many years, and the transmission performance of the interim evolving network was an important consideration [Fig. 21-30]. Objectives were derived for digital impairments, such as error rate, slips, and misframes. In addition, a loss and level plan for the interconnection of analog and digital facilities was formulated in which digital intertoll trunks between digital offices were assigned a loss of 0 dB, a departure from the via net loss plan for the analog network. In a toll connection that used digital facilities exclusively, the plan provided a fixed loss of 6 dB from end office to end office. The idle circuit noise was to be independent of length and would be the equivalent of a single encoder and decoder, typically less than 20 dBrnC at an end office.[69,70] Enthusiasm for the new plan was very high. With the

---

* Trademark of AT&T Technologies, Inc.

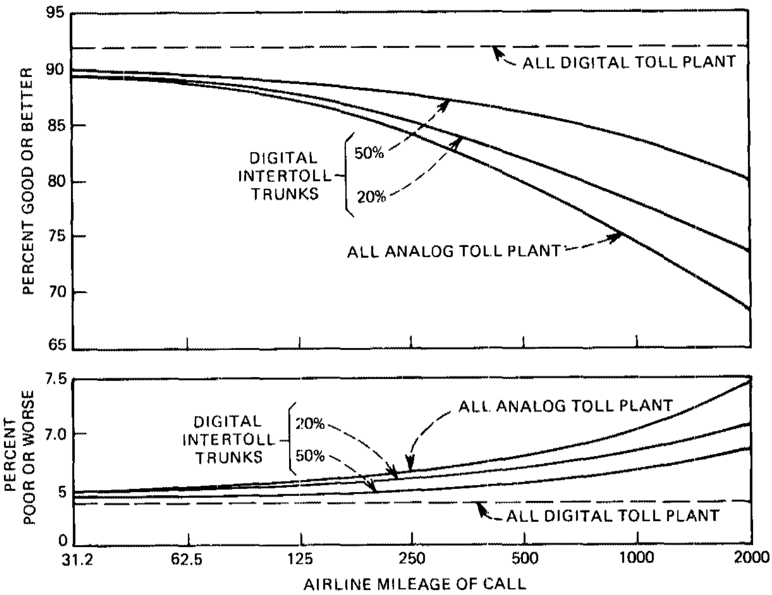


Fig. 21-31. Projected noise-loss grade-of-service of switched network with evolving digital facilities as a function of connection distance, 1977.

excellent quality of the improved digital terminals and the reduction or elimination of tandem analog-to-digital and digital-to-analog conversions, transmission quality would improve and become independent of distance as the network became increasingly digital [Fig. 21-31].

REFERENCES

1. L. J. Sivian, "A Telephone Transmission Reference System," *Elec. Commun.* **3** (October 1924), pp. 114-126.
2. R. V. L. Hartley, "The Transmission Unit," *Elec. Commun.* **3** (July 1924), pp. 34-42.
3. W. H. Martin, "The Transmission Unit and Telephone Transmission Reference Systems," *Bell Syst. Tech. J.* **3** (July 1924), pp. 400-408.
4. W. H. Martin and C. H. G. Gray, "Master Reference System for Telephone Transmission," *Bell Syst. Tech. J.* **8** (July 1929), pp. 536-559.
5. W. H. Martin, "Decibel—The Name for the Transmission Unit," *Bell Syst. Tech. J.* **8** (January 1929), pp. 1-2.
6. W. H. Martin, "Rating the Transmission Performance of Telephone Circuits," *Bell Syst. Tech. J.* **10** (January 1931), pp. 116-131.
7. F. W. McKown and J. W. Emling, "A System of Effective Transmission Data for Rating Telephone Circuits," *Bell Syst. Tech. J.* **12** (July 1933), pp. 331-346.
8. T. G. Castner, E. Dietze, G. T. Stanton, and R. S. Tucker, "Indicating Meter for Measurement and Analysis of Noise," *Trans. Amer. Inst. Elec. Eng.* **50** (1931), pp. 1041-1047.
9. R. S. Tucker, "Results of Noise Surveys: Part II, Noise in Buildings," *J. Acoust. Soc. Amer.* **2** (July 1930), pp. 59-64.

10. W. J. Williams and R. G. McCurdy, "A Survey of Room Noise in Telephone Locations," *Bell Syst. Tech. J.* **9** (October 1930), pp. 652-667.
11. H. S. Osborne, "A General Switching Plan for Telephone Toll Service," *Bell Syst. Tech. J.* **9** (July 1930), pp. 429-447.
12. A. B. Clark and H. S. Osborne, "Long Distance Telephone Circuits in Cable," *Bell Syst. Tech. J.* **11** (October 1932), pp. 520-545.
13. D. F. Seacord, "Room Noise at Telephone Locations," *Elec. Eng.* **58** (June 1939), pp. 255-257.
14. D. F. Seacord, "Room Noise at Telephone Locations—II," *Elec. Eng.* **59** (June 1940), pp. 232-234.
15. H. A. Affel, "Transmitted Frequency Range for Circuits in Broad-Band Systems," *Bell Syst. Tech. J.* **16** (October 1937), pp. 487-492.
16. A. H. Inglis, "Transmission Features of the New Telephone Sets," *Bell Syst. Tech. J.* **17** (July 1938), pp. 358-380.
17. N. R. French and J. C. Steinberg, "Factors Governing the Intelligibility of Speech Sounds," *J. Acoust. Soc. Amer.* **19** (January 1947), pp. 90-119.
18. H. Fletcher and R. H. Galt, "The Perception of Speech and Its Relation to Telephony," *J. Acoust. Soc. Amer.* **22** (March 1950), pp. 89-151.
19. H. Fletcher, *Speech and Hearing in Communication* (Princeton, NJ: D. Van Nostrand Co., 1953).
20. A. H. Inglis and W. L. Tuffnell, "An Improved Telephone Set," *Bell Syst. Tech. J.* **30** (April 1951), pp. 239-270.
21. L. B. Bogan and K. D. Young, "Simplified Transmission Engineering in Exchange Cable Plant Design," *Trans. Amer. Inst. Elec. Eng.—Pt. I. Commun. Electron.* **73** (November 1954), pp. 498-502.
22. A. B. Clark and H. S. Osborne, "Automatic Switching for Nation-wide Telephone Service," *Trans. Amer. Inst. Elec. Eng.* **71** (September 1952), pp. 245-269.
23. H. R. Huntley, "Transmission Design of Intertoll Telephone Trunks," *Bell Syst. Tech. J.* **32** (September 1953), pp. 1019-1036.
24. H. H. Felder and E. N. Little, "Intertoll Trunk Net Loss Maintenance under Operator Distance and Direct Distance Dialing," *Bell Syst. Tech. J.* **35** (July 1956), pp. 955-972.
25. O. H. Coolidge and G. C. Reier, "An Appraisal of Received Telephone Speech Volume," *Bell Syst. Tech. J.* **38** (May 1959), pp. 877-897.
26. V. Subrizi, "A Speech Volume Survey on Telephone Message Circuits," *Bell Lab. Rec.* **31** (August 1953), pp. 292-295.
27. A. J. Aikens and D. A. Lewinski, "Evaluation of Message Circuit Noise," *Bell Syst. Tech. J.* **39** (July 1960), pp. 879-909.
28. See reference 19.
29. W. T. Cochran and D. A. Lewinski, "A New Measuring Set for Message Circuit Noise," *Bell Syst. Tech. J.* **39** (July 1960), pp. 911-931.
30. I. Nåsell, "The 1962 Survey of Noise and Loss on Toll Connections," *Bell Syst. Tech. J.* **43** (March 1964), pp. 697-718.
31. D. A. Lewinski, "A New Objective for Message Circuit Noise," *Bell Syst. Tech. J.* **43** (March 1964), pp. 719-740.
32. Bell Telephone Laboratories Members of Technical Staff, *Transmission Systems for Communications*, 3rd Ed. (1964), and 4th Ed. (1970).
33. J. R. Cavanaugh, R. W. Hatch, and J. L. Sullivan, "Models for the Subjective Effects of Loss, Noise, and Talker Echo on Telephone Connections," *Bell Syst. Tech. J.* **55** (November 1976), pp. 1319-1371.
34. T. C. Spang, "Loss-Noise-Echo Study of the Direct Distance Dialing Network," *Bell Syst. Tech. J.* **55** (January 1976), pp. 1-36.
35. W. B. Macurdy and A. E. Ritchie, "The Network: Forging Nationwide Telephone Links," *Bell Lab. Rec.* **53** (January 1975), pp. 4-15.

36. H. E. Ives, "Television," *Bell Syst. Tech. J.* **6** (October 1927), pp. 551-559.
37. P. Mertz and F. Gray, "A Theory of Scanning and Its Relation to the Characteristics of the Transmitted Signal in Telephotography and Television," *Bell Syst. Tech. J.* **13** (July 1934), pp. 464-515.
38. M. E. Strieby, "Coaxial Cable System for Television Transmission," *Bell Syst. Tech. J.* **17** (July 1938), pp. 438-457.
39. G. N. Thayer, A. A. Roetken, R. W. Friis, and A. L. Durkee, "A Broad-Band Microwave Relay System Between New York and Boston," *Proc. IRE* **37** (February 1949), pp. 183-188.
40. M. W. Baldwin, Jr., "The Subjective Sharpness of Simulated Television Images," *Bell Syst. Tech. J.* **19** (October 1940), pp. 563-586.
41. M. W. Baldwin, Jr., "Subjective Sharpness of Additive Color Pictures," *Proc. IRE* **39** (October 1951), pp. 1173-1176.
42. A. D. Fowler, "Observer Reaction to Low-Frequency Interference in Television Pictures," *Proc. IRE* **39** (October 1951), pp. 1332-1336.
43. A. D. Fowler, "Observer Reaction to Video Crosstalk," *J. Soc. Motion Pict. Telev. Eng.* **57** (November 1951), pp. 416-424.
44. P. Mertz, "Perception of Television Random Noise," *J. Soc. Motion Pict. Telev. Eng.* **54** (January 1950), pp. 8-34.
45. P. Mertz, "Influence of Echoes on Television Transmission," *J. Soc. Motion Pict. Telev. Eng.* **60** (May 1953), pp. 572-596.
46. P. Mertz, A. D. Fowler, and H. N. Christopher, "Quality Rating of Television Images," *Proc. IRE* **38** (November 1950), pp. 1269-1283.
47. Bell Telephone Laboratories Members of Technical Staff, *Transmission Systems for Communications*, 4th Ed. (Bell Telephone Laboratories, 1970), pp. 698-701.
48. A. M. Lessman, "Subjective Effects of Delay Difference Between Luminance and Chrominance Information of the NTSC Color Television Signal," *J. Soc. Motion Pict. Telev. Eng.* **80** (August 1971), pp. 620-624.
49. J. R. Cavanaugh and A. M. Lessman, "Subjective Effects of Differential Gain and Differential Phase Distortions in NTSC Color Television Pictures," *J. SMPTE* **80** (August 1971), pp. 614-619.
50. A. A. Alexander, R. M. Gryb, and D. W. Nast, "Capabilities of the Telephone Network for Data Transmission," *Bell Syst. Tech. J.* **39** (May 1960), pp. 431-476.
51. J. H. Fennick and I. E. O. Nasell, "The 1963 Survey of Impulse Noise on Bell System Carrier Facilities," *IEEE Trans. Commun. Technol.* **COM-14** (May 1966), pp. 520-525.
52. J. H. Fennick, "Transmission Objectives For Voiceband Data," *Bell Lab. Rec.* **50** (January 1972), pp. 18-21.
53. T. C. Anderson, "The PAR Meter," *IEEE International Convention Record*, Part 5 (1964), pp. 155-165.
54. M. D. Balkovic, H. W. Klancer, S. W. Klare, and W. G. McGruther, "High-Speed Voiceband Data Transmission Performance on the Switched Telecommunications Network," *Bell Syst. Tech. J.* **50** (April 1971), pp. 1349-1384.
55. H. C. Fleming and R. M. Hutchinson, Jr., "Low-Speed Data Transmission Performance on the Switched Telecommunications Network," *Bell Syst. Tech. J.* **50** (April 1971), pp. 1385-1405.
56. R. R. Riesz and E. T. Klemmer, "Subjective Evaluation of Delay and Echo Suppressors in Telephone Communications," *Bell Syst. Tech. J.* **42** (November 1963), pp. 2919-2941.
57. G. K. Helder, "Customer Evaluation of Telephone Circuits with Delay," *Bell Syst. Tech. J.* **45** (September 1966), pp. 1157-1191.
58. E. T. Klemmer, "Subjective Evaluation of Transmission Delay in Telephone Connections," *Bell Syst. Tech. J.* **46** (July/August 1967), pp. 1141-1147.
59. A. B. Clark and R. C. Mathes, "Echo Suppressors for Long Telephone Circuits," *J. Amer. Inst. Elec. Eng.* **44** (June 1925), pp. 618-626.

60. O. B. Blackwell, "Time Factor in Telephone Transmission," *Trans. Amer. Inst. Elec. Eng.* **51** (1932), pp. 141-147.
61. See reference 56.
62. J. W. Emling and D. Mitchell, "The Effects of Time Delay and Echoes on Telephone Conversations," *Bell Syst. Tech. J.* **52** (November 1963), pp. 2869-2891.
63. P. T. Brady and G. K. Helder, "Echo Suppressor Design in Telephone Communications," *Bell Syst. Tech. J.* **42** (November 1963), pp. 2893-2917.
64. M. M. Sondhi, "An Adaptive Echo Canceller," *Bell Syst. Tech. J.* **46** (March 1967), pp. 497-511.
65. J. L. Kelly, Jr. and B. F. Logan, Jr., U.S. Patent No. 3,500,000; filed October 31, 1966; issued March 10, 1970.
66. D. L. Duttweiler and Y. S. Chen, "A Single-Chip VLSI Echo Canceler," *Bell Syst. Tech. J.* **59** (February 1980), pp. 149-160.
67. K. E. Fultz and D. B. Pennick, "The T1 Carrier System," *Bell Syst. Tech. J.* **44** (September 1965), pp. 1405-1451.
68. H. E. Vaughan, "An Introduction to No. 4 ESS," *International Switching Symposium Record,* Cambridge, MA (June 6-9, 1972), pp. 19-25.
69. F. T. Andrews and R. W. Hatch, "National Telephone Network Transmission Planning in the American Telephone and Telegraph Company," *IEEE Trans. Commun. Technol.* **COM-19** (June 1971), pp. 302-314.
70. J. E. Abate, L. H. Brandenburg, J. C. Lawson, and W. L. Ross, "The Switched Digital Network Plan," *Bell Syst. Tech. J.* **56** (September 1977), pp. 1297-1320.

Chapter 22

Services

I. INTRODUCTION

Up to this point, this history has been concerned with the evolution of the transmission network primarily in terms of the systems from which the nationwide network of facilities was constructed. It seems advisable at this point to consider briefly some of the uses to which the network was put, that is, the actual communication services provided to end customers. The purpose is not a comprehensive review of services over the years but only an indication of where the needs of a service had significant influence on the technology.

II. VOICE SERVICES

The Bell System plant carried many other kinds of signals besides voice, but the provision of voice-communication services to the general public remained by far the most important part of the telephone business. Voice services are categorized in several different ways. They include both calls within a local service area and long distance, or toll, calls between such areas. A distinction is also made between residential and business services, mainly for purposes of billing, although some business services place special requirements on transmission. One of the most useful classifications for an account of the technology is a grouping into ordinary and special services. Ordinary services include residence, public-telephone (coin telephones), mobile, and basic individual-line business services. All other services are considered special services. Special services require special treatment with respect to transmission, signaling, switching, or billing and are used mostly by business customers. As of 1985, the overall demand for special services was growing about twice as fast as the demand for ordinary telephone service.[1]

2.1 Ordinary Services

2.1.1 Message Telephone Service

Message telephone service is also known more familiarly as plain old telephone service (POTS). It is what most commonly comes to mind when telephone service is mentioned, and accounted for about 90 percent of Bell System revenue in 1975. Its task is to interconnect on demand any two customers within reach of a telephone. The growth of message telephone service in the United States over the years from 1925 to 1975 was spectacular [Fig. 22-1].

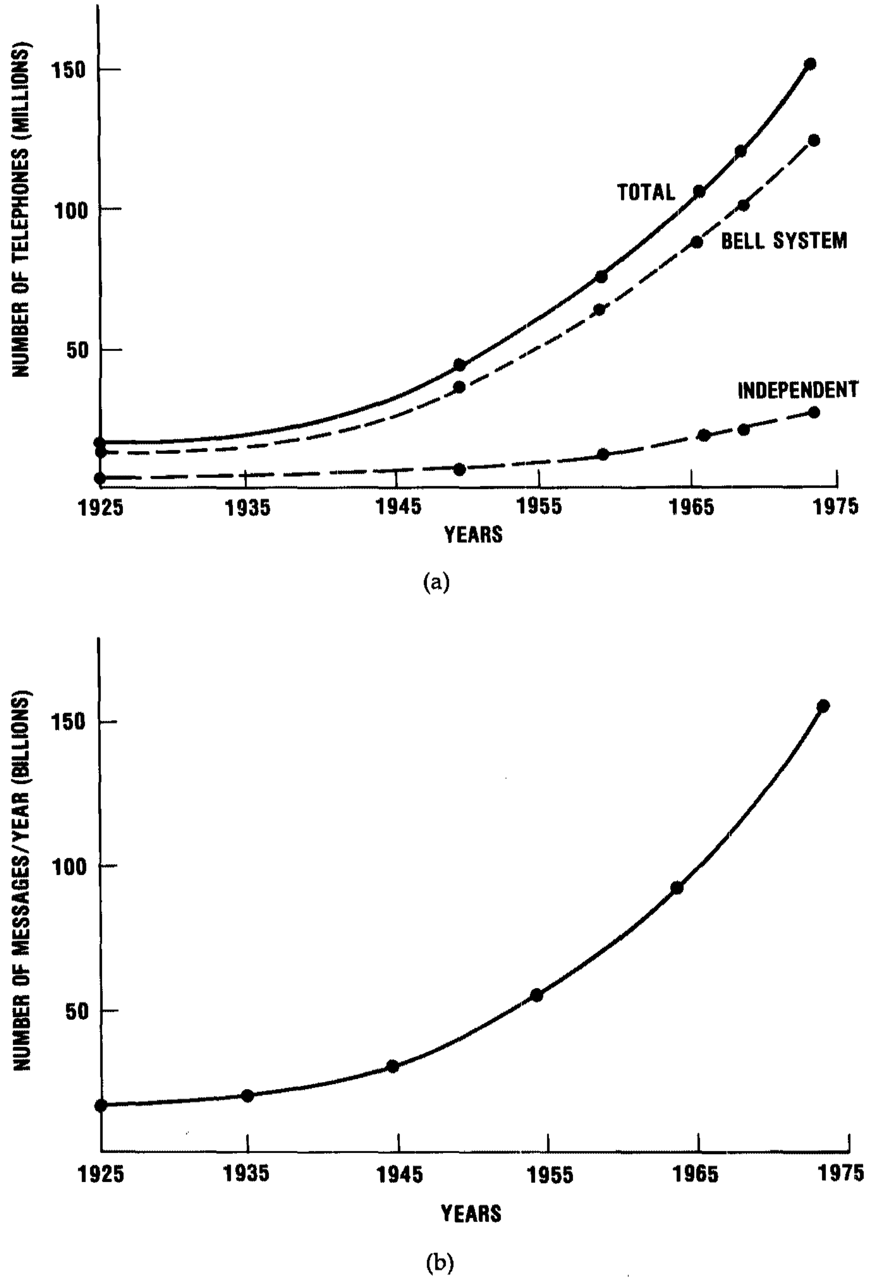


Fig. 22-1. Growth of telephone service in the United States, 1925-1975. (a) Growth in the number of telephones. (b) Growth in messages per year.

The evolution of message telephone service has been characterized not only by dramatic growth but by improved quality and increased convenience. This evolution was most directly visible to the customer in the telephone set used to make or receive calls. Behind the successively more modern appearance of each design lay solid improvements in response and efficiency [Fig. 22-2]. The 500-set design of 1950 was also strongly influenced by the consideration of human factors; it was easier to use and 25 percent lighter than previous sets. Many mechanical improvements were also made in what proved to be an effective, inexpensive, and extraordinarily long-lived design. It set a pattern imitated around the world [Fig. 22-3].

2.1.2 Dialing and Signaling

In telephone parlance, signaling refers to the system of control signals necessary to establish, monitor the status of (supervise), and take down a connection

Fig. 22-2. The evolution of the telephone station set. *Top*, the candlestick deskstand of 1919; *left center*, the handset of 1927; *right center*, the 300 type combined set of 1937 (the ringer was included in the base); and, *bottom*, the 500 type set of 1950.

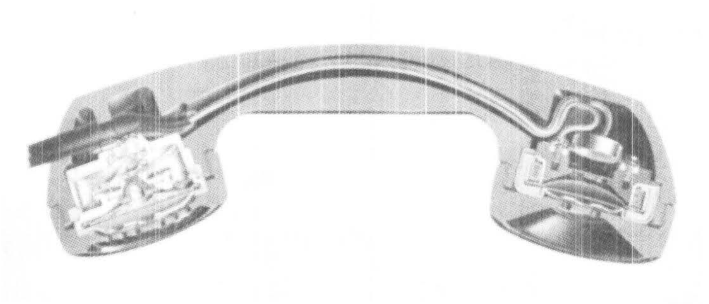


Fig. 22-3. Cross section of the 500 type handset, 1950.

through the network switches. The means of signaling evolved along with the transmission methods. On the earliest toll lines, DC signaling, essentially identical to local methods, was used. Around 1902, the desire to use the DC path for telegraphy caused a change to 135-Hz AC signaling, above the telegraph but still below the normal telephone-frequency range. The introduction of carrier telephony led to the invention of so-called in-band systems, and several patents were issued in 1921. In 1922, a 1000-Hz tone interrupted 20 times a second was used. A vacuum-tube detector was used for sensitivity with the output operating a relay. The relay was mechanically tuned to the 20-Hz interruption rate to guard against voice interference into the signaling circuit by speech signals (talk off), and the response was made slow so that only a continuing signal would cause operation. This system was used on some long circuits in 1923 and on the southern transcontinental line in 1924. (See another volume in this series, *The Early Years (1875–1925)*, p. 635.)

The widespread deployment of machine switching central offices in the 1920s and 1930s heralded the end of the era of operator-established local calls. With their dials, customers could now control their own connections with a consequent reduction in the time required to set up calls. For such service to be effective, it was necessary to provide trunk groups between the offices large enough so that circuits were available essentially on demand, even during the busy hour. Dial service spread rapidly in urban areas before World War II. For one thing, it was not economically feasible to invest much operator time on flat-rate calls, and people making large numbers of local calls were not tolerant of delays. In 1925, 1.5 million stations or 12 percent of the total were dial equipped; by the late 1930s 50 percent were dial phones. These were heavily concentrated in the urban areas.

2.1.3 Direct Distance Dialing

Until the 1950s, all toll calls were still placed with operators controlling the actual connections. In the 1920s and early 1930s, customers were still queued for calls over long toll connections, but waiting times were steadily reduced [Fig. 22-4]. By the late 1930s, most connections were completed while the caller

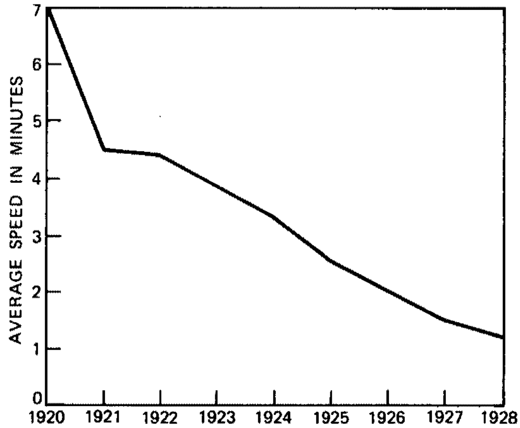


Fig. 22-4. Improvement in average time required to establish a toll connection.

held the line. The rapidity with which long calls were established on request was the subject of a popular Bell exhibit at the 1939–1940 New York World's Fair. Limited toll dialing by operators was possible by 1943,[2] and, after the hiatus of World War II, work was resumed in earnest on automatic toll switching. Nationwide dialing by operators began in the early 1950s, and, in 1951, toll dialing by customers was started. Direct distance dialing (DDD) spread rapidly following its introduction and by 1960 was available to more than 54 percent of the Bell System. Push-button keyed, multifrequency dialing (Touch-tone) was introduced in 1963, making voice-frequency in-band signaling a feature of the loop as well as the trunk plant. By the late 1960s, the Bell System was essentially completely on an automatic-dial basis. Overseas dialing by operators started in 1963 and by customers in 1970.

The principal impact of these developments on transmission was a host of new signals on the network and much more stringent requirements on the uniformity and continuity of transmission between switches [Fig. 22-5].[3,4] Supervision on carrier trunks was originally provided by a single-frequency tone at 1600 Hz. The frequency was chosen to be as high as possible but still to fit within the band of the narrow split-channel emergency-band banks introduced during and immediately after the war. Discrimination against voice interference into the signaling circuit was easier at higher frequencies, however, and, when full-band channels were again universally available, the single-frequency tone was moved to 2600 Hz. The original N carrier design provided a separate out-of-band signaling slot at 3700 Hz for complete protection against talk off, but the loss of flexibility in patching circuits led to a strong preference for an in-band arrangement. Supervisory signals and dial pulses were transmitted in either case by keying the tone on and off in synchronism with the open or closed status of the loop. In some cases, in addition to single-frequency supervision, multifrequency address signals were generated by a selection of two out of five frequencies in the range from 700 to 1700 Hz.

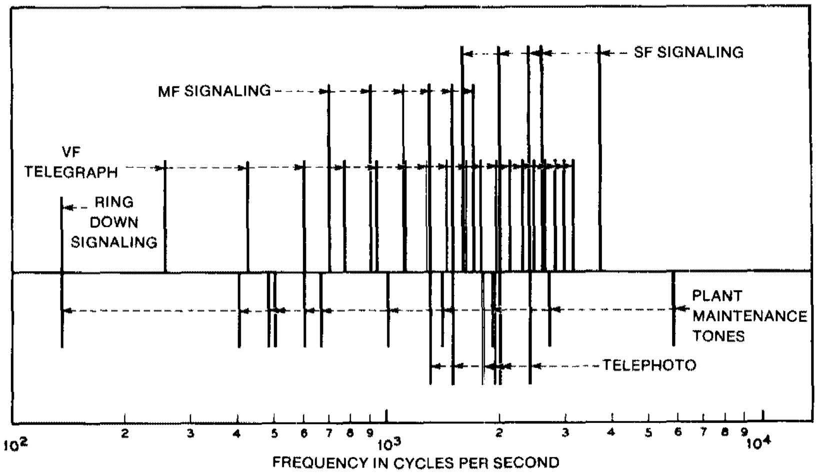


Fig. 22-5. Tones for signaling, telegraph, telephoto, and plant maintenance, 1960.

After 1950 the analog systems were designed to much tighter trunk-loss-variation limits than had formerly been necessary in order to transmit these tones within permissible loss ranges. In addition, tones were a potential source of intelligible crosstalk, which had to be made completely inaudible. The 2600-Hz signaling tones were especially troublesome because the signaling developers wanted them on at all times in an idle channel. If the tone was absent when a channel was idle, there would be no way of detecting a failure until the tone was needed and failed to appear. Tone-generated crosstalk limits often set the linearity requirements in analog systems.

Signaling equipment that was sensitive enough to detect fairly low-level tones, capable of distinguishing between the correct tones and speech, and able to translate from the line tones to the indications necessary to control a switch was expensive. The cost of signaling equipment was a significant part of the cost of carrier terminals and was an obstacle in the effort to use carrier in the exchange area, where terminal costs dominated the economics of application. As already noted, the inexpensive method of signaling in T carrier was a major key to its success.

Removing the control signaling completely from the message channels was an old idea and had been tried in the early days of automatic switching. However, in-band, per-trunk signaling was the dominant method for many years. In the 1960s, the advent of electronic switching offices with computer-like common controls and the availability of high-speed data channels renewed interest in the idea. There were many difficulties in the way; but, in May 1976, the first common-channel-interoffice-signaling (CCIS) link, which used a dedicated data circuit between the electronic switches and completely removed the control signals from the message circuits, was installed between Madison, Wisconsin and Chicago, Illinois. This marked the first step in a process that seemed likely to bring about a complete separation between the transmission of customer messages and control signals.[5,6]

By the late 1970s, CCIS and increasingly versatile, software-controlled switching offices were making a host of new voice services possible, such as conferencing, call storage, and call forwarding. These services seemed likely to be of great interest to telephone customers, but their implementation became entangled in the many regulatory and political issues preceding the AT&T divestiture of the local telephone companies. In general, they appeared to present no special problems to transmission technology.

2.2 Special Voice Services

Special service lines include central office to PBX trunks, foreign-exchange lines, private lines, private networks, and a number of other services. In 1975, there were about 25 major categories of special services, only a portion of which placed any special requirements on transmission. Where special transmission performance was needed, it often did not require any advance in technology but frequently required nonstandard packaging or arrangements. These special arrangements created many problems in administration and maintenance. An example is a foreign-exchange line, where a loop and station in a distant office are connected so as to appear as a subscriber in the nearby local office. The interoffice connection may be many miles long and require much of the electronics associated with a long trunk, but it must appear on the loop side of the local switch and operate with standard customer-loop-type signaling.

In the late 1960s and through the 1970s, methods for applying advanced technology, especially computers and software, to reduce the cost and improve the efficiency in the operation and maintenance of the huge existing plant became a major concern at Bell Laboratories. The design of equipment arrangements and methods for providing and maintaining special services were part of this effort. By the 1970s, specially designed terminals for voice-frequency pairs (metallic facility terminals), analog carrier circuits (analog facility terminals), and digital carrier circuits (digital facility terminals) made it possible to provide a number of frequently needed arrangements (e.g., for signaling, gain, and equalization) by means of single plug-in circuit cards. The terminating cards could often serve a range of conditions by means of switches that could be set by prescription without circuit testing. In the 1980s, the D5 digital channel bank consolidated many of the options and accomplished the same purpose by means of built-in microprocessors and software.

2.2.1 Private Lines and Private Networks

A private line is a leased line dedicated to a particular customer's exclusive use. Unlike the public message service where each call is set up by the network switches, a private line usually connects two fixed endpoints on a more or less permanent basis. Where a customer has a multiplicity of locations with a great deal of calling between them, a private network with PBXs at major locations and dedicated interlocation switching capability may be established. A customer is motivated to lease a private line or network because, for heavy usage between specific points, the costs are often less than the use of the public message network. Furthermore, there is no competition for service with other customers, control can be exerted over the dedicated network, and it can be tailored and reconfigured to the client's specific needs. It should be noted, however, that, while private lines or networks consist of communication channels dedicated to the leasing customer, these channels may use hardware, portions of which are used by other private-line customers or even by the public switched message service. Over the years, many private lines and networks were established and, in the post–World War II era, they became a major factor in the Bell System network. They ranged in size from single lines between two points to nationwide networks serving hundreds of locations.

There are special considerations that must be taken into account in engineering and maintaining facilities carrying special services. In the case of the switched public service, the grade of service is a statistical result of interconnecting a number of facilities in a variety of connections that are likely to be different even for successive calls to the same location. In the case of a private line, the connection is more or less permanently "nailed up." There are advantages and disadvantages to this arrangement. The line can be engineered with foreknowledge of the performance of each of the components making up the connection. It is not necessary to take into account the wider range of characteristics resulting from a variety of interconnections. It is also possible to provide special equalization, for example, to make higher data-transmission rates possible. On the other hand, the customer will always encounter the same facilities every time a call is placed to a particular destination. For an ordinary voice line, the performance must be significantly better than the lowest quality acceptable in the distribution that characterizes the public network.

Furthermore, unlike the public-network user, a special-services customer is not automatically protected from the effect of a failure in the dedicated facilities by multiple paths and alternate routing. Nor can the nailed-up lines be tested by the automatic test equipment that routinely checks switched lines via the office switches. Extra maintenance precautions and/or special protection arrangements are therefore necessary to provide adequate special-services reliability.

Starting about 1970, a method was devised for dealing with the maintenance and reliability demands of special services by the use of consolidated equipment arrangements. These combined channel banks, pads, amplifiers, equalizers, and signaling in a single bay. The arrangements simplified monitoring of the condition of circuits and eliminated multipoint testing within an office by locating all the required equipment together in one place. In the late 1970s, sophisticated and highly automated testing was made available to many special-service circuits by providing switched-access remote testing equipment as part of the consolidated arrangements.[7]

Some of the largest and most complex private networks were designed and assembled for government agencies and the military. Some of these were worldwide networks that used facilities from a number of foreign administrations as well as those of the Bell System and independent companies in the United States. In the early 1960s, the Department of Defense undertook to consolidate a number of facilities into an automatic voice network, called AUTOVON, to permit calls between military bases anywhere in the world. A unique feature of this network was that four-wire transmission was provided all the way to the station set, making echo control much easier and permitting many special transmission features. Four of the earliest electronic switches, specially modified for four-wire transmission, were dedicated to this service.[8]

III. NONVOICE SERVICES

Although the construction of the Bell System transmission network was motivated by the needs of the voice-transmission business, once in existence, it was found to be useful for many other types of signals. These uses ranged from telegraph and program-quality audio to television and, in more recent times, computer-to-computer data transmission.

3.1 Program Audio

The use of Bell System program audio channels to distribute radio-program material between a station and its affiliates began officially when the Long Lines Department established rates for program networks in 1926. Program transmission quickly became a well-established service, and the Bell System continued to offer the service to commercial radio networks and wired music systems into the 1980s. When television networking became established in the early post–World War II years, the accompanying audio was treated as a program signal and handled over a sound channel completely separate from the video signal.

Program signals include speech and a wide range of musical material. The Bell System offered facilities with a range of bandwidth and price to broadcasters. The first program-transmission networks had a band of frequencies from about 100 to 5000 Hz, compared with the nominal 200- to 3500-Hz band of a standard voice channel. Later, program services with bands of 50 to 8000 Hz and 50 to 15,000 Hz were offered, but these were rarely used. Besides bandwidth, program signals differed from ordinary telephone speech in having a wider range of transmitted-sound power, i.e., greater dynamic range. Less distortion was allowed and better signal-to-noise ratio was required. In the early audio frequency systems, the higher quality was achieved by using heavier-gauge conductors and more powerful amplifiers.[9] In the carrier systems, the signal was permitted to use more than a proportional share of the line's signal-power capacity and to occupy two or more voice channels.

Some program signals were more susceptible to delay distortion than telephone speech and required better equalization. Operating characteristics differed from ordinary speech lines as well. Operation was usually one way, holding times were very long, and extreme diligence in maintaining continuity and restoring service promptly in the event of a failure was required.

3.2 Telegraph

The rapid growth of telephony quickly led to a widespread network that could be, and was, used for telegraph transmission. Initially, this led to fierce competition between the Bell System companies and the companies that had previously built specifically telegraph-transmission systems. By 1900, most telegraph business was handled by the Western Union Telegraph Co., Postal Telegraph, Inc., and the Bell System companies. Later, as the telephone business continued to grow, AT&T acquired a substantial interest in Western Union and gained control of the company in 1909. The affiliation continued until 1913 antitrust rulings required a separation of the two companies. (See an earlier volume in this series, *The Early Years (1925–1975)*, Chapter 2.) The Bell System continued to transmit telegraph as a carrier for the telegraph companies, who served the end customer. AT&T also provided telegraph service over private lines with printers (later teletypewriters) so that specially skilled operators were not required. Teletypewriter switched exchange service was started in 1931 and placed on a customer-controlled dial-up basis in 1962 [Fig. 22-6].[10] The public switched service was sold to Western Union in January 1969, but private-line teletypewriter service continued through the 1970s.

On voice-frequency lines, the "DC" telegraph signals (actually low frequencies, with a band extending to less than 100 Hz) were separated from voice signals on the same pairs by simple high-pass/low-pass networks in so-called simplex or composite arrangements. Later a voice band was divided by filters into narrow slots of 170 Hz or so. Within each slot, a telegraph signal was transmitted by keying a carrier on and off.[11] Originally, 12 of these signals (later as many as 17) could be transmitted in a voice band. By confining the signal frequencies to the voice band and controlling signal amplitudes, the entire package could be treated the same as a voice signal and transmitted over any facility, including carrier systems, capable of voice transmission. Speeds up to 75 bits per second were generally available for many years. Signaling speeds of 150 bits per second became available starting about 1965. Throughout the 50 years from 1925 to 1975 telegraph was a distinct service that provided

Fig. 22-6. Typical teletypewriter subscriber station, 1936.

a printed record of the message and did not require that the communicating parties be on the line simultaneously.[12]

3.3 Television

There were several prewar demonstrations of television over coaxial cable, but the first transmission of the standard 525-line signal for broadcast was on December 1, 1945. On that date, the Army-Navy football game, played in Philadelphia, was transmitted over the coaxial cable for broadcast by the National Broadcasting Company in New York. Although the audience in the stadium probably outnumbered the television viewers, it was a harbinger of things to come.

Several other demonstration transmissions followed. AT&T made coaxial lines available for experimental purposes to any qualified broadcaster, and several made use of the lines. When the TDX experimental microwave-radio line was opened from New York to Boston in 1947, television was transmitted, as well as multichannel voice, in the first public demonstrations. Commercial rates were established on May 1, 1948, when AT&T announced that broadcasters would be charged $35 a month per airline mile for eight consecutive hours a day. Incremental hours could be had for $2 a month per mile. Rates were also established for terminal connections and occasional short-period usage. To whet appetites, AT&T pointed out that the impending interconnection of its separately established East Coast and Midwest networks made an audience of 40 million potential viewers accessible for a single program.

Acceptance was rapid, but one year's technological wonder was the next year's commonplace. In 1950, AT&T felt obliged to defend its rates publicly, pointing out that television could displace up to 600 voice channels on coaxial or radio routes, while charges were only seven times that for a full-time program circuit. Moreover, Bell System networking charges were only about five percent of the total cost to a sponsor for many common types of programs.

The television network grew rapidly. By the end of 1949, 8400 video-channel miles were in service, connecting 25 cities along the East Coast and in the Midwest. This grew to 15,000 channel miles and over 40 cities in the following year alone. By September 1951, two-way coast-to-coast links were available to transmit the opening of the Japanese Peace Treaty Conference in San Francisco. By 1953, 50,000 channel miles were in service, and, by 1957, a comprehensive national network was in existence [Fig. 22-7].[13] The network continued to expand, plateauing at a total of about 120,000 video-channel miles in the early 1970s.

Essentially all television transmission was on microwave radio as video transmission over coaxial cables had proven to be less satisfactory than radio and as the radio network became more widespread as well. A video channel on microwave radio occupied an entire 20-MHz band on TD2 and a 30-MHz band on TH. Even though the number of telephone circuits per radio channel increased, corresponding increases in television capacity were not possible. While the video network was large, it was static; the telephone network, on the other hand, continued to grow until television transmission became a com-

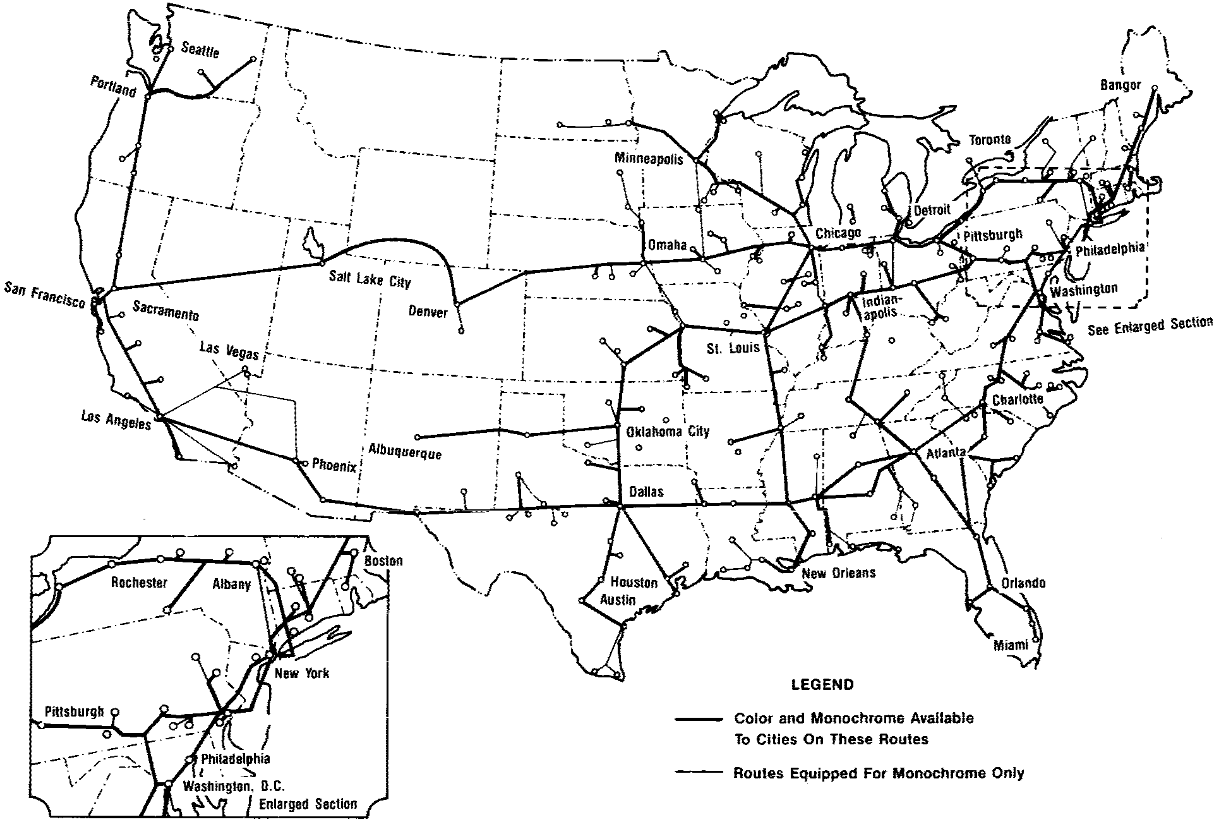


Fig. 22-7. The Bell System television network routes, 1957.

paratively small part of the signals carried. By the late 1970s, competing services, especially satellites (which Bell also offered), began to capture some of the business, and the total on Bell System terrestrial facilities slowly declined.

3.3.1 Color Television

The initial broadcasts and network-television transmissions were all monochrome (black and white), but motion pictures in color had become commonplace in the postwar years, and the early achievement of broadcast color television was a high-priority item in the minds of almost everyone in the industry. Bell Laboratories took part in and contributed to the discussions of the various methods proposed. In the late 1940s, Bell Laboratories collaborated with the Columbia Broadcasting System in experimental transmission of color signals from CBS's frame-sequential (color wheel) system.

When the National Television System Committee method for color transmission was adopted, AT&T collaborated with NBC in a demonstration of the system. On September 21 and 22, 1953, a number of color pictures were transmitted over a closed circuit between New York and Chicago for the benefit of the press and a number of advertising agencies. The Tournament of Roses Parade and the Rose Bowl football game were carried by microwave radio from Los Angeles to New York, and broadcast in color by NBC on New Year's Day of 1954. During the year that followed, programs were transmitted at a rate of one or two a week to gain experience, and, in subsequent years, they rapidly became commonplace. Network transmission of television signals in color was part of the planning for all postwar long-haul systems.

3.3.2 Television Operating Centers

Coordinating local television programs from studios and remote pickups was handled by the master control room of the television station concerned. However, when the station became part of a nationwide network, program material originating in different parts of the country had to be similarly coordinated. This task became a function of Bell System television operating centers (TOCs) located in many principal cities.

The TOCs permitted the rapid rearrangement of the transmission network to accommodate the needs of the television broadcasters. Not only were changes in the network needed from program-to-program, but commercials and items of local interest could be inserted into otherwise national or large-area hookups. A typical example is the broadcast of a local weather forecast in a network news program. TOCs equipped to carry out these functions quickly and efficiently were in service by 1957.[14] At a TOC, the incoming (source) lines and outgoing lines were arranged in matrix switches so that any source could be connected to any outgoing line. Both local baseband video links, such as lines to and from television-studio control rooms; and lines to the intercity facilities were available. By the late 1950s, matrices of 20 by 20 and 30 by 30 lines were in use [Fig. 22-8].

Elaborate precautions were taken against incorrect operation or equipment failures. The switches were controlled by buttons in horizontal and vertical arrays. Two buttons, one for the incoming and one for the outgoing line, had

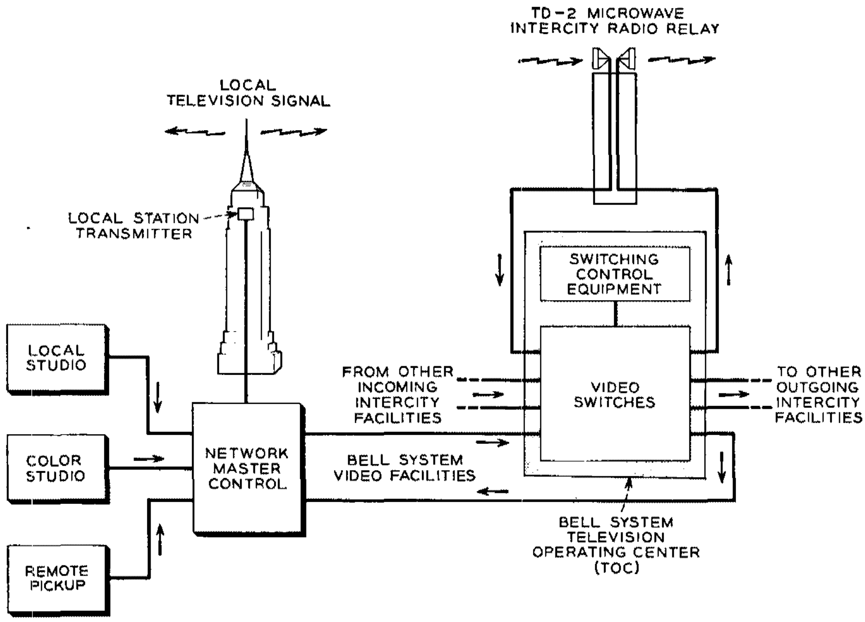


Fig. 22-8. A television operating center in relation to a local broadcaster and the national network, 1957.

to be depressed simultaneously to preselect a desired connection. The connection intended was indicated by lamps that could be checked before an EXECUTE button was operated at the moment the switch was to be made. It was also possible to preselect multiple connections, which could then be executed simultaneously by a salvo button. Finally, an ANNUL button was provided to cancel all preselections by a single action.

In addition to the switching function, the TOC provided monitoring and test positions for all lines that were served [Fig. 22-9]. From these, both the sound and picture on any line could be observed and communications established with people at the television studios or at the Bell System network facilities involved. Originally and for many years, the sound portion of a television program was transmitted separately from the picture over audio program circuits. Separate transmission of the two signals was necessary in the early days because the two signals were not compatible for simultaneous transmission over a single long-haul channel. The sound signals were originally switched by the broadcasters at their control centers, but the hazard of separate control of picture and sound were obvious, and, in later years, the functions were combined in the TOC. In time, as the signal-to-noise performance and equalization of the radio lines were improved, simultaneous transmission of video and audio became possible. In 1978, the Vicom Company Inc., extending an earlier development they had initiated, designed a system with two FM subcarriers above the video-signal frequencies. By the mid-1980s, the television

Fig. 22-9. Monitoring and test positions at a television operating center.

networks were beginning to broadcast stereo sound along with the picture from signals carried over the Bell System television network facilities.

3.3.3 PICTUREPHONE Visual Telephone Service*

The idea of adding visual to audio signals for two-way person-to-person communications is almost as old as telephony, although little could be done to bring it about before the electronics era.[15] The Bell System television-transmission experiments in the late 1920s and early 1930s were to explore visual displays as an adjunct to telephony. Interest shifted to network television for about 20 years, but, starting in the 1950s, the availability of less expensive video cameras, cathode-ray-tube (CRT) displays, and solid-state devices rekindled interest in the prospect for face-to-face communications, and serious work was resumed in Bell Laboratories.

In the mid-1950s, several possible two-way video telephone ideas were explored in the Bell Laboratories research area. In August 1956, a system employing a tiny picture tube with transmission over ordinary pairs was demonstrated to the Institute of Radio Engineers. The picture was semi-animated: it was still, but with a new update every two seconds [Fig. 22-10].[16] Work continued on a research and preliminary basis through the late 1950s. By October 1959, enough progress had been made in devices and understanding that a decision was made to develop a station set, later designated Mod I.[17]

---

* Registered service mark of AT&T.

Fig. 22-10. An experimental videotelephone, 1956. A still picture was updated every two seconds.

Person-to-person visual transmission raised a host of new questions for system planners. Among things to be decided were the picture size, the bandwidth (picture resolution), and the field of view (was a face-only picture adequate?). An early objective was for hands-free, speakerphone-type operation. It was felt that customers would not want to see someone speaking to them over a telephone but wished to talk face to face.

One of the principal tools for the continuing investigation was a variable-parameter video laboratory designed to carry out subjective tests on television pictures. The laboratory consisted of two television studios and a control room. Two-way transmission could be established over links in each of which major parameters could be varied over a wide range. Two sets of conditions could be compared by setting them up in advance and switching back and forth between them while viewing [Fig. 22-11]. The outcome of this work was a decision to develop a full-motion system with a 0.5 MHz bandwidth and to present a head-and-shoulders view of the speaker on a screen of about six inches in diameter.[18]

By early 1964, all the components for a full-system test were available. A number of terminal sets were constructed and provided to Bell Laboratories

Picture size	12 to 280 square inches
Aspect ratio	3:4 to 4:3
Frame rate	15 to 75
Interlace ratio	1:1 (no interlace) to 8:1
Lines per frame	50 to 1024
Bandwidth	1/8, 1/4, 1/2, 1, 2, 4, and 7 MHz

(a)

(b)

Fig. 22-11. The variable parameter two-way video system laboratory. (a) The range of the variable parameters. (b) The laboratory in operation, 1963.

employees at Murray Hill and Holmdel, New Jersey for the first operation in an office environment. Later the same year, a major demonstration based on the same design was conducted at the 1964 New York World's Fair. The exhibit consisted of six individual booths for calls within the exhibit and one booth that was linked by a microwave-radio channel to a booth at Disneyland in California. Subjects to try the system were selected at random from visitors to the Bell exhibit at the fair [Fig. 22-12].

In August and September 1964, about 700 users of the system at the world's fair were interviewed for their reaction. The response was encouraging. About 80 percent thought the picture quality was good or better, only a few thought it poor and about 50 percent thought that a visual service would be an important and useful adjunct to telephony. The interviews also showed, however, considerable sensitivity to hypothetical price levels for the service, especially for residential use. During the same year, exploratory commercial service was established between special public booths in the Chicago Museum of Science and Industry, the National Geographic building in Washington, D.C. and Grand Central Station in New York. The booths were operator attended and calls had to be set up by appointment. Except for a period of initial novelty, interest and use were low.

After the encouraging results at the New York fair, a more elaborate in-house trial was set up starting in April 1965, and continuing for six months. Twenty-eight sets were provided to Bell Laboratories management people at Murray Hill and Holmdel with two interlocation two-way trunks. During the same year, a product trial of the so-called Mod I design was carried out in cooperation with the Union Carbide Corporation, which had 23 sets installed in Chicago and 12 in New York. Late in 1965 and in 1966, sets were also installed at AT&T headquarters in New York. The Bell System installations were expanded in 1967 to a New York area corporate network, linking AT&T

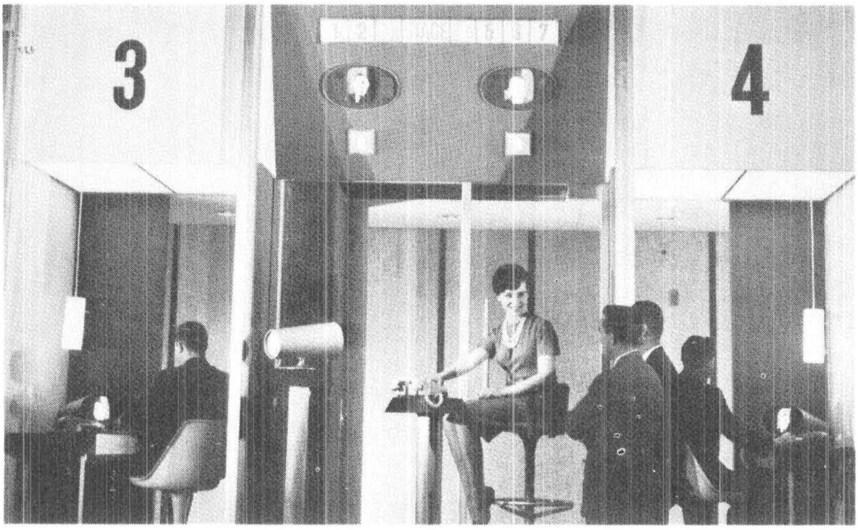


Fig. 22-12. The Bell System PICTUREPHONE visual television service exhibit at the 1964 New York World's Fair.

and Western Electric buildings in New York City with three Bell Laboratories New Jersey locations.

PICTUREPHONE visual telephone service was popular with the participants in these tests. The Union Carbide participants were reluctant to part with their sets when the tests were complete, and, in Bell Laboratories, 83 percent of calls that could be placed over the visual links used the system. At the same time, the Bell Laboratories viewers were more critical than the public at the World's Fair; only 65 percent judged the pictures good or better in quality. In addition, essentially all parties agreed that graphics were essential to supplement the face-to-face picture and that, for the graphics desired, a higher resolution was needed. Armed with these results, development was redirected toward a 1-MHz design (Mod II) with improved controls.

The development of the Mod II system was probably the largest development undertaken in Bell Laboratories directed toward a new service offering (as distinct from a new facility design). Almost every area was involved, from device development to station terminals to transmission and switching, all of which required new capabilities.[19,20] Local transmission was by means of a six-wire circuit—a two-wire pair for voice and a separate physical four-wire circuit for both directions of picture transmission [Fig. 22-13]. Switching of the picture signal was by an auxiliary switch slaved to the regular voice switch. Local transmission was by an analog signal over ordinary pairs, but with repeaters and special equalizers for the picture signal. Intercity transmission was to be accomplished by converting the 1-MHz analog picture signal to a 6.3-megabit digital signal using differential PCM. Three 6.3-megabit signals were combined in a special multiplexer to provide three PICTUREPHONE service trunks over a 20-MHz TD2 radio channel.[21] (See Chapter 20, Section 4.1.)

The subscriber station was entirely redesigned using the latest available cameras, display tubes, and integrated circuits. The picture was made up of 251 interlaced lines at 30 frames per second on a screen 5 by 5 1/2 inches. Desk-top graphics could be displayed by a flip-up mirror in front of the camera [Fig. 22-14].

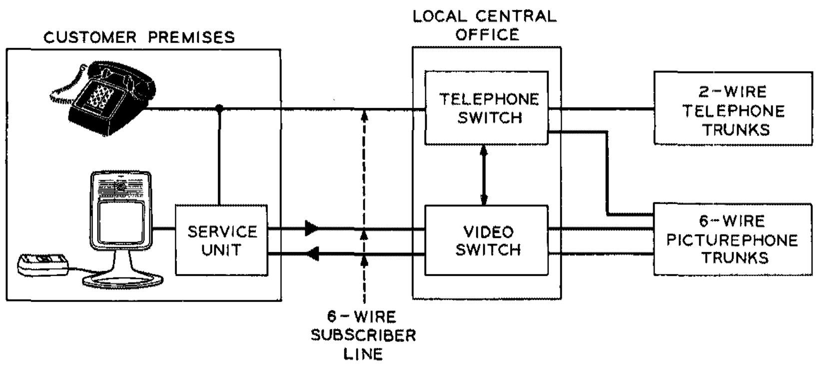
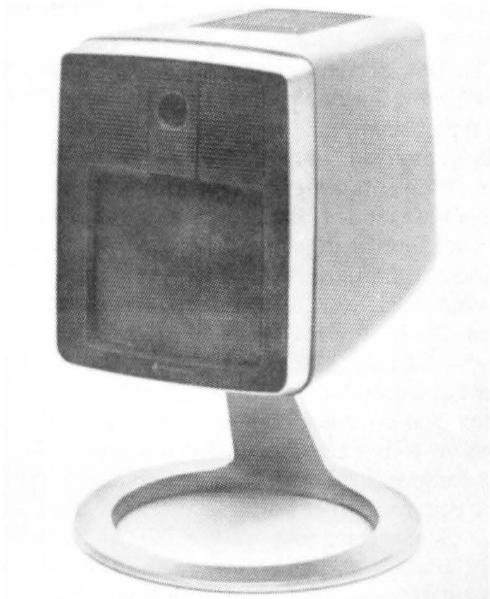
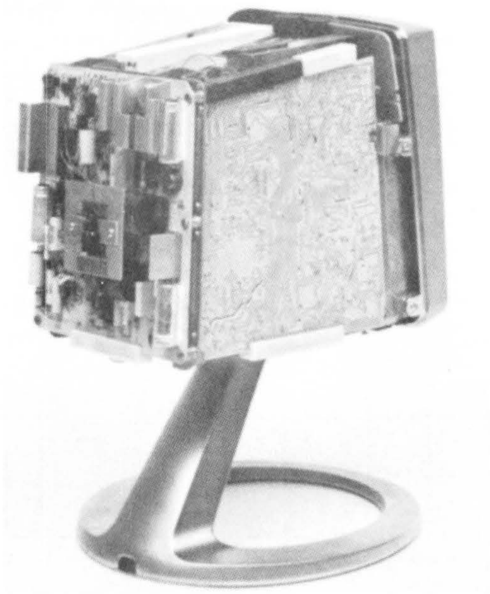


Fig. 22-13. Basic local arrangement for PICTUREPHONE visual telephone service.



(a)



(b)

Fig. 22-14. The Mod II PICTUREPHONE visual telephone service subscriber station. (a) The 1A display unit, front view. (b) The 1A display unit without the cover.

A six-month service trial was started in February 1969, with 40 Mod II terminals installed in the Pittsburgh and New York offices of the Westinghouse Electric Corp. The trial was judged to be successful, and commercial service was launched on a local basis in both Pittsburgh and Chicago in 1970. Service in New York was deferred because of severe service problems at that time.

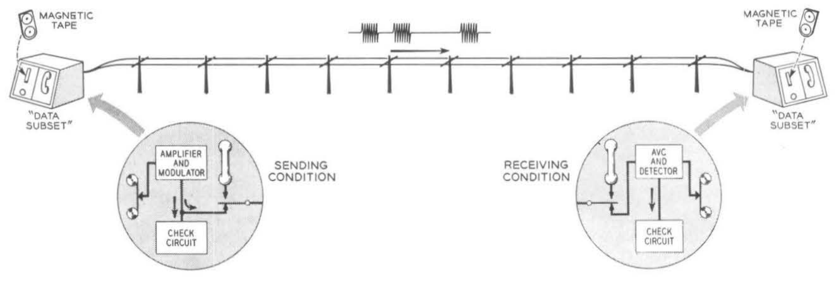
Between 400 and 500 stations were installed, mostly in Chicago, but subscribers were slow to appear. There was much debate about the cause. Clearly, the limited number of sets was a factor; it was difficult to interest customers in a new service when there were so few others to call. But the equipment was expensive and charges were high as well. Basic rates were in the vicinity of $100 per month, which included a fee for the terminal set and a connection charge. One-half hour per month of video calls was included, with additional time charged at $.15 per minute. Toll calls, when they became available, would have been correspondingly high. For whatever reason, the service did not become popular. When Illinois Bell drastically reduced the rates as an experiment and subscriptions still did not pick up much, active promotion was suspended. The technology was ready, but the public was not.

3.4 Data Transmission

The need for data communications at higher than the conventional telegraph rate of 75 bits per second arose about 1950. One of the first applications involved transmitting radar data from remote installations to government data-processing centers. By the mid-1950s, the Bell System was providing data links at 1600 bits per second for use in the SAGE system, an early-warning complex for the Strategic Air Command.[22] About the same time, with the growing use of digital computers, it became clear that there was going to be a commercial need for data communications at rates higher than teletypewriter speeds. To provide the service, two things were needed: (1) A data terminal, or modem (modulator/demodulator), to convert the binary baseband digits available from a computer into a form suitable for transmission over the nominal 250- to 3000-Hz band of a telephone voice channel and to reconvert the line signal to the computer format at the receiving end. (2) A characterization of the transmission plant to determine its ability to carry the new type of signal, both for dialed-up switched connections (DDD) and for full period private lines.

A research group at Bell Laboratories built and tested several different types of experimental modems during the mid-1950s. They also carried out analytic studies of the transmission characteristics of telephone channels and the best way to transmit the data signals over them.[23,24]

The early research terminals operated in two modes; data signals from a source could be converted and transmitted as the signals were generated or they could be recorded on magnetic tape and transmitted over the line at a uniform high speed [Fig. 22-15]. In 1958, a commercial service based on these concepts, using a data set transmitting at 1000 bits per second by means of frequency-shift keying (FSK), was offered on a limited basis [Fig. 22-16]. An extensive series of measurements on typical intercity voice channels in 1958 and 1959 provided the information on which the ability of the network to transmit voice-band data was based.[25]



(a)

(b)

Fig. 22-15. Experimental data transmission over telephone channels, 1957. (a) Transmission from a prerecorded magnetic tape via a data subset modem. (b) Electric typewriter and experimental equipment for use of magnetic tapes to store data for transmission.

By the early 1960s, an extensive line of data sets had been developed, ranging from narrow band sets (100 series) for transmission at the teletypewriter rate of 150 bits per second through several sets for voice-band data (200 series) for transmission typically at 1200 bits per second, to wideband sets (300 series) for transmission at 50 and 250 kilobits per second in the group and supergroup bands, respectively. Special sets were also developed for low-speed parallel transmission and other purposes. Voice-band data became by far the most general type of service, with the wideband higher rates and special sets limited to fewer customers with special needs. Terminals were either asynchronous, in which the data was not transmitted at a fixed rate but intermittently as generated, or synchronous, in which case the receiver expected a signal in each interval. Synchronous transmission was used for the higher rates. Frequency, phase, and amplitude modulation were all used in one set or another.

Frequency-shift keying was usually used for asynchronous data transmission

Fig. 22-16. Demonstration of 1000 bit per second data service, 1958.

where the keying rate could vary and was established by the customer. FSK was used in all low-speed, 100-series data sets, as well as in the medium-speed 202-type sets. In the latter, a constant-level carrier was shifted between 1200 and 2200 Hz at rates up to 1200 bits per second over switched connections and up to 1800 bits per second over private lines. (Service over the switched network had to take channels pretty much as they came. Private-line service could be more selective in avoiding the more limiting types of facilities and could tailor a channel for better performance.) Phase-shift keying (PSK) was used for synchronous data transmission, where the clock rate was established by the data set.

In theory at least, the signal-to-noise ratio in a typical telephone channel was good enough to support considerably higher information rates. Only binary, two-level signal states were used in the early data sets, and the signaling rates were deliberately kept fairly low to simplify the modem circuits and to provide a rugged error-free signal. For higher signaling rates the gain and delay characteristics of the channels had to be equalized. And, since the deviations in gain and phase varied from channel to channel, especially in switched connections, it was essential that the equalization adapt to the channel, that is, be automatic. This was first accomplished in 1964 by means of a tapped delay line equalizer (also called a transverse filter) in the receiving modem. As the distorted signal passed through a delay line, it was measured at a sequence of taps. The outputs of the taps, properly adjusted in amplitude, were summed and recombined with the original signal to minimize the distortion. A feedback control from the combined output was used to adjust amplifiers in each tap, automating the equalization. The process could be thought of as creating artificial echoes of the right magnitude, delay, and polarity to cancel out the distorting echoes from line irregularities [Fig. 22-17].[26]

Four-phase PSK was used in the mid-1960s in data sets of the 201 type, one version of which was capable of operating at 2400 bits per second over either the public switched network or private lines. Later, by the use of an

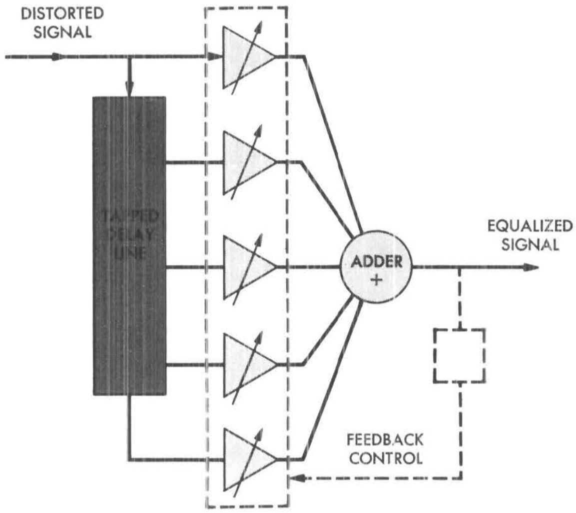


Fig. 22-17. A transverse filter tapped delay line automatic equalizer for data transmission, 1964.

adaptive equalizer, eight-phase modulation was possible in the 208-type sets. With these, a signaling rate of 4800 bits per second was possible over either private or switched connections [Fig. 22-18].

Frequency- and phase-shift keying were attractive because they avoided problems that could arise from amplitude nonlinearities in the channel, but the highest signaling rates were realized with amplitude modulation. By the early 1970s, 203-type data sets, using vestigial-sideband multilevel modulation, provided synchronous data rates up to 4800 bits per second on the switched network and 7200 bits per second on private lines. Finally, since double-sideband AM signals have no quadrature component in a channel that is symmetrical about the carrier, it was possible to transmit two orthogonally phased signals in the same channel. Using quadrature amplitude modulation, the 209-type data sets achieved 9600 bits per second (or two independent 4800 bit-per-second channels) over private-line facilities. Both the 203- and 209-type terminals had automatic equalizers.

Coincident with the technological changes that occurred during the 1960s, a major regulatory and policy change also took place regarding the interconnection of non-Bell System data sets to the switched network. Prior to 1968, except for certain private lines, it was illegal to connect any non-Bell equipment to the Bell System plant. This restriction was removed in 1968 as a result of a court decision in the landmark Carterphone case. From 1968 to 1975, it was permissible to connect non-Bell modems to the switched telephone network as long as power limitations specified in the tariff were observed and provided a Bell-furnished protective coupler was used to ensure adequate network protection. Starting in 1976, the protective-coupler requirement was also eliminated

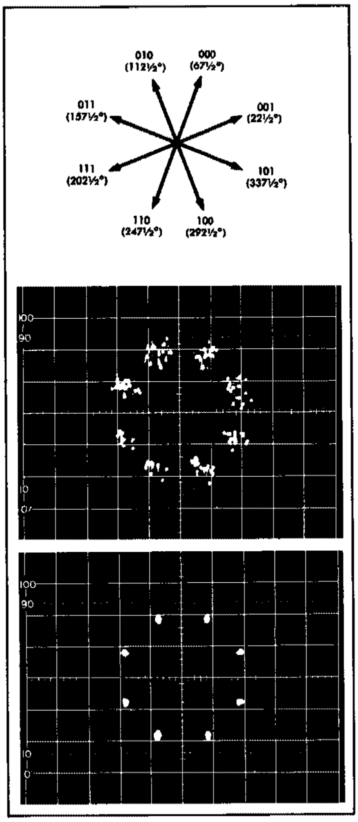


Fig. 22-18. Eight-phase modulation in which each phase shift conveys three bits of information, 1973.

and any data set registered with the FCC could be connected directly to the telephone network. The advent of freer interconnection led to increased competition and the entry into the marketplace of many new data-set vendors. It also meant that the switched network had to transmit data signals properly in an environment where the Bell System had less control over the characteristics of the signals it was required to carry.

The developments in data sets after 1970 were the result of improvements in devices rather than of improvements in modulation techniques and were in response to the growing need to transmit large amounts of data rapidly and practically error free as communication between computers became increasingly common. Starting in 1972, the Bell System introduced a new family of voiceband data sets that exploited integrated circuit technology to achieve better performance and dramatic size reductions compared to their predecessors [Fig. 22-19].[27]

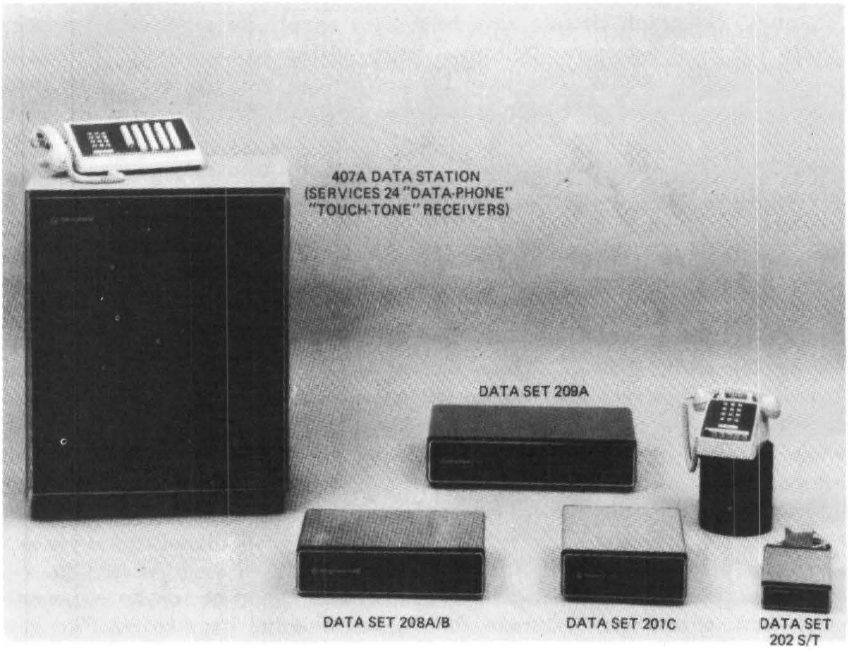


Fig. 22-19. Bell System data sets of the early 1970s.

3.4.1 Data Transmission Over Digital Facilities

Voice-band data signals were transmitted over T carrier channels in the same way as any other voice-band signal, although it always seemed extravagant to the designers to devote an entire 64-kilobit digital channel to transport data at a small fraction of that rate. Starting in the mid-1960s, some use was made of the high T carrier digital-line rate in special terminals, the T1WB (wide-band) terminals. With these terminals, a number of high-rate channels, ranging from eight at 64 kilobits to one at 512 kilobits, could be derived for transmission over the 1.544-megabit T1 line. The capacity was only one-third of what might have been expected, because it required three bits to identify the transitions of nonsynchronous incoming data.[28]

3.4.2 Digital Data Systems

In the late 1960s, as the use of computers expanded enormously, it was recognized that something more than voice-band data links or the limited availability of wideband services was highly desirable. This recognition led to the development of a purely digital system, the DATAPHONE* data communications service (DDS), which dispensed with all digital-to-analog and analog-to-digital conversions.

---

* Registered service mark of AT&T.

DDS provided a full-duplex, private-line network designed for point-to-point and multipoint digital data transmission at synchronous rates of 2.4, 4.8, 9.6, 56, and 1544 kilobits per second. The lowest three rates were referred to as subrates, as they were multiplexed up to a 64-kilobit rate before transmission over digital trunks to distant offices. The higher two rates occupied a single 64-kilobit T1 channel or the entire DS1 rate signal of a T1 line, respectively [Fig. 22-20].

A variety of new equipment was required to interconnect two DDS customer stations in different digital serving areas [Fig. 22-21].[29,30] The local-distribution portion of a DDS connection used metallic, twisted-pair cables for a full-duplex four-wire transmission path between the customer premises and the DDS serving office. A channel service unit (CSU), furnished as an integral part of a DDS channel, terminated the four-wire loop at the customer premises. The CSU included circuitry to permit signal loopback from a DDS test center. At the DDS serving office, the loop was terminated in an office channel unit (OCU). The OCU encoded the incoming data signals into an eight-bit byte format that added necessary control information and, regardless of the data service rate,

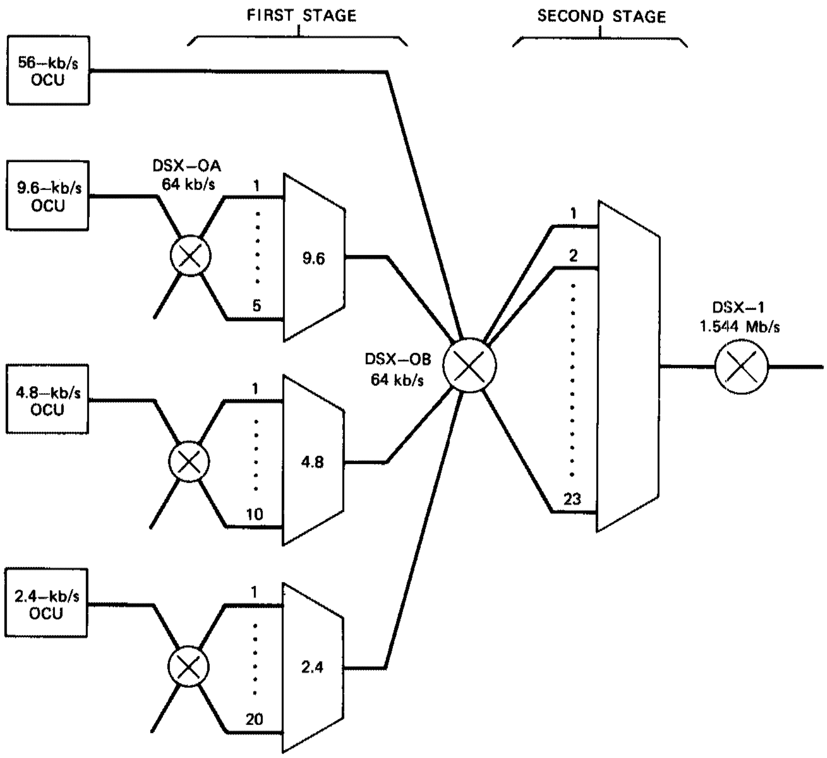


Fig. 22-20. Two-stage multiplex hierarchy for DATAPHONE data communications service. (OCU is office channel units; DSX indicates crossconnects.)

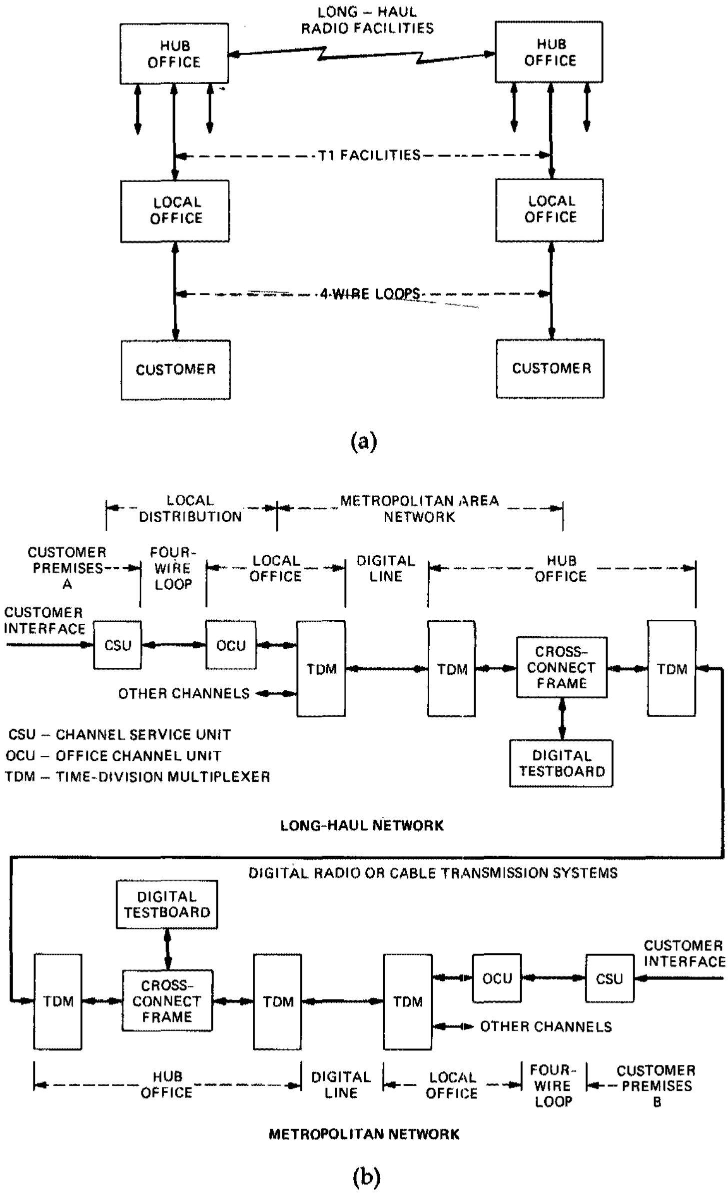


Fig. 22-21. Transmission arrangements for DDS. (a) Customer to office and interoffice connections. (b) Details of a DDS point-to-point connection.

built the signal up to a rate of 64 kilobits. For example, in the case of a 2.4-kilobit subrate data signal, the OCU added two bits to each block of six customer data bits to form an eight-bit byte and repeated the byte 20 times in succession (byte stuffing). For the 56-kilobit data rate, the OCU merely inserted a control bit after each block of seven customer data bits. By this means, up to twenty 2.4-kilobit, ten 4.8-kilobit, five 9.6-kilobit, or one 56-kilobit signal could be multiplexed into a single 64-kilobit T1 channel. The resulting 64-kilobit signal was used as the standard for interconnecting DDS transmission equipment. In a second stage of multiplexing, the 64-kilobit signals were combined in a special terminal (the T1WB4) for the T1 line. T1WB4 permitted up to 12 data channels and from 1 to 12 voice channels to be transmitted over a T1 line at the 1.544-megabit rate in the standard format. An alternative terminal provided up to 23 DDS channels.[31] (In the early 1980s, a single channel plug-in card [Dataport] provided a 64-kilobit interface that made it possible to provide a DDS 64-kilobit rate link over any T carrier facility from an otherwise standard D4 channel bank.)

In the early planning, intercity connections were to be provided over T1 lines for short distances or, with suitable multiplexes, over a T2 6.3-megabit line. For longer distances, channels of 13 megabits in a coaxial mastergroup band or 20 megabits in a microwave-radio channel were planned. In practice, however, by far the largest number of intercity connections for DDS was provided over the radio systems by a multilevel 1.5-megabits signal in a blank portion of the spectrum under the voice mastergroups [Fig. 22-22]. This system, the 1A-RDS (1A radio digital system), better known as DUV, for digits under voice, was implemented in time for the first DDS service in 1974. By 1984, over 90,000 route miles of TD and TH radio were equipped with DUV.[32,33]

In each major metropolitan area, a centrally located office was designated as a hub office for DDS. In addition to serving stations in its immediate vicinity, the hub provided a number of other functions. These included: (1) precision timing for network synchronization, ultimately tied to the master reference clock in Hillsboro, Missouri, (2) cross-connection facilities to assemble packages of channels for intercity transmission, (3) connection to a multipoint junction unit to permit single-to-multistation transmission, (4) access to the long-haul transmission facilities, and (5) test access to individual channels.

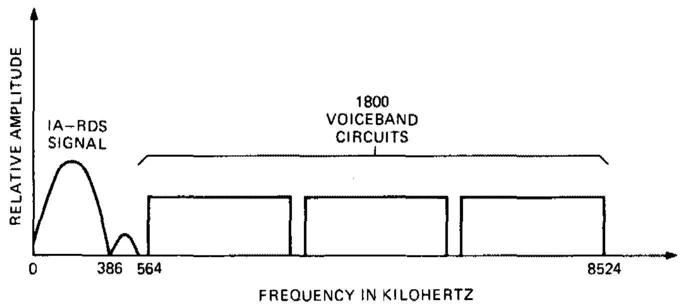


Fig. 22-22. The baseband spectrum of the 1A-RDS data under voice signal with three voice mastergroups, 1974. The entire band could be transmitted over TD or TH microwave radio systems.

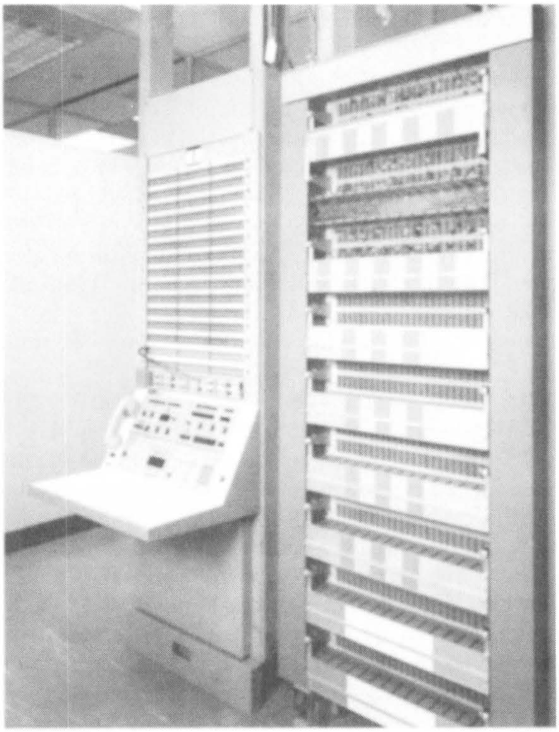


Fig. 22-23. The 950A test board and cross-connect bay for DATAPHONE data communications service.

Testing arrangements were an extremely important part of the DDS plan. The system was designed to provide a very high quality of service. The performance objective was error-free transmission in 99.5 percent of all one-second intervals, and availability, the complement of the average annual downtime, was to be at least 99.96 percent. To meet these objectives, all facilities were preservice qualified and the health of the 1.544-megabit links was continuously monitored in service. Digital facilities for DDS were protected by automatic switching to spare lines. Finally, centralized and remotely controlled fault isolation permitted end-to-end checking of a customer's link from a central test board [Fig. 22-23]. From the test board, loopbacks could be actuated at both the near end and far end of a customer's channel by a single tester without assistance. Single-point testing and trouble isolation were complemented by administrative arrangements that provided a single point of contact for a customer for all DDS problems.

DDS service to five cities started in December 1974. By the end of 1975, a backbone network of 24 cities was interconnected, and preparations were under way to expand the network to over 100 cities.

REFERENCES

1. R. F. Rey, ed., *Engineering and Operations in the Bell System,* 2nd ed. (Murray Hill, NJ: AT&T Bell Laboratories, 1984).
2. L. G. Abraham, A. J. Busch, and F. F. Shipley, "Crossbar Toll Switching System," *Trans. Amer. Inst. Elec. Eng.* **63** (1944), pp. 302-309.
3. H. R. Huntley, "Transmission Design of Intertoll Telephone Trunks," *Bell Syst. Tech. J.* **32** (September 1953), pp. 1019-1036.
4. H. H. Felder and E. N. Little, "Intertoll Trunk Net Loss Maintenance under Operator Distance and Direct Distance Dialing," *Bell Syst. Tech. J.* **35** (July 1956), pp. 955-972.
5. A. E. Ritchie and J. Z. Menard, "Common Channel Interoffice Signaling: An Overview," *Bell Syst. Tech. J.* **57** (February 1978), pp. 221-224.
6. C. A. Dahlbom and J. S. Ryan, "Common Channel Interoffice Signaling: History and Description of a New Signaling System," *Bell Syst. Tech. J.* **57** (February 1978), pp. 225-250.
7. See special issue on Streamlining Special Services Operations, *Bell Lab. Rec.* **54** (June 1976), pp. 142-174.
8. J. W. Gorgas, "AUTOVON: Switching Network for Global Defense," *Bell Lab. Rec.* **46** (April 1968), pp. 106-111.
9. A. B. Clark and C. W. Green, "Long Distance Cable Circuits for Program Transmission," *Bell Syst. Tech. J.* **9** (July 1930), pp. 567-594.
10. R. E. Pierce and E. W. Berris, "A Transmission System for Teletypewriter Exchange Service," *Bell Syst. Tech. J.* **15** (October 1936), pp. 529-548.
11. B. P. Hamilton, H. Nyquist, M. B. Long, and W. A. Phelps, "Voice Frequency Carrier Telegraph System for Cables," *J. Amer. Inst. Elec. Eng.* **44** (March 1925), pp. 213-218.
12. J. A. Duncan, R. D. Parker, and R. E. Pierce, "Telegraphy in the Bell System," *Trans. Amer. Inst. Elec. Eng.* **63** (1944), pp. 1032-1044.
13. J. M. Barstow, "Television and the Bell System," *Bell Lab. Rec.* **32** (March 1954), pp. 81-87.
14. C. A. Collins and L. H. Hofman, "Switching Control at Television Operating Centers," *Bell Lab. Rec.* **35** (January 1957), pp. 10-14.
15. G. R. M. Garratt and A. H. Mumford, "The History of Television," *Proc. IEE London,* Part IIIA, **99** (1952), pp. 25-42.
16. W. E. Kock and R. L. Miller, U.S. Patent No. 2,895,005; filed September 30, 1954; issued July 14, 1959.
17. D. N. Carson, "The Evolution of PICTUREPHONE Service," *Bell Lab. Rec.* **46** (October 1968), pp. 282-291.
18. A. D. Hall, "Experiments With PICTUREPHONE Service," *Bell Lab. Rec.* **42** (April 1964), pp. 114-120.
19. See special issue on PICTUREPHONE, *Bell Lab. Rec.* **47** (May/June 1969), pp. 134-187.
20. See special issue on the PICTUREPHONE® System, *Bell Syst. Tech. J.* **50** (February 1971), pp. 219-700.
21. D. W. Nast and I. Welber, "Transmission Across Town or Across the Country," *Bell Lab. Rec.* **47** (May/June 1969), pp. 162-168.
22. A. E. Ruppel, "SAGE Data Transmission Service," *Bell Lab. Rec.* **35** (October 1957), pp. 401-405.
23. P. Mertz and D. Mitchell, "Transmission Aspects of Data Transmission Service Using Private Line Voice Telephone Channels," *Bell Syst. Tech. J.* **36** (November 1957), pp. 1451-1486.
24. W. A. Malthaner, "High-Speed Data Transmission," *Bell Lab. Rec.* **35** (April 1957), pp. 121-125.

25. A. A. Alexander, R. M. Gryb, and D. W. Nast, "Capabilities of the Telephone Network for Data Transmission," *Bell Syst. Tech. J.* **39** (May 1960), pp. 431-476.
26. R. W. Lucky, "Automatic Equalization for Digital Communication," *Bell Syst. Tech. J.* **44** (April 1965), pp. 547-588.
27. E. R. Kretzmer, "The New Look in Data Communications," *Bell Lab. Rec.* **51** (October 1973), pp. 258-265.
28. L. F. Travis and R. E. Yaeger, "Wideband Data on T1 Carrier," *Bell Syst. Tech. J.* **44** (October 1965), pp. 1567-1604.
29. See special issue on Digital Data System, *Bell Syst. Tech. J.* **54** (May/June 1975), pp. 811-964.
30. C. R. Moster and L. R. Pamm, "The Digital Data System Launches a New Era in Data Communications," *Bell Lab. Rec.* **53** (December 1975), pp. 420-426.
31. E. T. Mackey, W. J. Maybach, and S. B. Pfeiffer, "Mixing Data and Voice on the T1 Line," *Bell Lab. Rec.* **53** (February 1975), pp. 136-142.
32. R. C. Prime and L. L. Sheets, "The 1A Radio Digital System Makes 'Data Under Voice,' A Reality," *Bell Lab. Rec.* **51** (December 1973), pp. 334-339.
33. R. R. Grady and J. W. Knapp, "1A Radio Digital Terminals Put 'Data Under Voice,' " *Bell Lab. Rec.* **52** (May 1974), pp. 160-166.

Chapter 23

Maintenance and Operations

I. INTRODUCTION

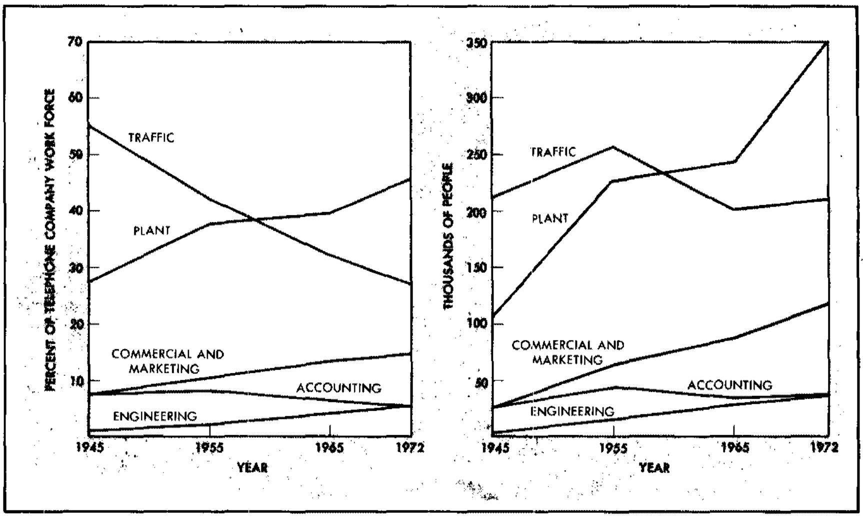
Although transmission systems were designed and manufactured to high-quality standards, the performance of which they were capable could only be realized if the systems were properly operated and maintained. Effective operation and maintenance required that problems were promptly identified, located, and eliminated. The trouble could be a sudden catastrophic fault or a more gradual drifting of one or more performance parameters outside the desired maintenance limits.

Routine tests to check that circuits were performing up to objectives were made from the earliest days, especially on long toll circuits. But in the late 1920s and 1930s when toll trunk groups were small and customers were queued up for access, the reports of operators who encountered problems in setting up calls or customer complaints were frequently the first notice of troubles. When most circuits were at voice frequency, the distinction was not very great between maintaining a trunk, the voice-band link between switching points, or a facility, a cable or open-wire line carrying several circuits. After carrier systems became common, especially the larger-capacity ones, many maintenance features unique to each facility system were built in. Broadband facility maintenance became a more specialized activity, and, although it was still intimately related to trunk maintenance, it was the concern of different groups of development and operating people. Facility-maintenance features became an integral part of all broadband system designs.

Local dialing became widespread during the 1920s and 1930s and almost universal by the end of the 1940s. Customer toll dialing (direct distance dialing—DDD) started in 1951 and spread very rapidly, becoming essentially universal by the end of the 1960s. From the viewpoint of the switching office plant people, the nature of the interoffice transmission facility was a matter of indifference. Their concern was that the trunk their switches seized could transmit signals to the distant office to establish and control the connection and that people could talk over it when the connection was made. Automatic dialing eliminated the operator's ability to screen out and report troubles before a customer encountered them. Starting in the 1940s and evolving over the years through several generations of design, means were developed for the routine testing of both local and toll trunks.

As traffic grew explosively in the years after World War II and the plant facilities required to serve it grew in extent and complexity, the number of people required to operate and maintain the plant grew as well. Customer dialing and the centralization of operators with the aid of the traffic service position system (TSPS) averted what would otherwise have been a crisis in providing a sufficient number of traffic operators for the expanding business. With local and long-distance customer dialing and TSPS, about the same number of operators were handling five times as many calls in 1975 as they did in 1945. But nothing comparable happened to limit the number of plant people required to operate and maintain the equipment. Bell System plant personnel more than tripled from 110,000 (28 percent of the total staff) to over 350,000 (about 45 percent of the total staff) in the same period. The increasing complexity of the plant was indicated by an increase by a factor of ten in the engineering force over the same period [Fig. 23-1].[1]

In the mid-1960s, the advent of so-called third-generation computers, both large central processors with enormous capacity for data storage and relatively inexpensive but powerful minicomputers, made it possible to bring the power of computers to bear on plant administration and maintenance problems. The new computers could store and quickly retrieve equipment records and related information. They could also be programmed to control much of the repetitive routine work associated with plant maintenance. By the mid-1970s, literally dozens of computer-based systems were in use to address these tasks and many more were in various stages of planning and development. Transmission facilities and the maintenance systems to keep them working properly came



(a) (b)

Fig. 23-1. The change in the composition of the Bell System work force, 1945 to 1972. (a) Percent of total in each category. (b) Total number in each category.

increasingly to be viewed as a single system that had to be designed and operated on an integrated basis. In turn, the computerized transmission operations-support systems had to be coordinated with similar computer-controlled operations systems for other plant facilities, such as switching, and all the plant facility systems with computer-aided systems for traffic, circuit orders, billing, and all of the multitude of activities necessary to the operation of the entire business. In the late 1970s, the hardware and software development of computer-based systems to assist in the day-to-day operation of the business, and the coordination of these operations systems for all the functions served, became a major activity in Bell Laboratories.[2] The automation of transmission testing, maintenance, and record keeping was a significant part of this effort.

II. MAINTENANCE OF THE TRANSMISSION NETWORK (1925–1940)

2.1 Local-Network Maintenance

In 1925 and for many years thereafter, noisy or inoperative loops were usually first detected by customers when they attempted to make calls. The customers reported the condition to operators, who, in turn, referred the problems to maintenance groups for trouble isolation and repair. Problems with local trunks in manual offices were generally detected by operators in the course of trying to establish interoffice calls. Here too, the operators would refer the problems to maintenance groups. The introduction of dial-operated local switches created a problem in detecting troubles in local trunks. Operators could judge the quality of transmission on trunks, but switches could not.

In step-by-step switching offices, trunk maintenance was especially difficult, as the switches pulsed forward blindly without any acknowledging signal from the far end. There was no indication that the connection had even been established, much less that it was a good one. Unless a faulty trunk was held by the calling party until the defective line was identified, the trouble could only be located by testing the trunks in sequence. The situation was better with the panel switches used in large metropolitan offices. Panel switches received revertive pulses from the far end while setting up a call. A failure in the sequence stopped the process and the defective trunk could be identified. There was little routine measurement of transmission. Local trunks were generally assumed to be in good order unless there was an indication to the contrary. Testing and troubleshooting, when necessary, were manual operations, with a tester at each end to transmit and measure.

2.2 Long-Distance Maintenance

In the pre-World War II long-distance network, toll operators established all calls. As in the manual-switching days of the local network, operators would determine if a long-distance trunk needed maintenance by simply listening to the quality of transmission as they were setting up calls. They then referred defective trunks to a toll-maintenance group.

If the problem was in a metallic circuit, it was localized by methods dating back to early telephone and telegraph days. A battery supply and voltmeter were used to determine if the fault was an open wire, a short between the wires of a pair, or from either to ground. If the problem was a short, a Wheatstone bridge was used to measure resistances and establish at least the approximate location. DC resistance measurements to the fault were compared to the resistance of a good pair over the known distance to the next test point. For opens, capacitance measurements between the wires of a pair using the same equipment as an impedance bridge furnished the necessary information. Since the capacitance per mile was known, the total capacitance indicated the distance to the break. Capacitance measurements were made at a frequency of only 4 Hz to avoid the transmission-time effects; the low frequency was especially necessary if the line was loaded. The measuring equipment was conveniently arranged in the toll test board, with patch cords to associate it with the line to be tested [Fig. 23-2].[3] Measurements from both ends and the experience and skill of the testers improved the localizing.

Even if the trunk in trouble contained repeaters or was on carrier, DC methods might still be used. The early voice-frequency repeatered lines and Type C carrier systems preserved DC continuity for telegraph circuits. But if the problem was in the electronic equipment, other procedures were necessary. Until the introduction of Type K carrier in 1938, repeaters were at attended stations. When a fault occurred, the staff at the terminal station responsible for trunk maintenance in that section conducted transmission tests from the terminal toll test board to each repeater location in turn. When the problem had been sectionalized to a particular repeater section, maintenance personnel at the adjacent attended stations working with the toll test board would first check their local equipment and, if this was found in order, cooperate in using DC and capacitance measurements to localize a fault in the intervening metallic span.

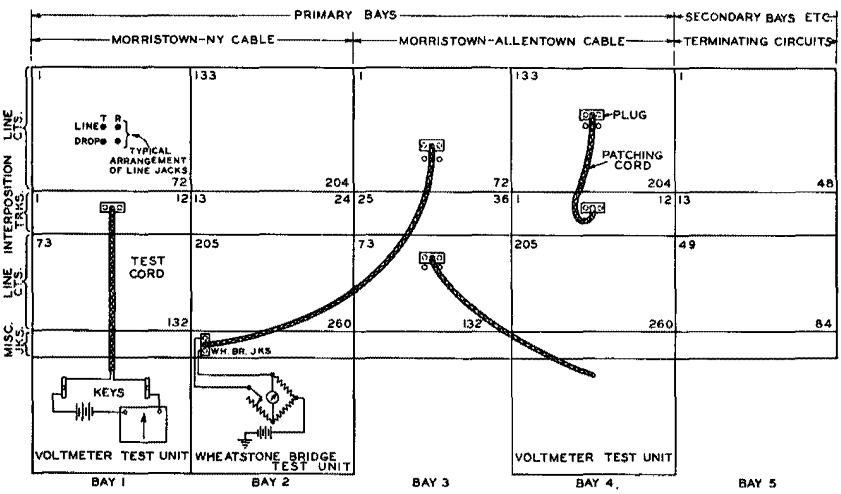


Fig. 23-2. The arrangement of a battery-voltmeter test unit and Wheatstone bridge at a toll test board, 1928.

III. TRANSMISSION MAINTENANCE (1940–1975)

Long-haul circuits were more expensive and generated greater revenues than local ones, and the loss of a circuit out of a small trunk group affected a higher proportion of calls. Restoring service on failed long toll circuits was, therefore, more urgent than in the local network where alternate facilities were more available. Locating and correcting a problem was often time-consuming, so the first order of business was usually to patch in spare equipment at a terminal or repeater site or even to restore the trunk by patching in a completely new line. The patch might be to different pairs on the same route or even to a circuit over a different route between the same endpoints.

3.1 Order Wires

All this activity, of course, required a great deal of communication and cooperation among maintenance people, often at several locations. In the early days, the working voice circuits were precious, and maintenance communication was usually by means of telegraph order wires. Not all maintenance people were skilled telegraph operators, however, and, as circuit cross sections grew and especially when toll cables became common, telegraph links became inadequate. Teletypewriters continued to be used for many years because of the value of their recorded messages, but dedicated telephone order wires for maintenance became the rule. Both two- and four-wire circuits were used, the latter where the line served was long and repeaters were needed in the order wire. Talking connections at the terminal test boards were by standard operator telephone sets. At intermediate points, talking paths could be made via monitoring points provided in the repeaters.

The major difference between order wires and ordinary lines was in the signaling. An order wire was like a long party line with a large number of subscribers. There usually were several locations to be reached within each terminal office, as well as multiple points along the line. It was desirable to be able to alert the person needed at any location without needlessly disturbing other busy people. One solution was a separate telegraph circuit dedicated to signaling. In a typical arrangement, 24 stations in eight different offices could be alerted selectively, but the equipment was modular and, with add-ons, up to 174 locations in 21 offices could be signaled.[4] It is not difficult to imagine that these party lines became very busy during major failures such as might occur during a big storm. As systems grew, it was common practice to break a long route into sections, each served by a local order wire, while an express order wire ran the entire length of the route, with appearances only at major maintenance-control points.

Order wires were provided by one means or another along all long-transmission systems. With paired cables or coaxials, pairs in the same sheath were generally used.[5] Radio routes presented special problems, and order wires were often established over completely separate facilities. It was highly desirable, however, to have an order wire provided with the same equipment that was used for the circuits in service. In some of the short-haul radio systems where the baseband signal appeared at every repeater, methods were devised to have the order-wire signal ride along with the main signal. Unfortunately, many kinds of failure that caused the loss of the main signal would wipe out the order wire as well.

3.2 Service Protection and Restoration

Reliability and continuity of service have been a tradition and matter of pride among telephone people since the founding of the business, and this tradition was always a major force in the design of transmission systems. The first repeatered lines used vacuum tubes of extraordinary reliability and life; many of them operated continuously for decades in an era when a lifetime of a few months was considered normal. When the 12-channel Types J and K systems became available in the late 1930s and the Type L1 coaxial system with hundreds of channels in 1941, elaborate precautions were taken so that most foreseeable failures would not cause a service interruption at all.

In the J and K systems, vacuum tubes were paralleled in each repeater stage. Failure of either tube would not interrupt transmission, and a defective or marginal tube could be replaced without affecting service. In addition, patching arrangements were provided at both repeaters and terminals so that working and spare equipment units could be brought to a common point and the substitution made by a single key operation. When the working item had been serviced or replaced, transmission was restored to the normal path with a barely perceptible hit or click.[6] However, patching and manual switching were too slow for the higher-capacity transmission systems coming into service. The L1 coaxial system provided a working spare line that was automatically switched into service when trouble occurred on a regular line.

On the first L1 lines, there were one working and one spare coaxial unit in each direction. The spare carried the signal as well as the working line, and, when monitoring equipment sensed the loss of the main pilot or a drop in its level below a preset value, the switch to the spare was actuated in a few milliseconds. In later coaxial system designs and on systems with multiple coaxials, a single coaxial unit continued to protect the working lines. Initiation of the switch was still by detection of abnormal pilot level on a working line at the receiving switch, but two-way communication over maintenance pairs and some fairly complex circuitry were necessary to ensure that the correct working coaxial was switched to the protection line and to make a near-simultaneous switch at the transmitting and receiving ends.[7,8]

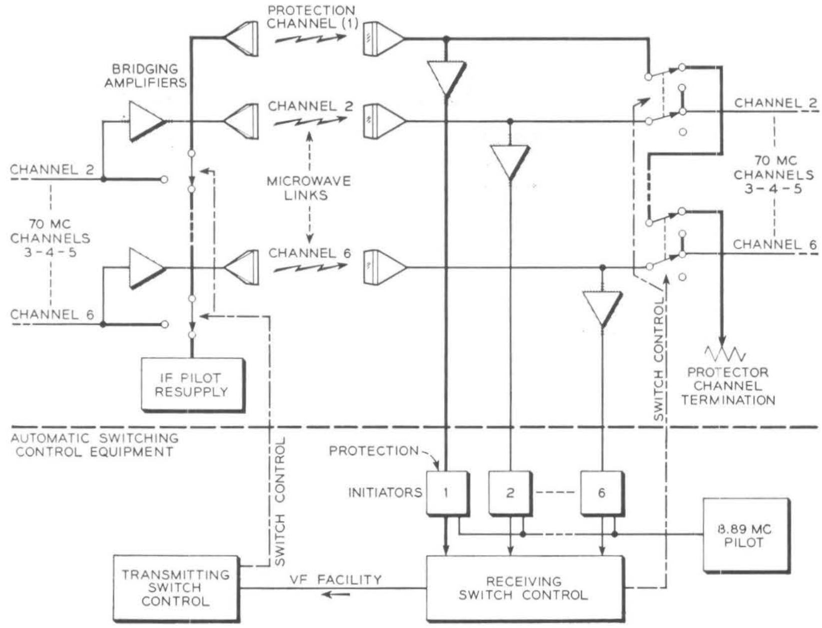
Ultimately, as many as ten working coaxials pairs in a 22-coaxial sheath were protected by these means in the L4 and L5 systems. The system was elaborated in several ways. In the 1800-channel L3 system (1953), the switch action was fast (15 milliseconds) when all pilots were lost, but slowed to 50 to 100 milliseconds if only one pilot faltered or for slow deterioration. The object was to prevent needless switches on transient hits. The delayed action was also built into other systems, such as the T2 digital line (1972), to prevent switches on transients from lightning surges.[9,10,11]

A few years after the first microwave systems went into service, automatic protection-switching systems were also provided for them as standard equipment. Protection switching in radio systems, however, not only protected a service in the event of equipment failures but greatly improved the continuity of transmission during periods of selective fading. In fact, in many areas, transmission continuity would not have been satisfactory without the improvement provided by the fast switch to a protection channel whenever a working channel suffered a deep fade. Radio switching sections could be as long as ten hops of 25 miles each, but were often much shorter because of route branches or the location of terminals to serve cities.[12,13]

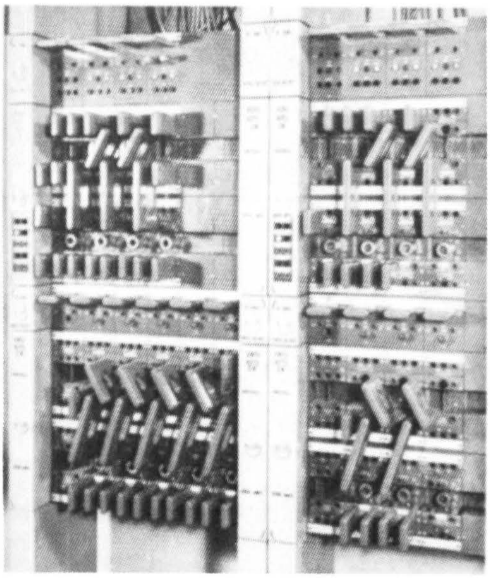
Radio protection-switching systems had many special features compared to cable systems. For example, the FM systems did not change their baseband to baseband loss with fades. The switch initiators instead monitored the signal-to-noise ratio at a frequency above the message spectrum and operated when it deteriorated to a preselected value. Tones above the signal band were transmitted over the protection channel to indicate whether it was available or in service [Fig. 23-3]. In the TD2 radio system the communications required between the ends to coordinate the switch to the protection channel of a failed or excessively faded channel were over land wire circuits. In the first version of the 6-GHz TH radio system, concern for failures and fading led to the provision of two protection channels for six working channels. Communications to control the switch, as well as to transmit voice and alarms, were over two narrow-band auxiliary radio channels at opposite ends of the 5925 to 6425 MHz band.

The first TD2 protection system, put in service in 1955, provided one protection channel for the remaining five radio channels in the 4-GHz band. When six additional interstitial channels became possible about 1960, one of them was also devoted to protection in a separate one-for-five switching system. In the 1960s, as the radio spectrum became more crowded, the Federal Communications Commission pressed to reduce the number of channels used for protection. The Bell protection arrangements were accordingly modified to provide one-for-eleven protection at 4 GHz and one-for-seven protection at 6 GHz. In the early 1970s, as TD2 and TH systems continued to expand and frequently occupied the same buildings and antennas along the same routes, a new protection system, the 400A, was developed. In the 400A, two 30-MHz channels in the 6-GHz band, widely spaced to minimize simultaneous fading, protected the remaining 18 working channels in both bands. Provision of two protection channels, accessible to any of the 18 working channels in a single switching complex, improved transmission continuity and reduced the equipment required compared to separate systems in each band.

All of the protection switching systems allowed maintenance personnel to initiate a switch manually. Most protection channels were on a first-come first-served basis, but the L5 coaxial line protection-switching system of the early 1970s also had a priority feature. This allowed operators to designate one coaxial pair that, in cases of a multiple coaxial-unit failure, was to be the one restored by the protection line. The most vital services could then be routed over the priority line. By the mid-1970s, automatic protection lines not only protected the transmission system they were on, but also could be used to restore service of large blocks of channels, such as mastergroups, from other failed systems through alternate routing. The protection switches could be controlled and the channels connected in tandem from a central location through telemetry and control systems.



(a)



(b)

Fig. 23-3. Microwave radio protection switching. (a) TD2 radio one-for-five protection switching. (b) TD2 protection switching equipment showing coaxial switches, bridging amplifiers, and coaxial patching plugs.

3.3 Remote Alarm Monitoring and Control

With the start of service on the 12-channel Type K carrier system on paired cable in 1938, repeater spacing was reduced to about 17 miles. For the first time, unattended auxiliary stations housing considerable electronic equipment were put in service. Maintenance service when needed was to be from the nearest attended location. Routine maintenance at scheduled intervals was established to keep the equipment in proper adjustment, but unanticipated failures had to be considered. If, for example, a primary power failure required a station to switch to battery back-up, the entire transmission system would be out of service if the battery power was exhausted before maintenance personnel could restore the main power. Clearly, it was going to be helpful, indeed essential, to have as much information as possible about the location and nature of problems before sending people out to correct them. This need led to the provision of remote alarm-monitoring systems for all transmission systems with unattended equipment.

In the remote alarm system developed for Type K carrier, a failure at an unattended site was made to operate or release a relay, whichever was convenient, depending on the nature of the circuit monitored. This placed a ground on one contact of a ten-position rotary switch. The switch at the unattended site rotated to this position and signaled to an identical switch at an attended location, causing it to move to the same contact. Troubles and alarms were classified as major or minor and distinctive lamp colors assigned to the switch positions. It was possible to release and recycle the switches from the attended sites to check whether a momentary trouble had disappeared or whether maintenance activity had cleared it. Operation was over No. 19 gauge cable pairs up to a range of 60 miles. The system was applicable to open-wire Type J carrier over greater ranges or even to repeatered voice lines, which were also beginning to have unattended stations.[14]

In the L1 coaxial system auxiliary repeaters were needed every eight miles and power-feed and switching main stations every 100 miles or so. Even main stations could be unattended at least part of the time. Full-time maintenance people could be as much as 200 miles from a trouble location. A new and more elaborate alarm and control system was designed to assist maintenance under these conditions. A trouble at an auxiliary repeater would cause an alarm indication at the nearest main station. A new B1 alarm system at the attended maintentance center could monitor as many as 168 alarm conditions in each of ten unattended main stations. The presence of an alarm at any station was signaled to the maintenance center over a circuit that was immediately released to continue its surveillance of the other stations. The maintenance station could then initiate scans and receive detailed information over separate circuits that could be addressed to the station with the trouble. The pattern of alarms, which could be quite complex for some types of system problems, was registered in a matrix of neon lamps. A log sheet could be placed over the lamps and marked for analysis and a record [Fig. 23-4].

The control station, using the same system, could also address commands to a station if conditions required. The commands activated a selected combination from among 100 order leads at the remote station to accomplish such actions as firing up a standby gas engine for power, or forcing a switch to the

Fig. 23-4. Recording troubles indicated by lamps under a log sheet at an L1 coaxial system maintenance center, 1950.

spare of a selected line.[15] A somewhat similar system, the C1 alarm and control system, adapted to the special conditions of microwave radio, was developed in 1952.[16] The communications link for the radio-system alarms was over wire pairs similar to those used to control protection switching. The TH radio system utilized the narrow auxiliary channels in the microwave band for alarm and control as well as for automatic protection switching.

The later coaxial systems (L4 and L5) required adjustable equalizers distributed along the line at unattended points. The alarm and control systems developed for these systems permitted remote adjustment of equalizers and the measurement of amplifier gain in addition to the usual alarm indication and fault-location procedures.[17,18] (See Chapter 10 for details of these functions.)

3.3.1 E Telemetry

All the maintenance systems had characteristics in common: they required sensors at remote points to measure quantities or to register troubles, they had to transport data to a central control, and they had to transmit control commands from the attended center to the remote point. They were, in fact, telemetry and control systems, similar in many respects to types that were becoming common outside of telecommunications in the space and missile age. A group of Bell Laboratories designers in New Jersey had a crash course in telemetry in 1961 and 1962 while devising the elaborate system used in the Telstar satellite project. Building on this experience in the later 1960s, they designed a standard system, E telemetry, for use in transmission maintenance. With E telemetry, transmission between remote points and control centers was over voice-frequency private lines or by dialed-up DDD lines.

Initially, the remote equipment for E telemetry was developed in modular form for alarm and control use on long-haul transmission systems, such as L4 coaxial and TD3 radio. Since the remote units could be used in several other applications, engineers were soon using the options available to custom-engineer field installations for a variety of purposes. After studying the use of telemetry remote terminals in the field, however, the designers standardized arrangements for a range of uses. The savings in time and costs led to the use of E telemetry for almost any application calling for centralized monitoring and maintenance of remote locations. By 1975, E telemetry was used to monitor cable gas pressure indicators, gather traffic data, and help to maintain electronic switches, as well as for transmission maintenance. It played an important role in the growing trend to centralized and automated maintenance of all types of plant facilities [Fig. 23-5].[19]

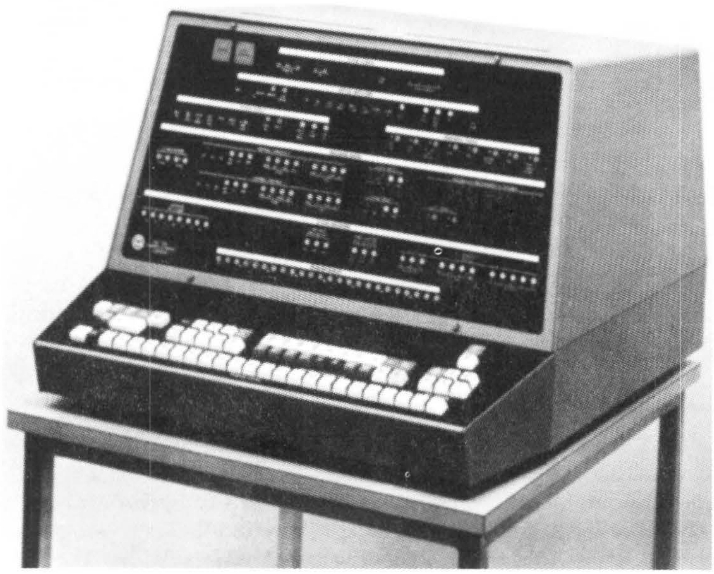


Fig. 23-5. Test console incorporating an E telemetry module for the maintenance of 1ESS switches from a central location, 1974.

IV. TRUNK TESTING

4.1 Local Trunks

To replace the manual operators' transmission test that was lost as offices were converted to dial operation, a number of trunk test arrangements were developed for successive generations of automatic switching. In the panel system used in large metropolitan areas in the 1920s and 1930s and later, the switch made a so-called trunk-guard test. If the test failed, the panel sender "stuck," and an operator was required to release the connection. In the process, the defective trunk was identified. When crossbar switching was introduced in the late 1930s, more elaborate tests and diagnoses became possible. Crossbar switches, like earlier ones, could be manually stepped through a trunk group. When a defective trunk was encountered, they provided not only its identity but information on the nature of the problem as well. Some of these procedures were automated for routine testing shortly after the introduction of dial switching, but the information provided on the transmission properties of trunks was very limited. In general, passing the test merely indicated that the switching control signals had been transmitted and received in a form adequate for the operation of the switch.

By 1954, automatically progressing trunk-test circuits were designed for the No. 5 crossbar switch. These test circuits accessed trunks through the regular switch train instead of through a master test frame. The test circuit was controlled by a perforated teletypewriter tape and a standard teletypewriter reader that selected the trunks to be tested and directed which tests were to be made.[20] The system could also call up a test line in a distant office to test trunk functions in the incoming direction. The teletypewriter recorded results or exceptions as they occurred. When the 1ESS* electronic switch was introduced in the mid-1960s, the same types of tests were performed under the instruction of the switch's stored program.

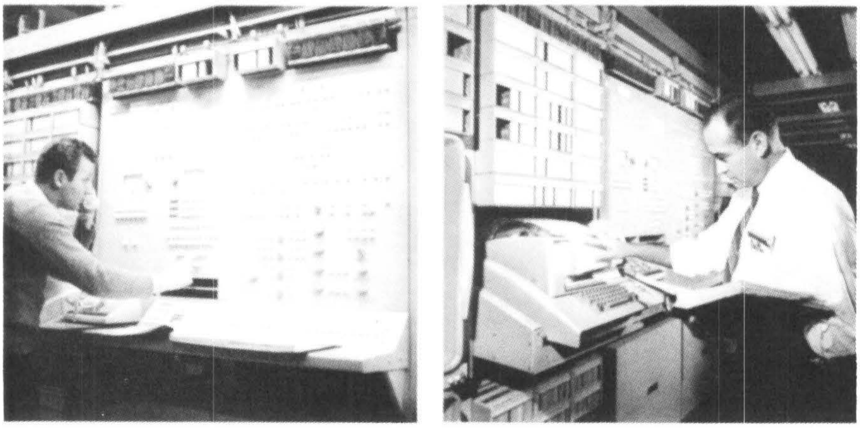
The 1ESS switch automatically tested its own trunk seizure and control equipment each time a call was dialed into or out of an office. In addition, on a scheduled basis, interoffice trunks were tested about twice a day and customer loops about every other day, with the problems found recorded by a teletypewriter [Fig. 23-6].[21] The 1ESS switch routinely checked insulation resistance, and trunks could be made available to test positions for manual tests of loss, but, in general, the routine tests performed by the local switches were functional rather than transmission tests.

4.2 Toll-Trunk Tests

In contrast to local trunks, which were short and usually passive, toll trunks required much more routine maintenance. They could be established over a variety of different facilities in tandem, with scores or even hundreds of amplifiers and equalizers, and might traverse several multiplex terminals between switches. For many years toll trunks were tested manually, one at a time, with test tone generators and meters. To save eyestrain, toll test boards were equipped

---

* Trademark of AT&T Technologies, Inc.



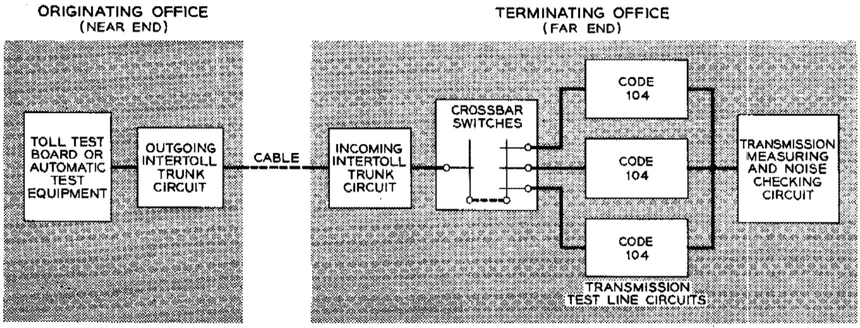
(a) (b)

Fig. 23-6. Local trunk testing in the 1ESS electronic switch. (a) The 1ESS test position. Routine automatic trunk tests were conducted every day, but little more than functional tests were normally made. (b) Results or exceptions were printed out by a teletypewriter.

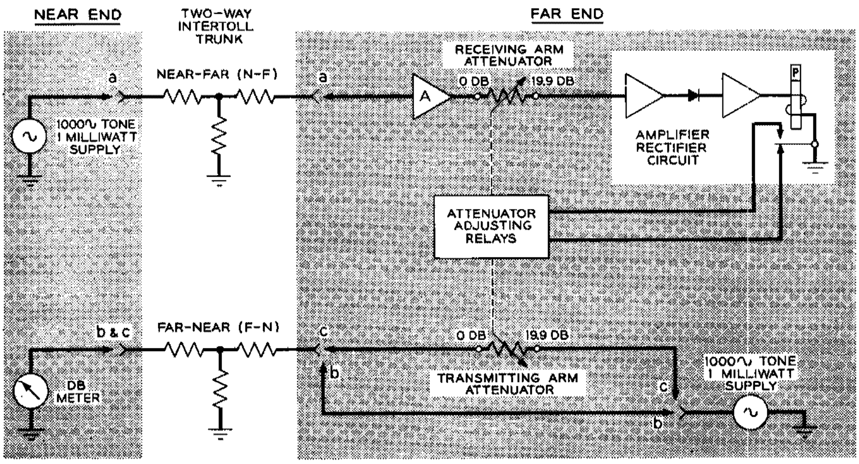
with galvanometers that projected their scales and pointers optically on large screens near the top of the bays in the test aisles. (These were affectionately (?) known to a generation of testers as the "Mickey Mouse.")

In 1956, a few years after customer toll dialing was introduced, a two-way intertoll trunk transmission-measuring system was developed, the automatic transmission test and control circuit. (A similar system was developed at about the same time by the L. M. Ericsson Company in Sweden.) Trunks were seized for test, either automatically through an outgoing trunk test circuit or manually at a toll test board. The system provided for two-way functional tests and improved loss and noise tests from one end. Although designed primarily for No. 4 crossbar switches, the far-end test lines and equipment worked equally well in No. 5 crossbar and step-by-step offices [Fig. 23-7].[22]

The system operated by dialing a test code (the 104 code) to associate the far end with a transmission-measuring circuit. With 1 mw transmitted from the near end, an attenuator following a 20-dB amplifier at the far end was automatically adjusted until the output to a detection circuit was also 1 mw. At the same time, an attenuator in the return side was adjusted to the same setting. In the next step, the far end first transmitted a 1-mw signal directly to the line, permitting a measurement of far-to-near-end loss, followed by transmission of the 1-mw tone through the previously set return-side attenuator. The change in level between the last two received signals permitted the near-end tester to determine the outgoing side loss. In a final step, high-gain amplifiers were placed in both ends for a noise check. Near-end noise could be read directly. The far-end amplified noise output was rectified and used to charge a capacitor for exactly five seconds. The voltage on the capacitor was then compared to a reference voltage and, if the reference was exceeded, made to



(a)

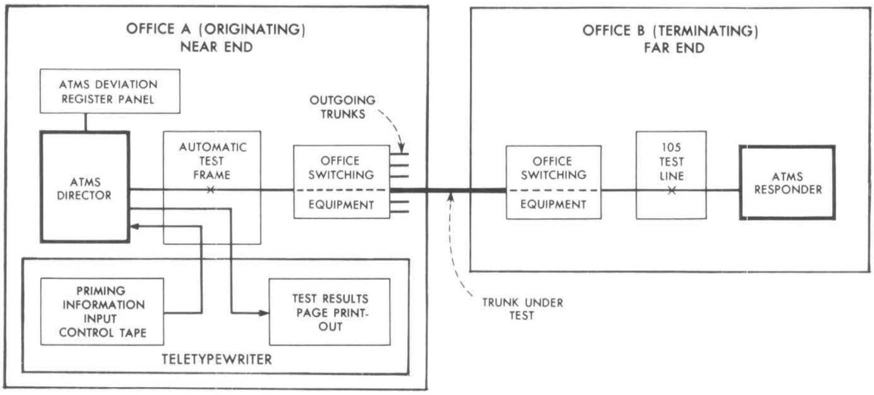


(b)

Fig. 23-7. The automatic transmission test circuit, 1956. (a) The test configuration was established by a dialed-up connection (code 104) to a far-end loss and noise checking circuit. (b) Two-way loss measurements were possible from one end by controlling the far-end signal source and an adjustable attenuator in the return side.

operate a relay. The relay operation could be detected at the near end, providing a simple go/no-go check of far-end noise. (More accurate noise measurements were provided in later designs.)

With this equipment, about 30 trunks per hour could be tested manually and 120 per hour with automatic equipment. Within a few years, the automatic testing equipment was associated with a simple computer to count the number of trunks in each loss range and a printer to record the losses and the mean and standard deviation of the distribution.[23]



(a)

(b)

Fig. 23-8. The automatic transmission-measuring system (ATMS), 1967. (a) ATMS seized trunks through the office switching equipment. Tests were programmed on teletypewriter punched tapes. (b) Loading punched tape for ATMS tests. Equipment in background is ATMS Director.

4.2.1 Automatic Trunk Measurements—ATMS and CAROT

The late 1950s and 1960s were a period of phenomenal growth in the long-haul plant, with circuit mileage more than quadrupling from 1955 to 1965. The trunk-measuring equipment developed in the early 1950s became inadequate for the expanded network. In response, in the late 1960s, a faster and more accurate system, the automatic transmission-measuring system (ATMS), was developed.

ATMS differed from the earlier system in several respects. The trunk to be tested was seized and held through the regular office switching equipment under control of an automatic test frame at the near end and a test line at the far end. The test frame was programmed by a punched paper tape and tape reader [Fig. 23-8]. The test frame also primed measuring equipment in an ATMS director at the near end and in a responder at the far end with information on conditions for the test, such as terminating impedances and the insertion of attenuator pads. Communications between the director and responder were by two out of six multifrequency tones in the outgoing direction and by a 2200-Hz data signal in the reverse direction. Speed of operation was improved by the faster communications and by a quasi-rms measurement of noise. In some of the earlier test arrangements, rms-noise power was measured with thermal-type indicators (thermocouples). These were very accurate but very slow, taking ten seconds or more to settle. The quasi-rms indicator, by contrast, gave a reading of adequate accuracy in less than 0.4 second. With these changes, loss and noise in both directions could be measured in less than five seconds, with the results printed out on a teletypewriter. ATMS was tested in the field in Norristown, Pennsylvania in 1965 and 1966 and was in production by 1968.[24]

Even while ATMS was in the final design stages, however, further developments in trunk testing were under way as part of a widespread move toward greater centralization and automation of plant-maintenance activities. ATMS could be used to obtain trunk measurements between two offices, both distant from the controlling office, by means of additional equipment called a remote-office test line (ROTL). For example, a ROTL in office B was dialed up from office A and made to act as a controller for tests between offices B and C [Fig. 23-9]. Operations were still controlled by punched tape from a test frame in a central office, however, and were slow and cumbersome compared to what could be done with the new minicomputers.

Around 1967, development started on a new system for centralized automatic reporting on trunks (CAROT). CAROT was designed to provide remote testing of trunks throughout a geographical area from a centralized location, not necessarily a central office. The area served could contain up to 100,000 trunks. A single controller, for example, might test trunks throughout the San Francisco Bay area, but one CAROT system might also cover one or more states in less populated regions [Fig. 23-10]. The plan with CAROT was to provide relatively inexpensive ROTLs for all types of central offices, with the function of the test frames and directors gathered at a central location in a CAROT controller [Fig. 23-11]. The heart of the controller was a minicomputer that stored data, such as test limits and the trunks to be tested; it also programmed the tests and

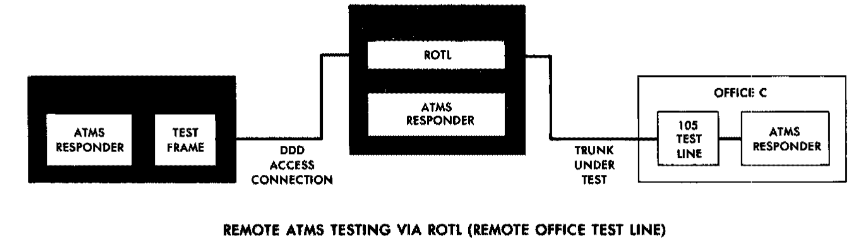


Fig. 23-9. The use of ATMS and a remote-office test line (ROTL) to obtain measurements on a trunk between two remote offices.

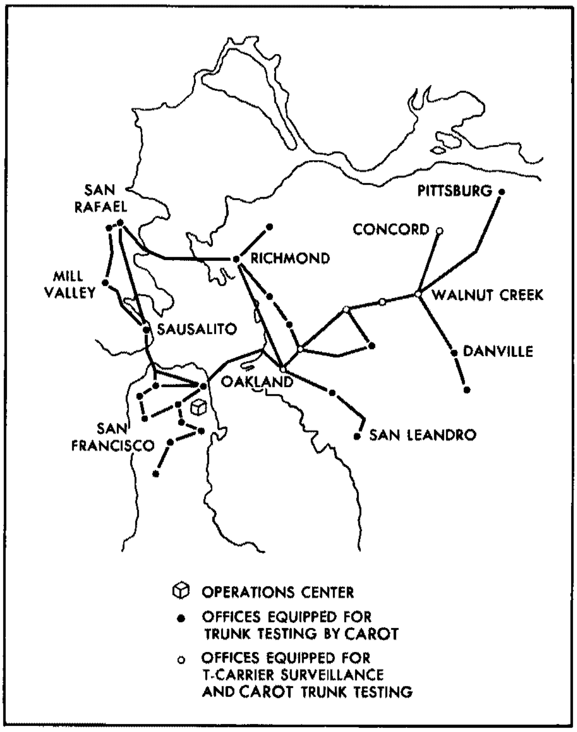
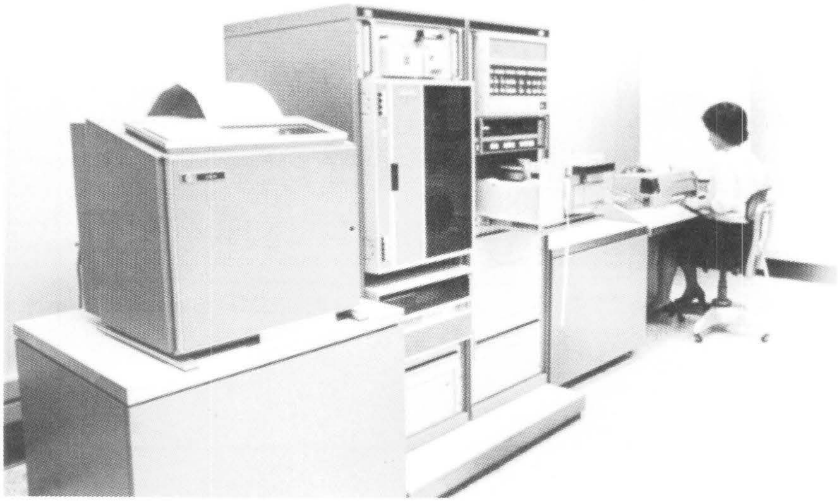
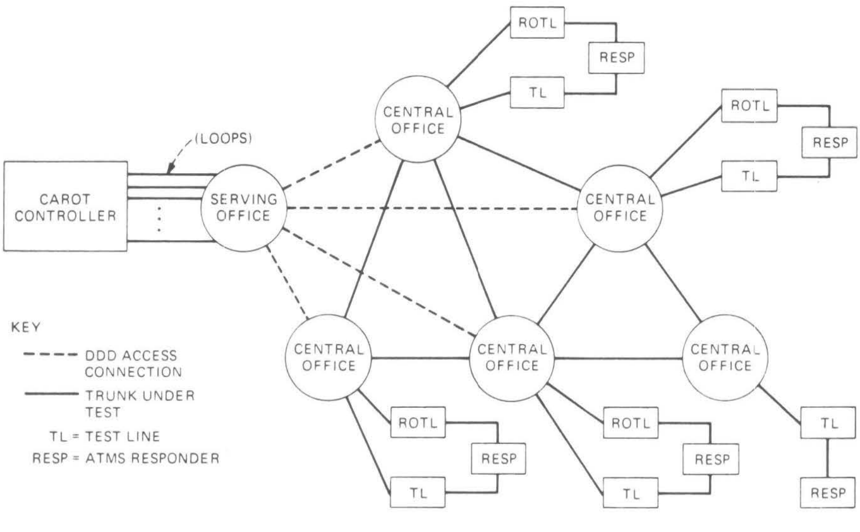


Fig. 23-10. Centralized automatic reporting on trunks (CAROT). In the 1972 CAROT trial in the San Francisco Bay area, 12,000 trunks were tested from a single operations center. The trial was conducted in conjunction with a field test of the T-carrier surveillance system.



(a)



(b)

Fig. 23-11. Area-wide trunk tests and measurements by CAROT. (a) The CAROT control center. The minicomputer-controlled center could perform tests on as many as 100,000 trunks over a wide area. Multiple access ports permitted testing up to 14 trunks simultaneously. The nightly test sequence was initiated from the keyboard, but testing was entirely automatic. (b) CAROT tested trunks between remote offices via dedicated loops and by dialed up connections to test lines (ROTLs) and responders.

stored and processed the results. Faulty trunks were identified and the nature of the problem recorded. In addition, average performance and other reports were automatically printed out for administrative purposes.

Access to the network was via controller ports and subscriber loops over which dialed-up connections to the selected offices were made. CAROT could conduct measurements on up to 14 distant trunks simultaneously. Programs written for a particular area directed the controller to select trunks and store their access and test-limit information. The software also established a schedule so that all trunks were tested at correct intervals. At the selected time, the controller transmitted information to the dialed-up ROTL indicating the trunk to be seized, the near-end equipment to be used, and on which far-end test line to terminate.

A field trial of CAROT was held in late 1972 in the San Francisco Bay area. The trial showed the ability of the system to improve trunk-maintenance and reduce costs but also revealed the need for several improvements. The trunk-maintenance data base was first stored on paper tape, but this was changed to magnetic-tape cassettes and later to disk files. Other improvements included expanding the repertoire of tests that could be performed by responders and extending the ability to perform tests via the controller to maintenance people at distant central offices.

CAROT quickly won acceptance. In a typical pattern, the system was used for demand tests and updated its files during the day. During the night, a CAROT center could test up to 14,000 trunks between a 100 or more offices, process the results, and forward reports on faulty trunks to the appropriate central office for action early in the morning before the busy hour traffic. By 1978, more than 30 percent of all Bell System trunks were being tested by CAROT with about 60,000 additional trunks being added every month.[25,26]

4.2.2 Testing Special-Service—SARTS and CMS3

CAROT was extremely effective in improving the maintenance of trunks in the public switched networks, but there were many other circuits requiring maintenance, such as toll-free wide-area telephone service (WATS) lines for clients calling a business from distant cities (800 numbers), and data links from a central computer to several remote locations. The problem was not trivial. In 1974, for the first time, special-service circuits accounted for more than half of the channels added between offices. These circuits presented special maintenance problems because they were often custom designed and, being outside the switched network, were difficult to access. During the early 1970s, considerable development effort was directed to the problem at Bell Laboratories, but the full story is beyond the scope of this history.

In brief, the approach was multipronged. Working with AT&T and several local telephone companies, Bell Laboratories developed the concept of a special services center to centralize administration and testing. Two minicomputer-based operations-support systems were designed to support this center. The switched-access remote test system (SARTS) provided low-cost switched maintenance access (SMAS) to the special-service circuits and allowed a single maintenance tester to test any point of a long and complex circuit from a central location [Fig. 23-12]. Complementing SARTS, the circuit maintenance system, CMS3, was used for the administration of circuit and service orders, circuit layout records, and trouble tickets for special services. At the same time, design methods for new circuits were simplified and standardized. The special services center concept, SARTS, and CMS3 were tested in San Diego, New York City, and Atlanta in the late 1970s and were a major component of the program to reduce maintenance expense undertaken in those years and into the 1980s.

4.3 Operations Systems Coordination

CAROT and the special-services transmission-maintenance systems were only a small part of a large family of computer-based systems developed in the 1960s and 1970s to assist telephone company operations and maintenance. In a series of studies starting in 1974, the coordination of computer-based operations-support systems evolved by 1978 to a Total Network Operations Plan (TNOP—Issue 1).[27] The plan included all the transmission-support systems, along with dozens of others. Parallel plans were developed for business and residence customer operations, reflecting the form of AT&T organization at that time. In all, over 100 systems were developed to assist in almost every aspect of day-to-day operation of the business.

In addition to the trunk maintenance systems discussed above, transmission-support systems were developed for loops, carrier terminals, cable pressure

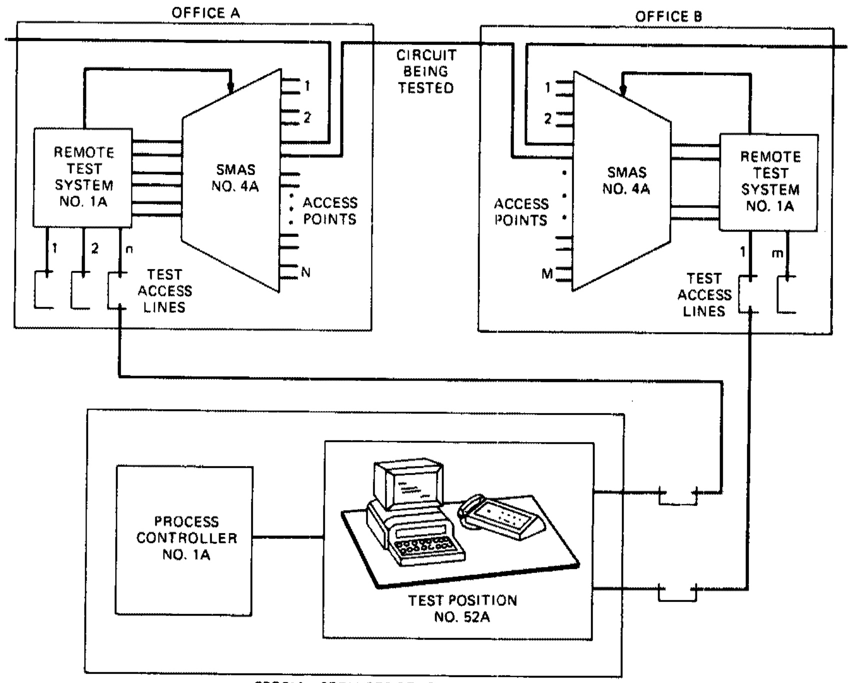


Fig. 23-12. The switched access remote test system (SARTS), 1978.

monitoring, facility surveillance and control, and T carrier administration.[28] Several of these systems had overlapping data bases and had to communicate and exchange large amounts of data with other support systems. The situation can be exemplified by the complex of transmission-related systems surrounding the electronic toll 4ESS switch [Fig. 23-13].[29]

A circuit-maintenance system (CMS1A) (not to be confused with CMS3, which was for special services only) functioned as the hub of transmission-related activities with a file of over 300 million bytes of information stored on disk. It provided a data-coordination function—updating files and storing test data for CAROT and furnishing data to a system to support multiplex terminal maintenance, the carrier terminal-maintenance system. CMS1A also provided

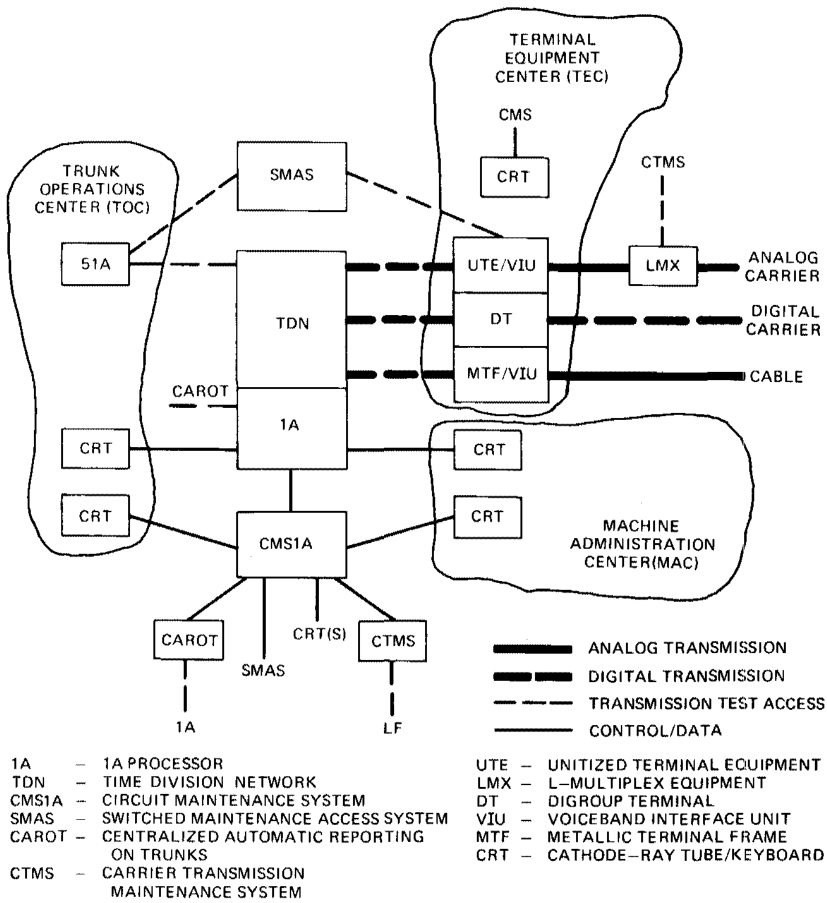


Fig. 23-13. The complex of computer-based support systems associated with the 4ESS toll switch, 1977. CAROT provided the means to test ordinary switched trunks and SMAS provided test access to special service lines. CMS1A furnished data and acted as a coordination hub for all transmission related activities.

the interface between these and trunk-access and maintenance software in the switch's 1A central processor.

Manual testing of trunks was possible from a trunk operations center (TOC) by means of a 51A test position. A CMS1A terminal at the 51A desk was the major input/output device during trunk testing [Fig. 23-14]. The 51A also had access, by SMAS switches, to analog special service lines that were not accessible via the time-division switching network. A terminal equipment center (TEC) was responsible for the maintenance of terminal equipment for voice-frequency, digital, and analog carrier lines terminating on the switch. In addition to a stationary CMS1A terminal with a cathode-ray-tube screen and printer, the TEC had a number of mobile units that permitted personnel to communicate with CMS1A while engaged in maintenance activities in the equipment aisles [Fig. 23-15].

Routine operational and transmission tests on trunks were performed by CAROT. A ROTL (but not so remote in this case) was connected to CAROT and the switching network. Working through the switch's 1A central processor, the ROTL could be associated with any trunk for tests by CAROT. Upon finding a faulty trunk, CAROT could request that it be removed from service by the 1A processor. The processor would turn down the trunk and notify CMS1A of the test result. CMS1A, in turn, would create a trouble ticket and add it to the work list at the TOC. If, upon localization, the trouble proved to be in the local terminal, the TOC would refer the trouble to the TEC. Personnel in the

Fig. 23-14. The trunk operations center for the 4ESS switch showing CMS1A CRT terminals at 51A test positions.

Fig. 23-15. A mobile CMS1A terminal in a transmission terminal equipment aisle of 4ESS switches.

TEC would then locate and repair the defective equipment, utilizing CMS1A for access and test information. They could then use CMS1A to verify that the problem had been cleared, to put the trunk back in service, and to clear the trouble ticket. The result was to create specialized, autonomous, but mutually supporting, systems.

REFERENCES

1. V. A. Vyssotsky, "Computers and Computer-Based Systems," *Bell Lab. Rec.* **53** (January 1975), pp. 90-95.
2. M. M. Buchner, Jr., "Planning for the Evolving Family of Operations Systems," *Bell Lab. Rec.* **57** (May 1979), pp. 118-124.
3. A. J. Pascarella, "Locating Faults on Toll Lines," *Bell Lab. Rec.* **7** (December 1928), pp. 161-164.

4. W. V. K. Large, "Telephone Order Wires for Toll Circuit Maintenance," *Bell Lab. Rec.* **9** (February 1931), pp. 283-287.
5. A. V. Wurmser, "Telephone Order-Wire Circuit for Type-K Carrier Routes," *Bell Lab. Rec.* **24** (May 1946), pp. 185-187.
6. H. A. Wenk, "Switching Devices for Toll System Maintenance," *Bell Lab. Rec.* **20** (November 1941), pp. 78-82.
7. K. C. Black, "Stevens Point and Minneapolis Linked by Coaxial System," *Bell Lab. Rec.* **20** (January 1942), pp. 127-132.
8. E. C. Thompson, "Automatic Line-Switching for L3 Carrier," *Bell Lab. Rec.* **36** (September 1958), pp. 340-343.
9. F. H. Blecher, R. C. Boyd, F. J. Hallenbeck, and F. J. Herr, "The L-4 Coaxial System," *Bell Syst. Tech. J.* **48** (April 1969), pp. 821-839.
10. J. H. Green and R. W. Sanders, "L5 System: Line-Protection Switching," *Bell Syst. Tech. J.* **53** (December 1974), pp. 2011-2034.
11. J. H. Bobsin and L. E. Forman, "The T2 Digital Line—Extending the Digital Network," *Bell Lab. Rec.* **51** (September 1973), pp. 238-243.
12. B. C. Bellows, Jr., "Automatic Protection Switching for TD-2 Radio," *Bell Lab. Rec.* **34** (March 1956), pp. 98-102.
13. R. H. Higgins, L. M. Klenk, and W. R. McClelland, "Automatic Protection Switching For The TH-System," *Bell Lab. Rec.* **40** (May 1962), pp. 172-177.
14. V. E. Rosene, "Alarm System for Auxiliary Repeater Stations," *Bell Lab. Rec.* **18** (July 1940), pp. 345-350.
15. L. A. Weber, "Coaxial's New Alarm and Control System," *Bell Lab. Rec.* **28** (March 1950), pp. 122-125.
16. G. A. Pullis, "Alarm and Control Features of the TD-2," *Bell Lab. Rec.* **30** (December 1952), pp. 477-481.
17. J. W. Mulcahy, W. P. Frawley, R. E. Maurer, and F. C. Kelcourse, "The L-4 Equalizing and Fault Locating System," *Bell Lab. Rec.* **45** (July/August 1967), pp. 230-236.
18. W. G. Scheerer, "L5 Main Stations Pilot the System," *Bell Lab. Rec.* **51** (November 1973), pp. 312-322.
19. R. J. Sanferrare, "E-Telemetry: Inside Story of Centralized Maintenance," *Bell Lab. Rec.* **52** (December 1974), pp. 345-350.
20. H. D. Cahill, "Automatic Progression Trunk-Test Circuit," *Bell Lab. Rec.* **32** (August 1954), pp. 313-318.
21. A. I. Besman and S. Bloom, "Trunk and Line Testing for No. 1ESS," *Bell Lab. Rec.* **45** (December 1967), pp. 369-372.
22. H. H. Felder, A. J. Pascarella, and H. F. Shoffstall, "Automatic Testing of Transmission and Operational Functions of Intertoll Trunks," *Bell Syst. Tech. J.* **35** (July 1956), pp. 927-954.
23. R. C. Nance, "Automatic Calculation of Transmission Deviations," *Bell Lab. Rec.* **35** (December 1957), pp. 503-507.
24. J. F. Ingle, J. J. Kokinda, and G. E. McLaughlin, "An Automatic Transmission Measuring System for Telephone Trunks," *Bell Syst. Tech. J.* **46** (November 1967), pp. 1935-1975.
25. M. M. Anderson, D. V. Helsing, and A. Mizrahi, "CAROT for Automatic Trunk Testing," *Bell Lab. Rec.* **50** (November 1972), pp 322-329.
26. M. M. Anderson, M. B. Einschlag, and E. G. Jaasma, "CAROT Revisited," *Bell Lab. Rec.* **56** (March 1978), pp. 71-75.
27. See reference 2.
28. Bell Laboratories Members of Technical Staff and Technical Publication Department, *Engineering and Operations in the Bell System*, 1st. Ed. (Bell Telephone Laboratories, 1977), pp. 563-602.
29. J. A. Giunta, S. F. Heath, III, J. T. Raleigh, and M. T. Smith, Jr., "No. 4ESS: Data/Trunk Administration and Maintenance," *Bell Syst. Tech. J.* **56** (September 1977), pp. 1203-1237.

Chapter 24

Economics

I. INTRODUCTION

The concept of universal service was first enunicated by Alexander Graham Bell and elaborated in the early 1900s into a company policy by Theodore N. Vail. One consequence was the effort to extend the network to all points and to make it possible to talk over circuits up to thousands of miles in length. Another consequence was a set of economic and operating policies: (1) The cost of entry was to be kept low. Anyone who wanted a telephone was to have one at a low cost and with minimum delay. (2) Local calling was to be generally flat rate, i.e., unlimited calling for a fixed monthly fee. (3) Long-distance calling rates were to be averaged. Tolls were based on airline mileage and were not related to the cost of the specific facilities required. (4) Connections were to be made when requested with a minimum delay. This meant the provision of large trunk groups so that the traffic could be served without queuing, even in the busiest hour. This was essential for customer direct distance dialing. (5) A corollary policy was that continuity of service would be assured. The lines were to be kept working even, for example, in the event of commercial power failure. Implementing this policy meant designing equipment to the highest standards of reliability and providing spare or standby facilities with means to switch to them in the event of a failure.

These policies and practices were seldom questioned and became deeply rooted parts of the Bell System culture. In the early days, the economics of a technical development that removed a major limitation on what could be done were often self-evident. Increasing the number of trunks on a route by superposing a carrier system on an existing facility might cost only a fraction of the expense of duplicating the facility. Even so, the correspondence files and records are full of careful cost analyses at almost every stage. Even though the gain was obvious, there were always questions of alternatives and timing. In later years, when most fundamental limitations to the realization of universal service had been overcome, each new development had to meet strict economic tests before final design and production were started, and continuing economic evaluation accompanied every stage of the process [Fig. 24-1].[1,2] The sections that follow indicate only the major relationships important to the economics of transmission technology.

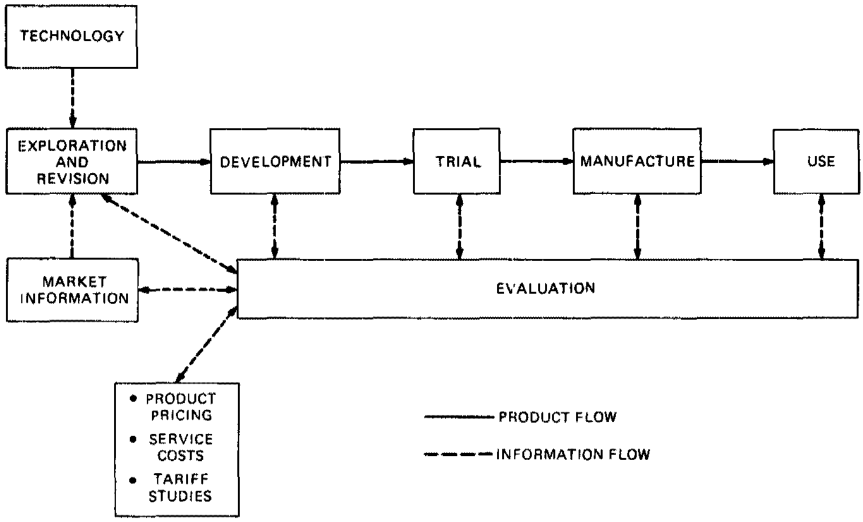


Fig. 24-1. The process of developing and manufacturing a new product involved economic analysis at every stage.

II. EXPANDING THE CIRCUIT CAPACITY OF EXISTING PLANT

Type C carrier, the first widely used carrier system, illustrated some of the basic relationships governing the economics of carrier systems. Carrier systems reduced per-channel line-haul costs by multiplying the number of channels on the medium. For a new facility, the average cost of a long circuit was lower because the cost of the copper wires, poles, and right-of-way were shared by a larger number of circuits. But the incremental cost of carrier circuits added to an existing line was lower still.

The opportunity to "mine" circuits out of existing plant, with the addition of relatively inexpensive electronics and only minor plant changes, was a major motivation for the development of most of the carrier systems. The 12-channel Types K and J carrier systems were applied to the same cable and open-wire pairs that had previously carried only a single voice circuit or, at most, a voice circuit plus three carrier channels. The successive generations of coaxial system designs were widely used to upgrade the capacity of earlier systems on existing cables, as well as for entirely new routes. The successively higher-capacity TD radio systems in most cases used the same roads, buildings, towers, and antennas installed for the earlier, more limited, designs. When the 6-GHz TH and single-sideband AR6A microwave radio systems were overbuilt on an existing 4-GHz TD route, it was necessary only to add the radio repeater bays and modify the antenna coupling hardware to add up to 42,000 circuits to the route capacity. Circuits gained by upgrading, with little change in the massive fixed plant, were by far the least expensive to be had.

Carrier systems, however, were not universally applicable. The cost of a voice-frequency (VF) line was almost directly proportional to its length for almost all distances (allowing for differences in terrain and route construction costs), and terminal costs for VF trunks were very low. Short VF trunks were very inexpensive. Carrier terminals, on the other hand, were relatively expensive, and carrier-derived circuits were not economical below a break-even distance where the savings in line costs equaled the cost per channel of the terminals [Fig. 24-2]. With Type C carrier, for example, the break-even distance in a typical case was not less than 250 or 300 miles.

Much of the history of transmission-system development from 1925 to 1975 consisted of a two-fold attack on costs: to transmit more and more channels per unit of line hardware and to lower the cost of terminals so that multichannel carrier systems could be economically applied at shorter and shorter distances. The increase in line capacity was enormously enhanced in the 1930s and 1940s by the introduction of coaxial cable and microwave radio. In many ways the terminal cost and short-haul problem proved a tougher nut to crack. This was frustrating, because there are a great many more short than long trunks, and carrier would have been a welcome alternative to running additional pairs in crowded urban conditions. In the 1960s, in the exchange area, more than 90 percent of the trunks were under ten miles in length and 50 percent less than five miles. It was estimated that reducing the break-even distance from ten to seven miles would more than triple the market for a short-haul carrier system.

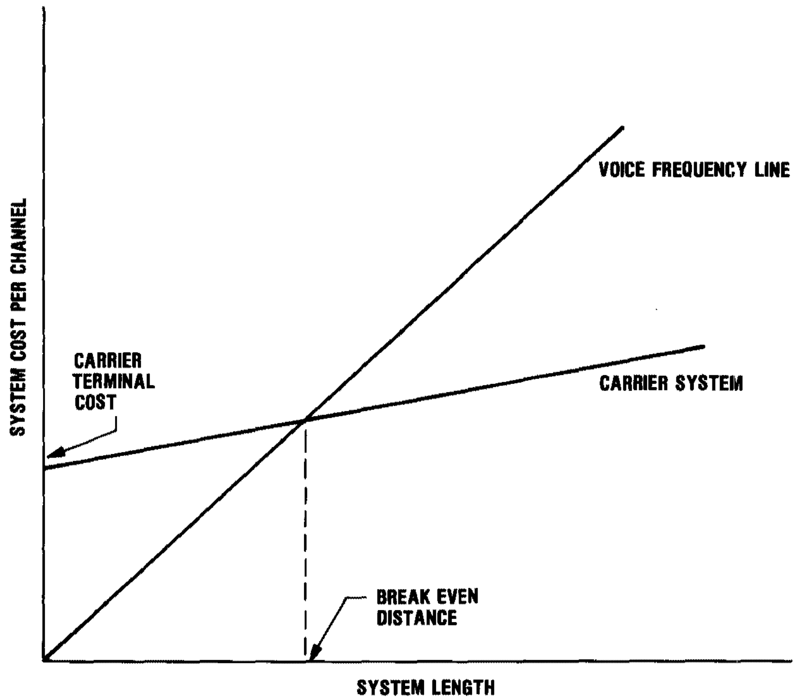


Fig. 24-2. Cost per channel versus system length.

III. ECONOMIES OF SCALE

Additional circuits obtained by multiplexing have little value if there is no market for them. Except for the years of the Great Depression, this was of little concern in the period from 1925 to 1975. In the late 1920s, the demand for circuits was backlogged, and reduced costs and rates stimulated it further. In the postwar years from 1945 to 1975, the demand for circuits and the plant capacity to meet it increased at an average compound rate of 14 or 15 percent per year. This led to several studies of circuit costs as a function of the capacity of transmission facilities. A study in the late 1940s, at a time when the early coaxial and radio systems were being widely installed, showed the striking results [Fig. 24-3]. A similar study in the early 1970s showed that the trend had continued [Fig. 24-4]. The unit cost in the expanded L5 (L5E) coaxial system, with a capacity of 132,000 channels in a 22-tube coaxial sheath, fell along the general-trend line, as did the projected costs for the two-level (WT4) and four-level (WT4A) millimeter-waveguide systems then under development. If the 6-GHz single-sideband AR6A AM radio system had been plotted at its maximum capacity of 42,000 channels, it would have been along the trend line of the radio systems. The trend of these figures, known as the Dixon-Clapp curve from two early authors, and the plot of each component should be taken only as broadly indicative of costs. Costs varied widely from time to time and from place to place. In addition, major factors such as the difference in costs between wholly new route construction and converting or overbuilding existing plant to mine circuits are not considered. Nevertheless, the broad trend is representative of actual cost differences.

The Dixon-Clapp curves show several things. First, and most striking, is the tremendous reduction in cost per circuit that had taken place, amounting to a

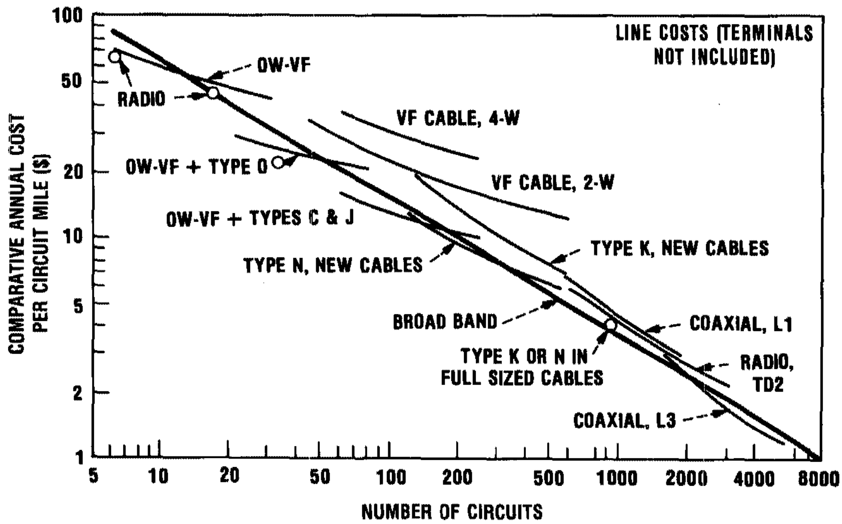


Fig. 24-3. Relative line costs versus system capacity from a 1950 study.

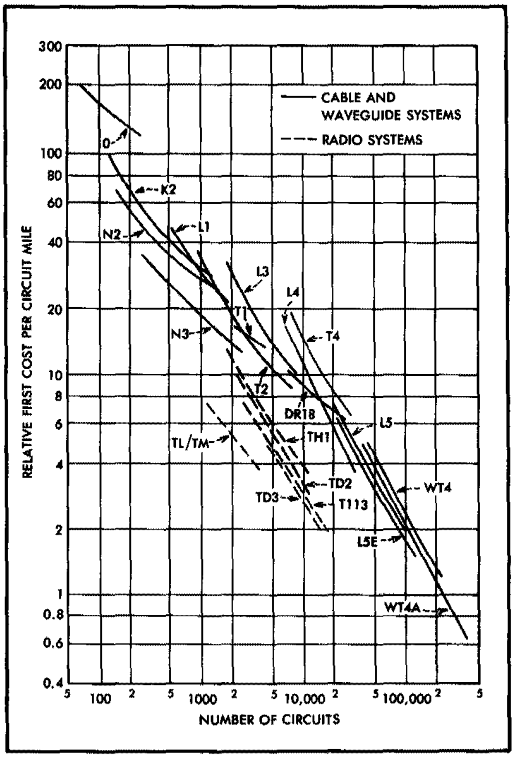


Fig. 24-4. The economy of scale relationship for transmission facility line costs, 1975.

factor of one hundred to one, from the earliest carrier systems to the highest-capacity radio and guided systems. The costs, moreover, are on a current-dollar basis. If the plots were normalized for inflation, the ratio of costs between the earliest and mid-1970s systems would have been increased by at least another factor of ten. The curves also show why radio systems provided most of the long-haul mileage. Where they could be used, they were the lowest-cost alternative during the era when systems with route capacities of 5000 to 15,000 channels met the need.[3]

IV. THE RENEWED ATTACK ON SHORT HAUL-T CARRIER

Control signaling arrangements for toll dialing significantly increased the cost of carrier terminals. In the 12-channel K carrier installations of the early 1950s, for example, the provision of signaling for automatic switching increased the break-even distance, compared to a VF line, from about 40 to 75 miles. The major motivation for the N carrier development with companding, double-sideband transmission and (relatively) inexpensive signaling was to reduce the break-even distance, so that the system would be economical in the exchange

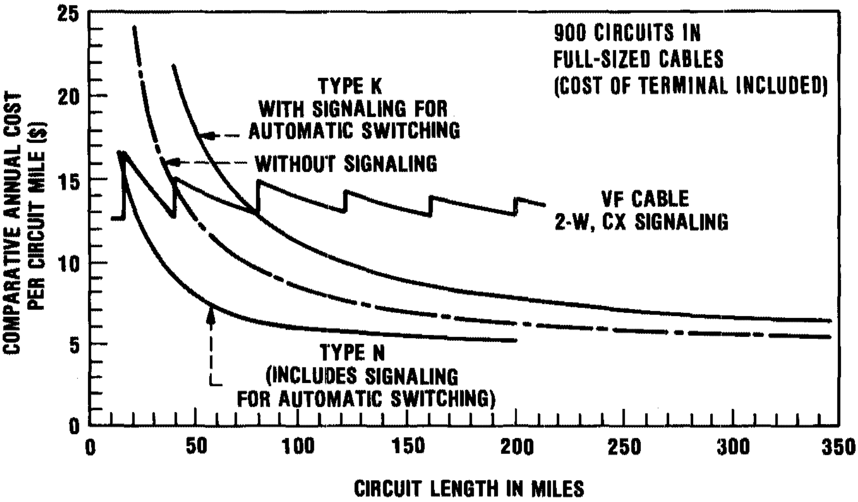


Fig. 24-5. Economic comparison of voice-frequency line to Types N and K carrier.

area. But the early N carrier systems were mostly used for short toll trunks, and relatively few systems were installed below 20 miles [Fig. 24-5]. The major breakthrough for carrier in the exchange area came in 1962 with the first digital system, T1. T carrier benefitted especially from the built-in signaling that used the eighth bit of the codes and required very little hardware beyond what was required for the voice-signal processing.

During the studies leading to the development of T carrier, a projected cost comparison indicated that the PCM system would be less costly compared to N carrier or voice-frequency lines below about 12 miles. Costs changed considerably during the course of the development and in the years that followed, but, by the 1970s, the per-channel cost of D1 multiplex terminals had been reduced below $500, and the break-even point against a typical voice-frequency line was below ten miles. This expanded the market so that by the mid-1970s over 2.5 million channels of T carrier were in service, more than the total of all other types of carrier channels combined.

Continued cost reductions and successive generations of design using larger- and larger-scale integrated circuits further reduced terminal costs until in the D4 bank, introduced in 1977, the cost was only about $300 per channel and the break-even point was at five miles or less. For a toll connecting trunk terminating on the digital 4ESS* switch, a digital trunk with a D4 terminal was most economical at any distance. In the 1980s, most new exchange trunks were being derived by T carrier. By that date, emphasis had shifted from lower first cost of the shipped equipment to features that actually increased the cost of the bank somewhat but that facilitated and reduced the cost of engineering,

* Trademark of AT&T Technologies, Inc.

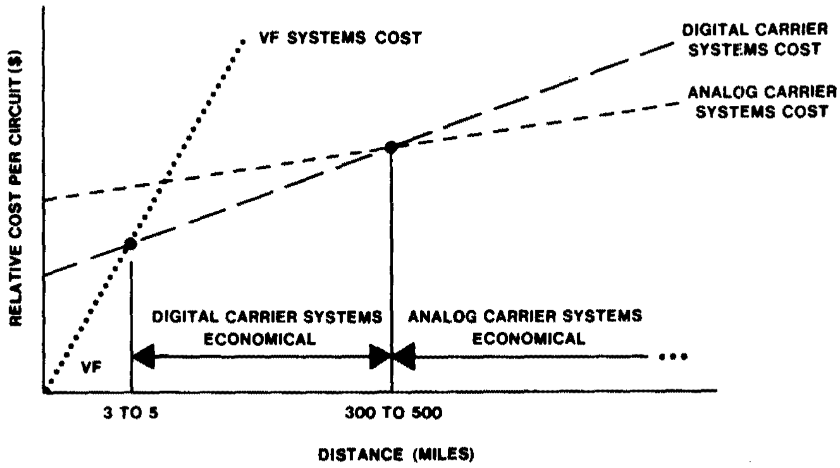


Fig. 24-6. Relative economics of transmission systems in the late 1970s.

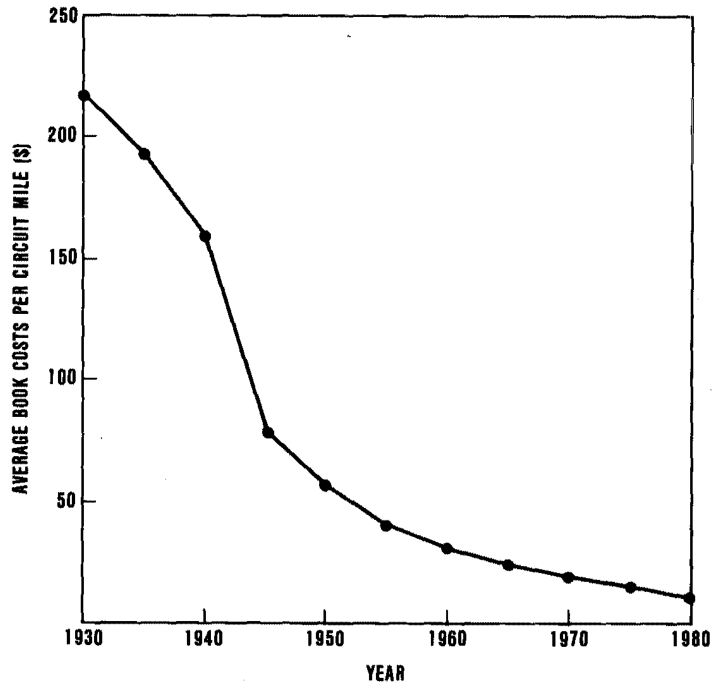


Fig. 24-7. Long Lines Department average embedded cost per circuit mile, 1930 to 1980.

installation, and maintenance, as well as the cost of providing and maintaining special services.

V. SUMMARY

The period from 1925 to 1975 saw a spectacular decline in the line cost of transmission circuits, approaching, in normalized dollars, a factor of 1000. Terminal costs were also reduced dramatically, but it was not until the introduction of digital systems in 1962 that carrier was economic for shorter exchange trunks. In 1975, the large-capacity analog systems were still least costly for the longest circuits, but digital systems were proving in at ranges up to a few hundred miles as well as for all but the shortest local trunks [Fig. 24-6]. The advent of optical systems in the 1980s seemed almost certain to provide the last link in an economical, nationwide, digital transmission network.

Each new system had to be added to the existing network in a compatible way. It was neither feasible nor economic to convert the entire system to the latest technology over-night. Nevertheless, the combined effect of rapid growth and the constant reduction of cost per circuit had a striking impact. This is exemplified in the Long Lines interexchange network, where the average net book cost per circuit mile decreased by a factor of 20, from $217 to $11 per circuit mile, over the 50 years from 1930 to 1980 [Fig. 24-7].

REFERENCES

1. Bell Laboratories Members of Technical Staff and Technical Publications Department, *Engineering and Operations in the Bell System*, 1st Ed. (Bell Telephone Laboratories, 1977), pp. 199-221.
2. American Telephone and Telegraph Company Construction Plans Department, *Engineering Economy: A Manager's Guide to Economic Decision Making*, 3rd Ed. (New York: McGraw-Hill, 1977).
3. E. F. O'Neill, "Radio and Long-Haul Transmission," *Bell Lab. Rec.* **53** (January 1975), pp. 50-59.

Epilogue

No explicit theme was stated for the first volume of this series, *The Early Years (1875-1925)*, but it might well have been the conquest of distance and the achievement of a national network. The theme of the current volume and its companions might equally well be the conquest of cost and the achievement of universal service.

In 1925, telephone subscribers could call just about anywhere in the United States. Every city and town of any consequence was connected to the nationwide network. With the cooperation of the local company (Bell System or independent), AT&T Long Lines, and two or more operators, any two telephones in the United States could be connected over a transmission path of decent quality in a reasonable time. By 1927, that connection could be extended overseas. Overseas calls were poorer in transmission quality, less convenient to place, and considerably higher in cost than a domestic long-distance call; but, by present standards, even the domestic call was not cheap or easy to place. Moreover, not everybody could make such calls, at least not from their homes, because many people did not have telephones.

In 1925, there were 16 million telephones in the United States or only about 14 telephones per 100 of population. During that year, subscribers placed 21 billion local and 1 billion toll calls. The toll calls were placed over a network consisting of 2.7 million voice-circuit miles, 98 percent of which consisted of voice-frequency trunks on copper open-wire or cable pairs. The sprouting, but still miniscule, carrier plant provided only 47,000 voice-circuit miles.

By 1981, there were 191 million telephones in the United States or 84 telephones per 100 of population. Ninety-six percent of the households had telephones and the subscribers placed 238 billion local and 20 billion toll calls that year. The 10- to 20-fold increase in calls does not indicate the full measure of increase in traffic. The average distance over which calls were placed, especially toll calls, and the holding times also increased. The calls placed in 1981 were carried over a trunk network consisting of more than one billion voice-circuit miles, over 300 times the network mileage of 1925. All but a miniscule part of that network was derived by carrier techniques, with more than 100 million voice-circuit miles on digital systems.

During the 25 years from 1925 to 1950, the toll transmission plant capacity increased from 2.7 million to 23 million voice-circuit miles, even with a conspicuous plateau during the depression years [Fig. EP-1]. Until 1940, transmission was largely at voice frequency; carrier on pairs still accounted for only 12 percent of the total. By 1950, two-thirds of the mileage was on carrier, with coaxial systems contributing 30 percent of the carrier total. Within five years of the start of their major deployment, coaxial systems provided half again as much toll mileage as the total that had been built in the first half century of the system's existence.

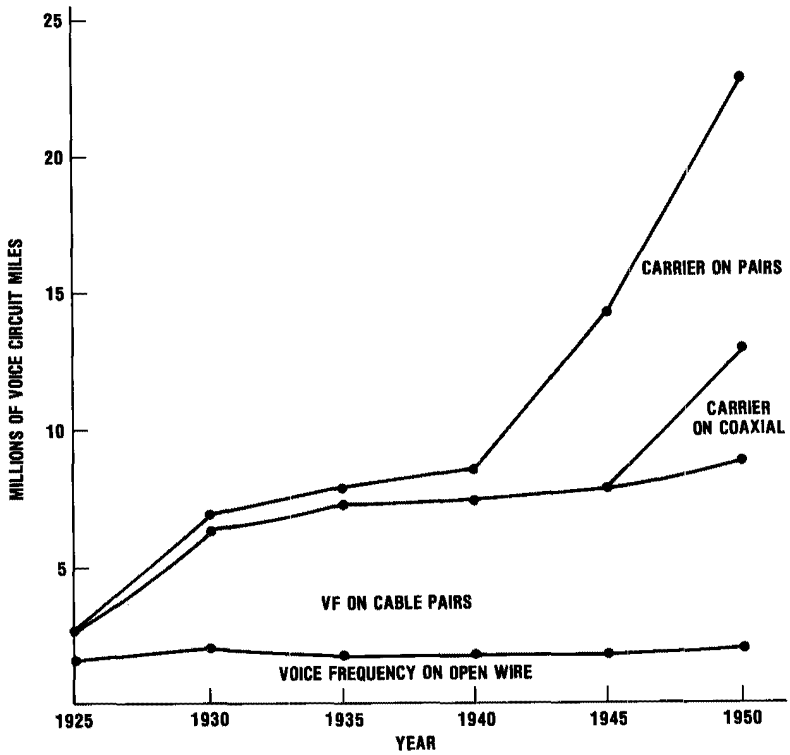


Fig. EP-1. Transmission plant mileage, 1925 to 1950.

But even the large increase between 1925 and 1950 turned out to be merely the stage setting for the enormous expansion of the next 25 years. The transmission capacity in 1950, which looked so large in the eyes of the people who designed and built it, can hardly be found on a plot of the growth to 1975 and 1980 [Fig. EP-2]. The policy of having circuits available on demand, the reduction in real rates that accompanied the postwar expansion, and the introduction of customer direct dialing transformed long-distance calling from an occasional necessity, surrounded by considerable ceremony, into a casual everyday thing. In response, the plant capacity expanded by a factor of over 30 from 1950 to 1975 and nearly doubled again to exceed one billion voice-circuit miles by 1980. The lion's share of this was provided by microwave radio, first introduced on a commercial basis in 1951. By 1965, radio supplied more than half the total and, by 1975, 62 percent. While coaxial systems furnished only 20 percent of the 1980 total, this still amounted to a healthy 200 million voice-circuit miles—greater than the entire plant capacity in 1965. A large part of the growth in the later years was due to upgrades that increased the capacities of radio and coaxial systems by mining circuits out of existing rights-of-way, cables, buildings, towers, and antennas, with a minimum of new construction.

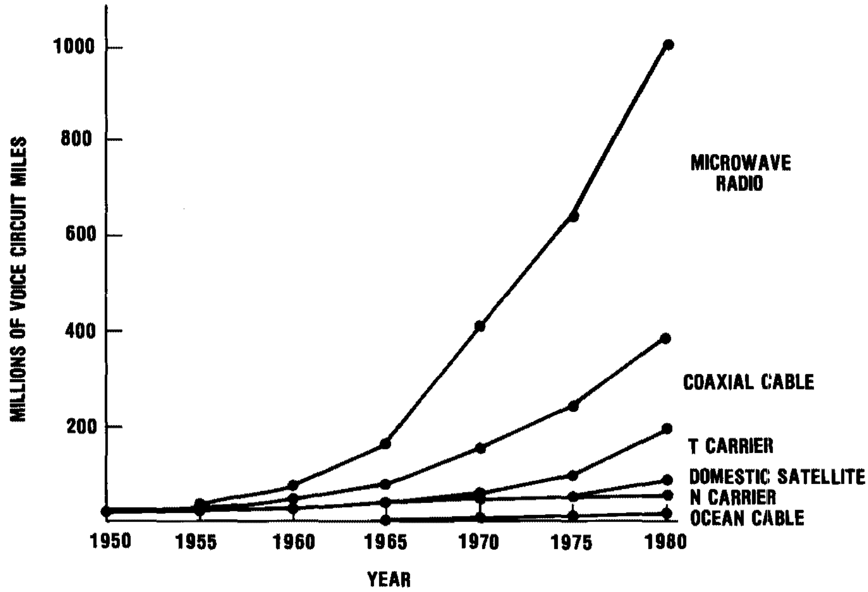


Fig. EP-2. Transmission plant mileage, 1950 to 1980.

The growth in the year 1980 alone was 139 million circuit miles, greater by 20 million miles than the entire system capacity only 17 years earlier, in 1963.

A hint of the shape of things to come was in the growth of digital systems to over 100 million voice-circuit miles, and to ten percent of the total by 1980. All of this was in local or short toll circuits. However, starting in 1983, digital long-haul transmission was launched on the fiber-optic system between New York and Washington.

Will the fantastic growth rates of the past 25 and 50 years, or rates even close to them, be sustained? If growth rates of similar magnitude to those in the past are to continue, it is not likely to be to furnish greater speech signal capacity. It is difficult to imagine that the nation's people can spend much more time on the telephone than they already do. However, major new services in data transport, information to homes by video display, or person-to-person high-definition video may well create a new surge of demand. If the demand arises, it appears that the technology will be ready for it. In early 1985, Bell Laboratories demonstrated ten lasers, each operating at 2 gigabits (2×10^9 bits) per second, wavelength multiplexed on a single glass fiber, with regeneration at 42-mile spacing. This was, of course, only a laboratory demonstration, but, if implemented on two fibers from New York to Chicago, such a system would provide capacity equivalent to over 200 million voice-circuit miles with only 20 repeater locations.

Whatever the future holds, the research, development, and construction of the Bell System transmission plant will surely rank as one of the major technical accomplishments of the 20th century. In 1925, the Bell System had 336,000 employees, about 240 per 10,000 telephones or 15 per million calls. In 1981, with 1,060,000 employees, it required only 55 per 10,000 telephones or 4 per million calls. There were very few areas of the American economy that demonstrated improvements in productivity as great as that. There is no question that most of the improvement resulted from the advance of technology fostered by and, in large part, produced within the system. Transmission technology made a large contribution to the overall improvement and an even more striking contribution to the reduction in long-distance costs.

In 1925, a three-minute station-to-station, coast-to-coast call cost $15.95 during the business day, $8.00 after 8:30 PM, and $4.00 after midnight (in 1925 dollars). Even at the lowest rate, it required 7-1/2 hours of the average factory worker's pay to place such a call. In 1983, a three-minute subscriber-dialed coast-to-coast call on Bell facilities cost $1.72 during the business day, $1.04 after 5 PM, and $0.69 after 11:00 PM and for most of the weekend. At the lowest rate, it would have cost just five minutes work at a factory worker's average pay to place such a call. By any measure, the half century assault on costs was enormously successful.

Abbreviations and Acronyms

Abbreviations and Acronyms

a	ampere
A/D	analog to digital
A/D/A	analog/digital/analog
AC	alternating current
ALBO	automatic line buildout
AM	amplitude modulation
AMPS	advanced mobile-phone service
ATMS	automatic transmission-measuring system
AT&T	American Telephone and Telegraph Company
AUTOVON	automatic voice network
BSRFS	Bell System reference frequency standard
C	centigrade
CAROT	centralized automatic reporting on trunks
CBIC	complementary bipolar integrated circuit
CCIS	common channel interoffice signaling
CLOAX	corrugated-laminated coaxial cable
cm	centimeter
CMS	circuit maintenance system
CODAN	carrier-operated device, antinoise
codec	coder-decoder
compander	compressor-expandor
Comsat	Communications Satellite Corporation
CONECS	connectorized exchange cable splicing
COTMS	computer-operated transmission measuring set
COZY	computer-operated Z/Y bridge
CRT	cathode-ray tube
CSU	channel service unit
DAME	design aid for multiplex equipment
dB	decibel
dBm	decibels referred to one milliwatt
dBm0	decibels above one milliwatt at the zero transmission level point
dBrnC	decibels above reference noise, with C message weighting
dBrnC0	decibels above reference noise, with C message weighting at the zero transmission level point

DC	direct current
DCT	digital carrier trunk
DDD	direct distance dialing
DDS	DATAPHONE* data communications service
DEPIC	dual-expanded polyethylene-insulated conductor
DIF	digital interface frame
DLL	dial long lines
DS	digital signaling
DT	digroup terminal
DTI	distortion-transmission impairment
DUV	digits under voice
EB	emergency band
EDT	ethylene diamine tartrate
EDST	eastern daylight savings time
F	Fahrenheit
FCC	Federal Communications Commission
FDI	feeder distribution interface
FDM	frequency division multiplex
FIT	failures in 10^9 device hours
FM	frequency modulation
FSK	frequency shift keying
GaAsFET	gallium arsenide field effect transmitter
GDF	group-distributing frame
GHz	gigahertz
GTE	General Telephone and Electronics
HF	high frequency
HIC	hybrid integrated circuit
Hz	hertz
IF	intermediate frequency
IMPATT	impact avalanche transit time
IMTS	improved mobile-telephone service
Intelsat	International Telecommunications Satellite Consortium
IRE	Institute of Radio Engineers
JB	jumbogroup
JFS	jumbogroup frequency supply
JMX	jumbogroup multiplex
JPL	Jet Propulsion Laboratory
kHz	kilohertz
km	kilometer

* Registered service mark of AT&T.

kw	kilowatt
LBO	line buildout
LED	light-emitting diode
LOCAP	low capacitance cable
LSA	limited space charge accumulation
mc	megacycles
MCVD	modified chemical vapor deposition
m	meter
μf	microfarad
μm	micrometer
μs	microsecond
ma	milliampere
MGDF	mastergroup distributing frame
MGT	mastergroup translator
MHz	megahertz
mm	millimeter
MMGT	multimastergroup translator
MMGTC	multimastergroup translator for cable
MMGTR	multimastergroup translator for radio
MMX	mastergroup multiplex
modem	modulator/demodulator
ms	millisecond
MTSO	mobile-telecommunications switching office
MUSA	multiple-unit steerable antenna
mw	milliwatt
MW	megawatt
NASA	National Aeronautics and Space Administration
NEXT	near-end crosstalk
NIC	negative-impedance converter
NPR	noise power rate
ns	nanosecond (10^{-9} seconds)
NTI	noise-transmission impairment
Ω	ohm
OCU	office channel unit
OMFS	office master-frequency supply
OTLP	zero transmission-level point
PAM	pulse-amplitude modulation
PCM	pulse-code modulation
pf	picofarad (10^{-12} farads)
PFS	primary frequency supply
PIC	polyethylene-insulated conductor
POTS	plain old telephone service
psi	pounds per square inch

PSK	phase shift keying
PUB	peripheral unit bus
PUC	peripheral-unit controller
pw	picowatt (10^{-12} watt)
QAM	quadrature amplitude modulation
RCA	Radio Corporation of America
REG	range extender with gain
RF	radio frequency
rms	root mean square
ROTL	remote-office test line
rpm	revolutions per minute
SAC	serving-area concept
SAGE	semiautomatic ground environment
SARTS	switched-access remote test system
SCARAB	submissible craft assisting recovery and repair
SLC	subscriber-loop carrier
SLM	subscriber-loop multiplex
SMAS	switched maintenance access
SSBAM	single-sideband amplitude modulation
STAR	standard tantalum active resonator
syndes	synchronizer-desynchronizer
TASI	time assignment speech interpolation
TE	transverse electric
TEC	terminal equipment center
TM	transverse magnetic
TMS	transmission measuring set
TOC	television operating center
TOC	trunk operations center
TSA	transmission-surveillance auxiliary
TSC	transmission-surveillance center
TSPS	traffic service position system
TWT	traveling-wave tube
TU	transmission unit
UHF	ultrahigh frequency
v	volt
VF	voice frequency
VHF	very high frequency
VIF	voice-band interface frame
VIU	voice-band interface unit
VODAS	voice-operated device, antisinging
VLA	very large array
VOGAD	voice-operated gain-adjusting device
WE	Western Electric
w	watt

Credits

Credits

Figure 3-4 from J. R. Poppele, F. W. Cunningham, and A. W. Kispaugh, *Proceedings of the IRE* **24** (August 1936), p. 1072. Copyright 1936 by the Institute of Radio Engineers, Inc.

Figure 3-6 from J. R. Poppele, F. W. Cunningham, and A. W. Kispaugh, *Proceedings of the IRE* **24** (August 1936), p. 1081. Copyright 1936 by the Institute of Radio Engineers, Inc.

Figure 3-7a from W. H. Doherty, *Proceedings of the IRE* **24** (September 1936), p. 1173. Copyright 1936 by the Institute of Radio Engineers, Inc.

Figure 3-10 from P. H. Smith, *Proceedings of the IRE* **35** (December 1947), p. 1557. Copyright 1947 by the Institute of Radio Engineers, Inc.

Figure 3-12 from P. H. Smith, *Electronics* **12** (January 1939), p. 31. Copyright 1939 by McGraw-Hill Publications Co., Inc.

Table 4-1 from H. W. Bode, *Symposium on Active Networks and Feedback Systems* **10,** Brooklyn, NY (April 19–21, 1960), p. 5. Copyright 1960 by the Polytechnic Press of the Polytechnic Institute of Brooklyn.

Figure 4-6 from H. W. Bode, *Symposium on Active Networks and Feedback Systems* **10,** Brooklyn, NY (April 19–21, 1960), p. 8. Copyright 1960 by the Polytechnic Press of the Polytechnic Institute of Brooklyn.

Figure 4-7 from H. W. Bode, *Symposium on Active Networks and Feedback Systems* **10,** Brooklyn, NY (April 19–21, 1960), p. 9. Copyright 1960 by the Polytechnic Press of the Polytechnic Institute of Brooklyn.

Figure 7-12 from *Post Office Electrical Engineers Journal* **24** (July 1931), p. 153. Copyright 1931 by the Institution of British Telecommunications Engineers, London.

Table 9-2 from G. W. Bleisch, *IEEE Transactions on Communications Technology* **13** (September 1965), p. 361. Copyright 1965 by the Institute of Electrical and Electronics Engineers, Inc. Reprinted with permission.

Figure 13-4 from D. C. Hogg, *Journal of Applied Physics* **30** (September 1959), p. 1417. Copyright 1959 by the American Institute of Physics.

Index

Index

A

B

D

G

H

I

J

K

N

S

T