AT&T Technical Journal
Vol. 64, No. 3, March 1985
Printed in U.S.A.

Selection of a Laser Reliability Assurance Strategy for a Long-Life Application

By F. R. NASH,* W. B. JOYCE,* R. L. HARTMAN,* E. I. GORDON,† and R. W. DIXON*

(Manuscript received March 5, 1984)

We are concerned with assuring the reliability of semiconductor lasers intended for an application in which the design lifetime is long, replacement or redundancy is impossible or impractical, and the failure of even a few lasers could be disastrous. In the search for a reliability assurance strategy that will meet our objectives, we have carefully examined the well-known and widely used bathtub and lognormal approaches. Based upon our understanding of the expected aging behavior of lasers, we propose an alternative reliability assurance strategy that we believe to be an improvement over the traditional approaches. The object is not how to make reliable lasers, but rather how to confidently predict, in a timely fashion, which lasers in a given population will endure beyond the intended system lifetime. Particular emphasis is placed upon initially imposed overstress regimes that address the anticipated presence of transient modes of degradation and infant failure modes with low thermal activation energies that may be invulnerable to detection during accelerated thermal aging. Since lasers degrade gradually rather than fail suddenly, comparable emphasis is placed upon monitoring the stabilized long-term degradation rates of the survivors of the overstress regimes so as to permit lifetime predictions of individual lasers.

I. INTRODUCTION

The problem that we faced some two years ago of assuring the reliability of the lasers to be used in the SL optical fiber submarine cable may be summarized as follows:

---

* AT&T Bell Laboratories. † Lytel, Inc., Somerville, New Jersey.

Given: A state-of-the-art device, i.e., a new semiconductor laser structure composed of new materials, which had no established record of reliability, which had never seen system use, and about whose details of operation there was a paucity of knowledge;

Goal: Design a reliability assurance strategy to determine which, if any, of a given group of these lasers may be confidently installed in a system which must operate for 25 years under circumstances where redundancy is too expensive, replacement is impractical, and where the failure of only a small number of lasers would be crippling.

To take advantage of reductions in the loss and dispersion of optical fibers, a low-threshold index-guided InGaAsP/InP laser operating at a wavelength of 1.3 μm was chosen instead of the higher-threshold gain-guided GaAlAs/GaAs laser that operated in the wavelength range 0.8 to 0.9 μm. Feasibility demonstrations of the potential for adequately long life had been made for some significant fraction of manufactured InP lasers. The size of the usable fraction and a strategy for selecting it prior to installation remained to be determined. In view of the aggressive schedule for deployment (1988), it was inevitable that the emphasis be placed upon the reliability assurance of an existing population and not upon the issue of how to design for reliability in the manufacturing process. Thus screening, not product improvement or failure mode analysis, is the primary goal.

Building in reliability from the start involves the complex technological problem of utilizing the details of innumerable specific failure mechanisms to modify the laser structure, design, growth, and processing stages of manufacture. The reliability literature is dominated by device-specific, failure-mode-specific studies, which, however useful for providing feedback to the manufacturer, do not provide an organized way of thinking about generic modes of failure and schemes tailored to their discovery. Even if expense and time were not constraints, it would be impossible, in any case, to arrive at a perfected manufacturing process that produced only long-lived devices. Thus, sooner or later, the problem that must be addressed is that of warranting the reliability of a given group of lasers, with fixed initial properties, irrespective of the failure mechanisms still operative in the manufactured population.

This paper charts the search for the reliability assurance strategy that was best suited for our goal. The intended audience is not the statistician or applied mathematician, but rather the device physicist who may some day be faced with our problem in the context of a laser or another new semiconductor device. The natural starting point for our inquiry is an examination of the strategies that are already known, in particular, the bathtub and lognormal approaches that have been widely and successfully used for some time. These traditional quality

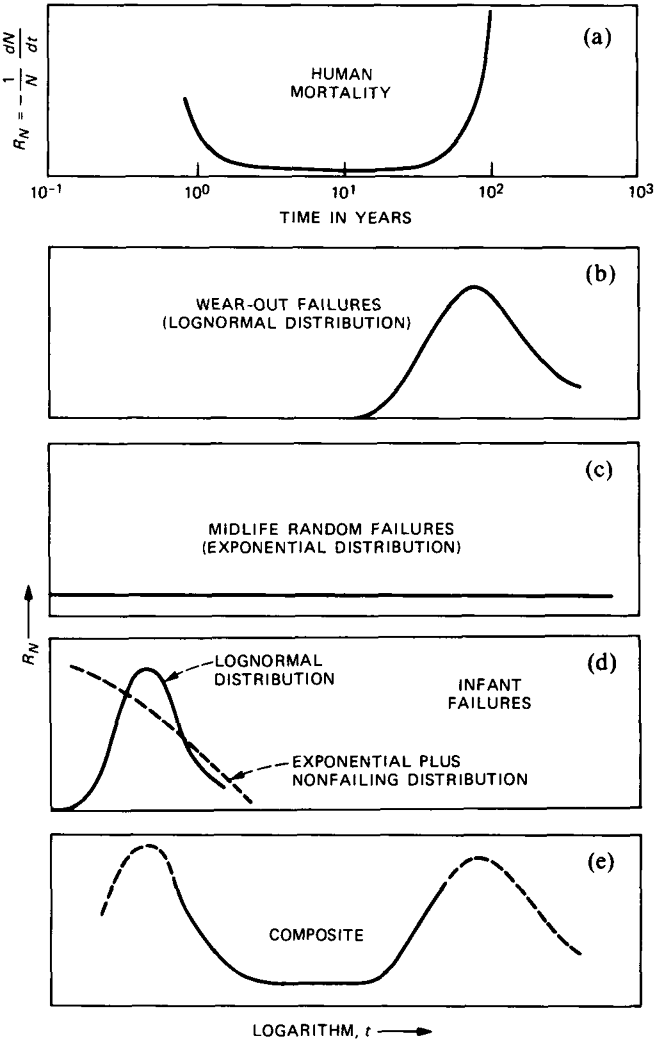

Fig. 1—(a) The hazard rate (bathtub) curve for human mortality. (b) A lognormal distribution characterizing the wear-out or long-term failures in a population of semiconductor devices. (c) An exponential distribution which characterizes random failures, the so-called accidents in the midlives of devices. (d) Solid curve—a lognormal distribution characterizing an infant population. Dashed curve—an exponential plus a nonfailing distribution used to generate a declining hazard rate for a subset of infant failures caused by random externally imposed stress aggregations. (e) A curve which is a composite of (b), (c), and the lognormal portion of (d); beyond the bathtub part (solid curve), there is rollover of the bathtub sides (dashed curve). Only in a restricted domain of time does the semiconductor failure of (e) resemble human mortality (a).

control approaches to reliability assurance, used separately or in concert, and which rely on inferences made from the behavior of previously aged-to-failure populations, will be shown to be inadequate for our stated goal. As shall be discussed, a preferable alternative approach, which is applicable to lasers, relies upon inferences made from the early behavior of each individual device in the very population intended for deployment.

So that a reader has the option to pick and choose among the various sections, depending upon his background in reliability assurance matters, the following "roadmap" and summary of our major points may prove useful. Section II reviews the bathtub model and a compilation of associated reliability terminology that it is important to define for purposes of making clear distinctions in subsequent discussions. Also discussed is a suggestion for a modification of the bathtub curve, which appears to be dictated by actual laser reliability data and a practical way in which to acquire bathtub data.

The traditional (human mortality) bathtub curve, a plot of the failure (hazard) rate versus time (see Fig. 1a), has a declining left-hand edge (period of infant mortality), a relatively flat bottom (period of useful life), and an increasing right-hand edge (period of inevitable failure or wear-out). By analogy, semiconductor devices are assumed to be susceptible to wear-out, but in addition some fraction of them is viewed as possessing additional failure mechanisms that cause early or infant device inoperability. If we had the traditional option of using only well proven-in devices with "nailed down" invariant manufacturing procedures and a well-established record of longevity in terrestrial applications, the bathtub reliability assurance strategy for a nonreplacement high-reliability use could be stated as follows: (A) use a mature technology that offers an intrinsic long life so that the wear-out failures (increasing right-hand side of Fig. 1a) are not present during normal (system) life, and (B) censor atypical (infant) devices (decreasing left-hand side of Fig. 1a). Simply put, the strategy is to install the survivors of the censoring process.

Competitive pressures have substantially eroded the traditional conservative approach. We are presently compelled to seriously consider the use of laser devices of an a priori unknown quality in a time frame that does not permit us to wait and see exactly what types of failures occur during some intended use. Furthermore, even were we permitted to establish some kind of reliability record, there would be no certainty that any understanding of the failures could, in feedback fashion, be translated in any timely way into corrective manufacturing changes. Our only course of action is to anticipate the worst and ask how the traditional approaches can be improved, since it is likely that the success of the entire fiber-optic undersea cable project will depend importantly upon the quality of the reliability assurance strategy employed.

Two of the potentially serious problems connected with any uncritical application of the bathtub strategy, (A) and (B) above, are examined in Section III. Using GaAlAs/GaAs laser reliability data as a guide, we show in our analysis that we must anticipate that the hazard rate curves for the infant and wear-out populations may overlap considerably so that the bathtub will not have any obvious flat-bottomed period of useful life. Thus, even if infant devices were successfully screened out, wear-out-type failures, perhaps in significant numbers, might be expected to occur during system life, and therefore (A) seems ruled out. With respect to (B), the usual procedure for promoting infant failures prior to deployment of the product is to accelerate their occurrence by operating the devices for a short period of time at an elevated temperature, the so-called elevated temperature *burn-in* screen. In Section III we emphasize the possibility, again supported by GaAlAs/GaAs failure data, that low thermal activation energy infant failure modes may exist that are invulnerable to detection in elevated temperature operation for a short duration. Therefore, it would be imprudent to rely exclusively upon a burn-in to accomplish (B).

Let us return briefly to the problem of wear-out failures occurring during system life. Assume for the moment that the wear-out failures occur suddenly and without any warning. An important addition to reliability assurance in the presence of this problem, which defeats (A) above, is the lognormal statistical approach, discussed in Section IV. It assumes that the wear-out failures are temporally distributed according to some law, in this case the lognormal. The hazard rate curve can be estimated from experimentally estimated values of the median lifetime and the variance. In a nonreplacement application, the hazard rate curve can be used to construct a redundancy scheme to allow for a certain percentage of failures during the system life. Thus in the event that the wear-out failures are unaccompanied by any detectable precursor degradation, the best reliability assurance strategy would appear to be a combination of the bathtub and lognormal approaches, to wit: (A′) statistically characterize the wear-out of a population to determine the expected number of failures during the system life and then provide sufficient redundancy, and (B) censor infant devices.

An obvious and unavoidable weakness of the bathtub and lognormal approaches, in the case of sudden wear-out failures, which is shared by any sampled-population or population-inferential approach for high-reliability applications, is that conclusions about the behavior of as-yet unmanufactured devices must be based upon aging to demise some early-manufactured population. Given the inevitable inadvertent and premeditated changes in growth and processing of devices, the assumption that all populations are statistically equivalent, no matter when manufactured, may not always be correct.

Beyond our tutorial critique of the reliability assurance strategies which have been used in the past, our forwarding contribution (Section V) to improving reliability assurance comes from the recognition of three properties of lasers: (1) lasers are prone to possess low thermal activation energy modes of failure; (2) lasers may also have initial modes of degradation which eventually saturate or stabilize and which are not indicative of long-term behavior; and (3) lasers, with rare exceptions, do not fail suddenly; instead they invariably degrade slowly. Consequently, we shall describe two procedures that we believe must be incorporated into any laser reliability assurance program. The first relates to the low thermal activation energy early or infant failure mode, which is not particularly susceptible to the traditional burn-in employing temperature as the degradation accelerant or driving mechanism. It is crucial to employ considerable overstressing, as well, of the injection current and optical power in order to purge (screen out) devices which possess this potentially devastating failure mode. Overstressing of current, optical power, and temperature are also considered important to compel rapid stabilization of certain initially occurring modes of degradation in order that the activation energy and degradation rates of the long-term wear-out mode be credibly determined in a timely fashion. The second thrust is to attempt to estimate lifetimes by monitoring the thermally accelerated wear-out rate of degradation of every individual laser that survives the overstress regimes (designed to eliminate infants and stabilize transient modes of degradation). Truncation of the surviving population will be accomplished by removing lasers whose initially observed long-term rates of degradation are too large.

Relative to the concerns of this paper, we call the reader's attention to several other papers in this volume. A successful implementation of the proposed strategy (purging, stabilization, truncation) for lasers is described by Nash et al. The truncation procedure is discussed more generally and quantitatively in papers by Joyce et al. and Eckler. Purging as a procedure for establishing the reliability of detectors is demonstrated by Saul et al.

II. THE BATHTUB MODEL AND DEFINITIONS OF SOME RELIABILITY TERMINOLOGY*

2.1 Description of the bathtub curve

In discussions of the reliability of products, it has been a common practice to plot the device hazard (failure) rate as a function of time. Imagine that a number of devices have been placed on test at $t = 0$

* The Glossary in the Appendix may be consulted for definitions of some reliability terminology also given in the text.

and that the failures are noted as time progresses. If, after the lapse of a time t, the expected number of survivors is N, then the rate at which devices can be expected to fail at time t is $-dN/dt$. The hazard rate is a normalized instantaneous failure rate and is defined as

$$R_N = -\frac{1}{N}\frac{dN}{dt}. \tag{1}$$

This can be shown[1,2] to agree with the conventional definition.[3] A slightly idealized version of a plot[4] of R_N versus time for human mortality, using data on which life insurance companies base computations of premiums, is shown schematically in Fig. 1a. This so-called bathtub curve for ages 0 to 100 years has the following interpretation. If a person survives the diseases of infancy (infant failures are associated with the decreasing hazard rate up to ~5 years), and then survives the accidents that cause demise in midlife (midlife failures are associated with the roughly constant hazard rate in the range 5 to 50 years), that person will die because of old age (wear-out failures are associated with the increasing hazard rate in the period >50 years). The useful lifetime of a surviving individual is the period up to ≈70 years.

The bathtub curve has been widely employed[1,3,5-10] as a conceptual device to model the reliability of mechanical, electrical, and semiconductor components and systems. It is intuitively obvious that the reliability of light bulbs, brake linings, tires, shoes, etc., will be described by such a curve.* Experimentally, bathtub-like hazard rate plots have been found for radar sets,[5] vacuum tubes,[11,12] resistors,[13] capacitors,[14] electronic equipment,[15] inertial guidance systems,[16] airplane engine starting systems,[3] and bipolar integrated circuits.[17]† We shall assume for the present that Fig. 1a is a plausible depiction of the hazard rate of some semiconductor devices in the course of an intended use. It is understood that none of the devices was subjected to any preoperation screening (e.g., burn-in‡ or accelerated thermal aging),

* Consider candles as an example. Infant failures are those candles which are made too long, too short, broken, or without wicks. Barring accidental breakage, all candles similarly made and burned are expected to live a well-defined period of time before wear-out (no more wick).

† For the first-made devices in any new and probably poorly controlled semiconductor device technology, the bathtub might not hold any water, i.e., it might not have a minimum, because of the presence of a multiplicity of failure mechanisms, each operative in different time frames, and each with a large variance in its temporal occurrence. After considerable effort using the physics-of-failure[18,19] approach (analysis of failed devices to discover the failure mechanism and feedback of this information to generate suitable controls for each mechanism by design changes or process control), a bathtub curve may result. Some infant mortality is invariably still present, and wear-out is inevitable.

‡ We define burn-in to mean normal operation at the intended use temperature. An elevated temperature burn-in will be equivalent to accelerated thermal aging.

for otherwise the infant failure population would have been largely eliminated.

2.2 Wear-out (preordained failure)—gradual or sudden

Thus far we have made one kind of distinction among device failures, i.e., when failure occurs (infant, midlife, wear-out). Another convenient *when* term is premature failure, which is inoperability prior to some desired application lifetime. This is a system-lifetime rather than a device-lifetime term. It is useful because the actual lifetime of an individual device is irrelevant if it greatly exceeds the system lifetime. Included in the premature class may be the infant, midlife, and early wear-out failures of the bathtub curve. We now wish to make another kind of distinction that relates not to when failures occur, but rather to why. We call the two classes preordained and chance.* The dominant long-term (wear-out) mechanism causes devices to degrade "deterministically at a rate preordained" but unknown at the moment of birth.[20] "The laser-to-laser variations in lifetime are the result of random slice-to-slice and laser-to-laser growth and processing variations which result in initially inequivalent lasers."[20] The distribution of the lifetimes of wear-out failures has often been modeled as being lognormal† (i.e., Gaussian in the logarithm of lifetimes) for semiconductor devices, e.g., germanium and silicon transistors,[2,18,21-23] low-noise GaAs Field-Effect Transistors (FET)s,[24] power GaAs FETs,[25] GaP Light-Emitting Diodes (LED)s,[26] GaAs LEDs,[27] and GaAs lasers.[20,28-33] The basis for this behavior is not understood,[20,34] although a proportional-effect model has been proposed.[35]

Preordained failures, of which wear-out failures are examples, may be either gradual or sudden (catastrophic); these are *how* terms. Thus, one speaks of gradual degradation in the sense in which some measurable parameter changes slowly or gracefully with time, and by monitoring that parameter a time-to-failure may be predicted.[20,24-26] The essence of the failure is a change in one or more parameters which exceed some specification. On the other hand, one may speak of catastrophic failure.[24-26] In this case the emphasis is upon the suddenness and the complete inoperativeness of the device as contrasted with a slow change that eventually exceeds some specification. Electrolytic corrosion of contacts[24] and burnout[25] of FETs are examples of sudden

* Chance or random failures will be discussed in Section 2.3.

† The Weibull model[3,36,37] has also achieved popularity. It is difficult to distinguish between Weibull and lognormal distribution functions; differences become significant only in the tails of the distributions, but actual observations of times-to-failure are sparse in the tails because of limited sample sizes.[38] "[T]he only practical way of progressing is to choose a simple function, test it empirically, and stick to it as long as none better has been found."[11,36]

failures. The suddenness refers to the fact that the complete failure occurred somewhere in between two times of observation; the actual failure might have occurred in a millisecond or a day. The failure was, nonetheless, still preordained, i.e., due to some initially present device weakness, but it was inconvenient or impossible to monitor the parameters that would have enabled failure to be anticipated. Even had monitoring occurred, the changes might have been too small to detect in the incubation period prior to reaching some threshold at which, e.g., a thermal runaway occurred.[25]

When the distribution of lifetimes is a Gaussian and the independent variable is the logarithm of time, the corresponding hazard rate curve is asymmetric with a positive skew.[21] Figure 1b depicts the hazard curve for a lognormal wear-out distribution where the ordinate is linear.

2.3 Midlife (chance) failure

In contrast to the preordained[20] or time-dependent-type[39] failure there is the so-called chance or random failure,[20] which has also been called event dependent.[39] The chance aspect exists in factors external rather than internal to the device. External events,* the most likely of which are electrostatic discharges[39] and random accumulations of operating stresses (e.g. current, mechanical vibration, etc.) that exceed device capacity,[1,7] can cause these failures to weak (susceptible to infant failure) and strong devices alike, and will occur even if all devices are initially equivalent in *all* respects. No preinstallation screen, burn-in, or accelerated thermal aging is capable of identifying which devices have a potential for chance failure, since all devices have the same potential. By definition, the chance failure does not display any symptoms of prior deterioration.†

By analogy to spontaneous radioactive decay, chance failure is usually associated with the Poisson or exponential distribution.[20] Thus, if at time t, the expected number of survivors in a population, whose initial number was N_0, is given by $N = N_0 \exp(-t/\tau_c)$, it follows from (1) that $R_N = \tau_c^{-1}$, where τ_c is the lifetime against chance failure. A population whose failure is governed exclusively by chance failures with constant R_N is depicted in Fig. 1c. This constant (or near constant) hazard rate is associated with the bottom of the bathtub and the so-called midlife or accidental failures.

---

* For a subcable laser system, an unlikely event would be the creation of current leakage paths by cosmic rays.

† Note that the sudden preordained failure and the sudden chance failure have in common a lack of observed precursor degradation. It is conceivable that a chance mechanical shock could be responsible for starting a gradual failure. Even in such an event, there still would have been no precursor degradation prior to the shock, so that the event is still of the chance variety.

2.4 Infant (preordained or chance) failure

So far, in the course of discussing preordained and chance failures, we have examined the bottom (midlife failures) and right-hand side (wear-out failures) of the bathtub curve, Fig. 1a. The left-hand side is associated with the so-called infant mortalities. It is not expected that the fundamental wear-out mode of degradation, presumably common to all devices, will actually control the failure of every device. On occasion, other mechanisms that are simultaneously and randomly present in the population of devices will cause early (infant) failures, substantially prior to the times-to-failure that would have occurred if wear-out had been the only degradation mechanism operative.

Unlike the wear-out mechanism, which is generally well characterized even though its manifestation is intended to be beyond the desired useful lifetime of devices according to Fig. 1a, the infant failures are poorly characterized[17-19] for several reasons: (1) the $t = 0$ point is arbitrary; (2) the infant failures often occur in such relatively short periods of time after the device turn-on that hazard rate data are difficult to acquire for each small increment of time during operation; (3) for the same reason, it is inconvenient or impossible to monitor a suitable device parameter during the lifetime to distinguish preordained from chance failures; (4) infant failures may not show up in small-sample life tests and large samples may be difficult to obtain in a rapidly evolving technology; (5) the major concern is with knowing the screening techniques required to eliminate the infant class and not with knowing the precise failure distribution; and (6) evidence[17,18] suggests that the infant failures have a number of origins, so that, for example, only a fraction may be susceptible to thermal acceleration. The activation energies associated with infant failures in an integrated circuit are typically low (0.25 to 0.40 eV).[17] The operating time duration in which infant failures occur appears to be $\sim 10^4$ hours.[17,39]

One view[17,18] of infant failures is that they represent workmanship-like defects (e.g., cracked chips, actual or incipient opens or shorts) which require some time of operation to be detected. Such devices are initially substandard* and are called weak because they are destined

* There is another class of substandard devices which may be called odd or eccentric. Although they are not obviously connected with premature device failure, their initial characteristics are sufficiently different from the main population to warrant rejection for reasons of eccentricity. This is one of the prices paid for assuring reliability in any demanding application. Imagine that for a particular wire bond, there is an air space over one-half of the potential attachment area. Imagine also that mechanical vibration is largely absent in the intended operating locale. A harsh vibration or pull test might very likely reveal this half-attached bond. To be sure, it might be considered wise to identify the laser as a so-called odd or substandard device, but it is far from obvious that this device would ever fail, during the cable lifetime, much less be a candidate for premature failure.

for early failure. The cause of failure would be a random accumulation of operating stress(es) (e.g., current transients) in excess of the normal, and thus would exceed the capacity of the weak, but not the strong devices. Even though the weaknesses are inherent in the devices, the failure-causing events are random and extrinsic, and hence the failure time distribution will be exponential. To derive a declining hazard rate (the left-hand edge of the bathtub curve, Fig. 1a), assume that the infant fraction of some population is governed by $N_{\text{in}} = N_{0,\text{in}} \exp(-t/\tau_{\text{in}})$, while in the relevant time frame, the remaining number (N_r) of devices do not degrade, or do so negligibly. From (1) it may be computed that

$$R_{N,\text{in}} = \frac{1}{\tau_{\text{in}}} \frac{1}{1 + \dfrac{N_r}{N_{0,\text{in}}} \exp(t/\tau_{\text{in}})}, \tag{2}$$

which continually decreases from $t = 0$. Such a function, plotted against the logarithm of time, is depicted schematically in Fig. 1d as a dashed curve.

Two observations from our years of experience with the elevated temperature (70°C) burn-in of GaAlAs/GaAs lasers at normal optical outputs, however, suggest that the characterizations of some infant failures are different. First, the infant failures are most often of the preordained type; the operating current to maintain a constant output increases smoothly and significantly over a short period of time (tens of hours). Second, the times to infant failures are susceptible to characterization by a lognormal distribution.[40]* Early failure distributions found during accelerated thermal aging, and consistent with lognormal characterization, have also been found for transistors[2,18,23] and Complementary Metal-Oxide Semiconductor (CMOS) integrated circuits.[41] In summary, early or infant failures have the potential for being either of the chance or preordained types.

A hazard rate signature for a lognormally distributed infant population is shown in Fig. 1d. It displays an increasing hazard rate region, which to our knowledge has never been experimentally observed. Connected with some of the reasons, already given, concerning the poor characterization of infant mortality, there is another reason why an increasing hazard rate is unlikely to be observed even if the failure distribution were lognormal. This is related to the dispersion (σ) in the logarithm of lifetimes for such a distribution. Figure 2 shows a temporal plot of the hazard rate[21] for $\sigma = 0.5$ (good quality control)

---

* In Ref. 40, the plotted data of Ref. 20 were supplemented by 14 additional early failures, and several additional long-term lifetimes.

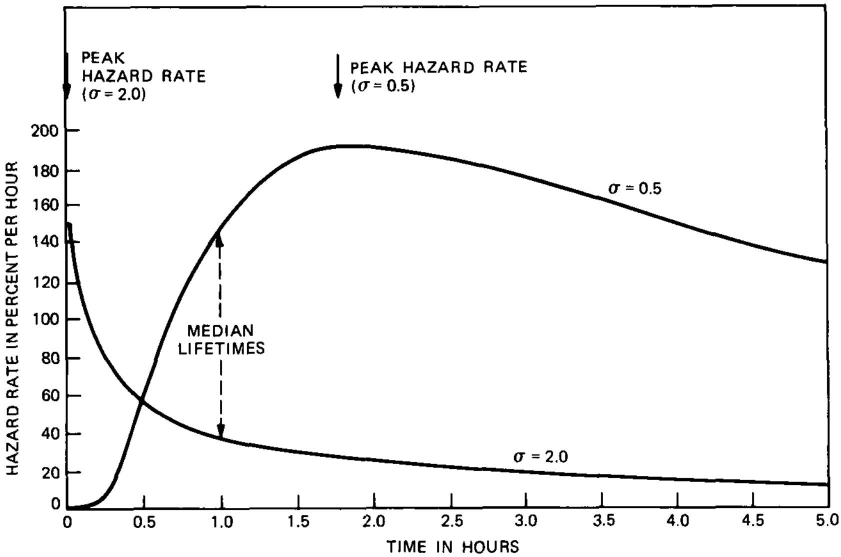

Fig. 2—Hazard rate curves for two lognormal distributions with the same median lifetime (τ = 1 hour) but with different dispersions (σ = 0.5 and 2.0). If the σ is large enough, the initially increasing hazard rate cannot be observed (see Ref. 21).

and a median infant lifetime, $\tau = 1$ hour. The peak hazard rate occurs at $t_p \approx 1.8$ hours, which on the time scale shown should be easy to detect. With poorer product quality control, a lognormal infant distribution ($\sigma = 2$ and $\tau = 1$ hour) has a hazard rate that peaks at $t_p = \tau/100 = 36$ seconds, a very short period of time in which to record failures, especially with small sample sizes. Therefore, depending upon σ, hazard rate curves can take a form that is practically indistinguishable from that of a decreasing function of time.[21]

2.5 Comparison of bathtub curves for semiconductor and human mortality

A composite of the two lognormal (see Figs. 1b and d) and the exponential hazard rate curves, Fig. 1c, is shown in Fig. 1e. Bathtub-like characteristics are displayed by the practically accessible central region (solid curve). The rollover (maximum in hazard rate) at short time, if it exists, may be impractical to detect. The rollover at long times has been seen in the failures of digital bipolar integrated circuits[17] and inferred for the failures of high-quality submarine cable transistors.[2] In general, wear-out rollover will not be seen in the field use of high-quality devices because it occurs too far out in time. The rollover at long times for the lognormally distributed wear-out of semiconductor devices is a clear departure from the analogy to the wear-out of human beings, which has an always increasing hazard rate and which is typical of a normal distribution. The steady-state or low hazard rate region (bottom of the bathtub) has been observed[17] and inferred[42] for integrated circuits. As a practical matter, the testing of semiconductor devices is usually terminated at a point (left-hand corner of Fig. 1a) where the hazard rate, though small, is continuing to decline.[19] Since it cannot fall below some random-event-determined level, it is generally taken for granted that were testing continued, an essentially constant* hazard rate would be recorded.[19] A modification of the human mortality bathtub curve (see Fig. 1a), which is similar to our Fig. 1e, has been independently proposed.[37]

2.6 Reliability assurance strategy suggested by the modified bathtub curve

We assume that some particular semiconductor device fabrication technology has been perfected to the point where a bathtub-like curve with a flat-bottomed period of useful life accurately describes the unscreened reliability of devices in field use. No matter how hard one tries to additionally perfect the design, growth, choice of materials, and processing, there will always be at least some infant mortality. Experience has also demonstrated that eventual wear-out appears inevitable for devices which were not infant failures. To the extent that the wear-out times-to-failure are credibly described by a lognormal law, then Fig. 1e, rather than Fig. 1a, is the anticipated bathtub curve; the exact shape of the infant distribution is of minor interest in constructing a reliability strategy, since regardless of its shape, the infant subpopulation must be eliminated prior to deployment.

The traditional strategy for reliability assurance, which is suggested by either bathtub curve, Fig. 1a or e, is the following. Operate (burn in) the unscreened devices in the laboratory, under the field conditions used to obtain the bathtub curve in the first instance, until the hazard rate becomes constant, *or* falls below some acceptable value, *or* becomes unmeasurably small. This burn-in procedure (accelerated perhaps by the use of an elevated temperature ambient)[43] is the usual way to censor the infant mortality subpopulation, and is widely and successfully employed.[42,43] The surviving devices are then ready for use. As the time for wear-out approaches, the devices are replaced. If replacement is not possible, then the technology must be sufficiently mature so that wear-out failures are not present during system life. This bathtub strategy does not involve characterizing the wear-out failure distribution, it only requires knowing when wear-out failures start to become important.

---

* In principle, if not in practice, an advantage in using the hazard rate, which is normalized to the instantaneous expected number (N) of survivors is that the minimum of the bathtub appears flat. If the normalization factor had been the total initial population (N_0), both the infant and chance hazard curves would be decreasing functions, making it difficult to distinguish one from the other.

(If the wear-out distribution is characterized, and if it is credibly lognormal, then an alternative strategy is specifically suggested by the rollover in Fig. 1e. The majority of the population which survived the infant screening could be aged to failure. The remaining few devices might be expected to live indefinitely if one were sufficiently far out on the declining right-hand edge of Fig. 1e.*)

2.7 Timely determination of bathtub data—accelerated thermal aging

For a mature technology whose development has produced millions of device hours of testing, data sufficient to estimate the bathtub hazard rate exist. For a new technology under a qualification constraint, not only is there an absence of bathtub data, but there is no time to acquire it directly. To appreciate the time issue, assume that only a long-term gradually occurring wear-out mode of degradation is operative and consider the subcable laser that must operate for at least 25 years = 220 kh in an ≈10°C ambient. Aging to failure at 10°C is out of the question. If $I(t)$ and $I(0)$ are the operating current and its initial value, respectively, and if one defines the lifetime (τ) as the first time at which $I(t) = 2I(0)$, then the lifetime is related to the assumed uniform rate of increase of I by $\tau\, dI/dt = I(0)$. Setting $\tau =$ 220 kh shows the initially normalized degradation rate to be 0.45 percent per kh; if $I(0) = 30$ mA, then $dI/dt \approx 0.1$ mA/kh. This would be very difficult to measure in a laboratory-convenient time period. Even if one could age at 10°C for 2.5 years, there would remain the problem of predicting behavior for a time period which was ten times longer than the test period in the face of the uncertainty about whether the aging would continue to depend linearly upon time.

An indirect approach must be found to relatively quickly obtain the required data, or at least the confidence that we know what it will look like. Fortunately, semiconductor devices are susceptible to a more sophisticated approach than the brute-force testing for long periods of time at the low-stress conditions of the intended use.† The use of elevated temperatures to accelerate failure[18,22] compresses the useful life-span of a device into the time scale of a laboratory experiment. By comparing the times-to-failure at various elevated temperatures, a thermal activation energy (acceleration factor) can be determined[44] that may permit a reliable estimate‡ of device lifetimes at much lower temperatures where actual measurements would be of intolerably long duration. The implicit assumption is that a unique temperature-time

---

* This particular strategy will be discussed again at the end of Section IV.

† This was the approach taken for electron tubes (see Ref. 18).

‡ The axiomatic difficulty with this procedure is that of showing, for example, that a prediction of times-to-failure which exceed the life-span of the data analyst will indeed come true (see Ref. 2).

relationship with a single value of the activation energy is applicable for all temperatures of interest.

Accelerated thermal aging is a key ingredient of all of the reliability assurance strategies that we shall consider. If a single mode of failure is controlling the demise of all devices at some elevated temperature, then the lifetimes may be modeled by a single lognormal distribution.[2,18,21-33] The absence of any evidence of infant failures, at the accelerating temperatures, in the cited examples[2,18,21-33] may be attributed variously to the smallness of the sample size, special selection, prescreening, or to the fact that the infant failure mode, though not controlling at the accelerating temperature, would become predominant in the relevantly afflicted devices at some lower temperature.* The random selection of large samples appears to more commonly produce a bimodal distribution,[18-20,23,27] an example[20] of which is shown in Fig. 3. Although it characterized the lifetimes of GaAlAs/GaAs lasers operating at 70°C, it strongly resembles the S-shaped lifetime distributions found during the accelerated thermal aging of transistors[18,23] and CMOS intergrated circuits.[41] It is usual[18,23,41] to define the early failure group in accelerated thermal aging as the "freak" population. The remaining larger population is assumed to be controlled by a single longer-term (wear-out) mode, since the relevant failures are lognormally distributed.[18,20,23] Where the point of inflection in the (dashed) curve is sharply defined, as it is in Fig. 3, the freak and wear-out distributions overlap, only slightly, so with some confidence it may be said[18,19,23,40,41] that the freak distribution in Fig. 3 comprised 10 percent of the total population, and that if aging at 70°C had been performed for only 100 hours, the surviving population would be controlled by a single degradation mechanism.† Knowledge of the latter point is of considerable importance, since one of the main goals of accelerated thermal aging is the determination of the activation energy of the ultimately controlling long-term mode, and confidence in the value obtained is heightened if the main population's demise is controlled by a single mechanism.

Freak failures (i.e., the earliest anomalous failures in accelerated thermal aging) have been experimentally found to be lognormally

* The fact that elevated temperature aging might not reveal the presence of a potent low activation energy infant mortality mechanism which would be significant at the ultimate device operating temperature will be dealt with in Section 3.3.

† Considerable complexity and possible misinterpretation would exist if more than two different and temporally overlapping failure mechanisms coexisted with similar activation energies. Different combinations of median lives, dispersions (i.e., the variations in the potencies of failure mechanisms from one laser to another), contributing fractions, and activation energies will produce many forms of cumulative failure plots in accelerated thermal aging regimes, and only in the event of bimodality will the sorting out be tractable.

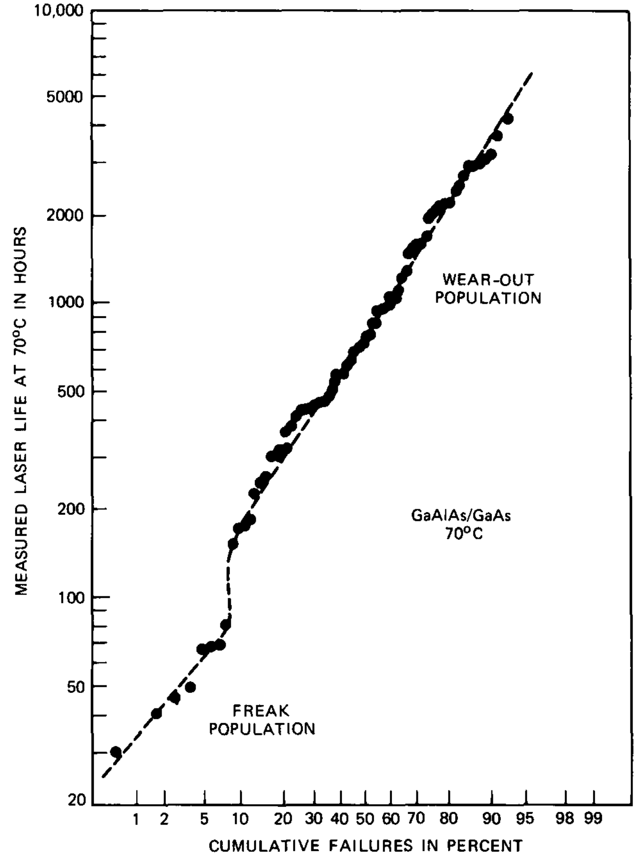

Fig. 3—Wear-out and freak distributions of the failures observed in 70°C ambient aging of (Al,Ga)As lasers (see Ref. 20).

distributed for some of the first commercially available plastic-encapsulated transistors[2] and for lasers.[40] Generally, however, not enough data are available for the unambiguous characterization of freaks, so it is often assumed[18,23,41] that such failures are lognormally distributed.

III. POTENTIAL PROBLEMS WITH THE BATHTUB STRATEGY AND THE USE OF ACCELERATED THERMAL AGING

3.1 Failure to exploit the potential of accelerated thermal aging

When only a single lognormal failure distribution results from aging studies at a single elevated temperature, accelerated thermal aging is only a characterization procedure, i.e., it permits the median lifetime (τ) and the dispersion (σ), which is the standard deviation in the logarithm of lifetimes, to be determined for what is presumptively the wear-out mode. If failure distributions (each being single) are established at two elevated temperatures, then the activation energy of the wear-out mode may be determined as well.[44] If it is plausible to assume that this activation energy applies at the lower intended field use temperature, then τ may be calculated for this temperature, at which direct measurements are not possible. In the event that it is found that τ exceeds the system lifetime in a high-reliability application, no more than feasibility has been established, a fact that we assumed to exist prior to embarking on our quest for reliability assurance. Knowing that the wear-out mode is tolerably slow acting in some unknown subpopulation of devices is not equivalent to a confident prediction of reliability assurance because of early wear-out failures and the infant mortality problem. The latter, which must exist for some fraction of a population, has not been adequately dealt with in past studies. Alternatively put, the problem of reliability assurance in a high-reliability application is, to an important degree, the problem of identifying those devices in a population for which an early (infant) mode of degradation, rather than the simultaneously present wear-out mode, is controlling at the use temperature.

Single lognormal failure distributions (i.e., absence of any anomalous early failure distribution), established during accelerated aging, probably result from inadequate population sizes,[2,18,21-27,29,31,33] preselection,[28,30] and screening tests.[24,32] Demonstrations of anomalous early failure modes[18,20,23,31,41,45] conform to the expectation that all semiconductor devices (including lasers) are potentially susceptible to non-wear-out failures and that large test populations will be required if elevated temperature aging is to achieve optimum utilization. Aging bimodal populations at several temperatures will permit the activation energies of both modes of degradation, as well as the contributing fractions, median lifetimes and dispersions to be determined.[41] (With few exceptions,[20,40,41] statistically significant data on early anomalous failures have not been recorded.) From these data a hazard rate plot (bathtub curve) may be calculated. If the infant and wear-out distributions are very well separated, as in the ideal case of Fig. 1e, then the bathtub strategy is employed, i.e., burn in the mixed population at some elevated temperature until the infant class has been eliminated, and then install the survivors.

3.2 The problem of overlap in a bimodal failure distribution

An intended use-temperature burn-in test of a large population of nominally good devices will produce, in the absence of preselection or screening, an infant population. Wear-out will generally be unobservable since it occurs largely in the remote future. Accelerated thermal aging of semiconductor devices will significantly compress the time scale in which wear-out becomes important, because its activation energy is universally found to be large (0.5 to 1.5 eV) and thus wear-out is usually viewed as a strongly thermally activated process. The question to be considered, if a bimodal failure distribution is found at an elevated temperature, is what form this distribution will assume at the lower use-temperature. If the two separate components of the bimodal distribution overlap significantly at the use-temperature, then the bathtub could appear "all filled in." There would be no obvious period of useful life. The traditional bathtub strategy (see Section 2.6) would not work.

The most realistic estimate about whether the bathtub strategy was likely to be applicable to our problem was thought to derive from the best available reliability data[20] on GaAlAs/GaAs lasers, as shown in Fig. 3. An important feature of the lognormal distribution of the long-term failure mode in Fig. 3 is its large dispersion, i.e., the standard deviation in the logarithm of lifetimes[20,40] is $\sigma = 1.1$ to 1.2. It is even larger ($\sigma = 1.3$ to 3.0) for other long-term laser reliability studies.[28-30] The range, $\sigma = 1$ to 3, has, in the past, also been typical[2,18,22,23,41] of thermally accelerated transistor failure distributions. The data of Fig. 3 have been analyzed[40] as a mixture of two lognormal distributions with the following results for the percentages, median lifetimes, and dispersions of the freak and wear-out populations, respectively: 15 percent, $\tau_f = 1$ hour, $\sigma_f = 1.2$; and 85 percent, $\tau_w = 680$ hours, $\sigma_w = 1.25$. By using these values and parameterized failure rate curves,[21] we can translate the lognormally distributed lifetimes[20] of Fig. 3 to the hazard rate plot of Fig. 4. The peak hazard rate for the freaks is ~100 times larger than that for the wear-out group.* The bathtub of Fig. 4 barely holds any water, i.e., it seems filled in. The prime question now is, what is the lower room or use-temperature version of Fig. 4?

Unfortunately, no reliability data from devices comparable to those of Fig. 3 exist at any temperature other than 70°C, so that no bathtub curve can be constructed for 30°C operation. To the extent to which the activation energy of the freak population is the same† as that of

* The reason for this follows from an order-of-magnitude estimate in which eq. (1) is applied to each population separately. The hazard rate for the freak population, where the measuring interval is the median freak lifetime, is $[(0.15\ N_T/0.85\ N_T)(\tau_f)]$, since 15 percent of the initial total number (N_T) failed, leaving a wear-out population equal to 0.85 N_T. The hazard rate for the wear-out population is $[(0.85\ N_T)/(0.85\ N_T)(\tau_w)]$. The ratio of the former to the latter, with $\tau_f = 1$ hour and $\tau_w = 680$ hours, is 120.

† The claim[17,18] has been made that transistor freaks have approximately the same activation energy (≈1.0 eV) as the main (wear-out) population. By inference the freaks are not a subset of the infant group, whose activation energies are lower.[17] Published data[41] for CMOS integrated circuits reveal activation energies for the main and freak populations to be 1.3 and 0.9 eV, respectively. The large activation energy and lognormal distribution for the freak population[41] would suggest that it was an early wear-out distribution.

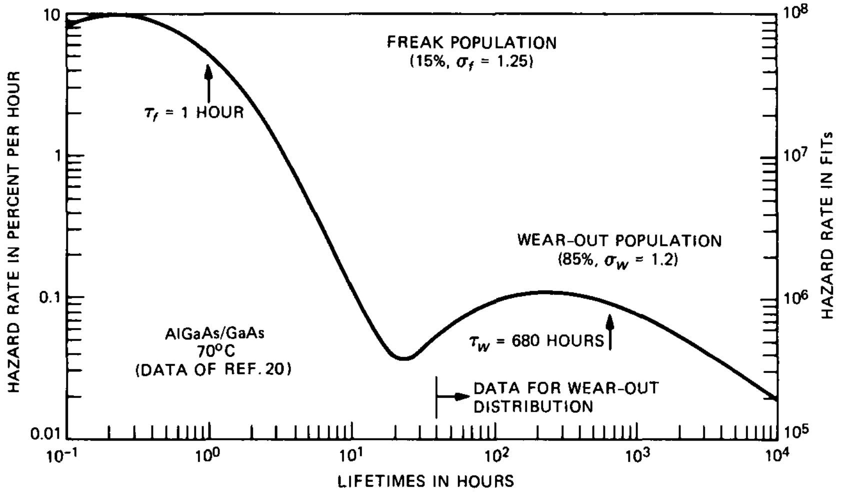

Fig. 4—The hazard rate plot corresponding to the failure distributions in Ref. 20. The contributing fractions, median lifetimes, and dispersions of the two lognormal distributions, which were necessary for the hazard rate calculation, were determined in the analysis of Ref. 40.

the wear-out group, which appears to be approximately true for certain transistors,[17,18,41] then the shape of the bathtub curve at room temperature would be approximately the same as Fig. 4. The substantial variance ($\sigma_w = 1.2$) of the wear-out group would make an early wear-out failure as important as any freak failure in an application in which even one premature failure would be disastrous. The overlap of distributions defeats any strategy for reliability assurance that is based upon accelerated thermal aging until the freaks are eliminated. No matter when the screening is terminated, that point in time will be well into the wear-out distribution, so that statistically, some fraction of the wear-out group will become an early failure group, just subsequent to any installation of the survivors. The usual bathtub strategy would therefore be inadequate in this event.

Although freak failures in the realm of transistor reliability appear to be due to high thermal activation energy mechanisms, this is not thought to be generally true in the realm of laser reliability. For example, for GaAlAs/GaAs lasers (see Figs. 3 and 4), the dominant early failure mechanism is the Dark-Line Defect (DLD).[46,47] In GaAlAs/GaAs LEDs, the DLD have been found[27,48] to be *weakly* temperature dependent, and in optical pumping of wafers, the activation energy was determined to be 0.2 to 0.3 eV.[49] (For InGaAsP/InP lasers, however, mixed results have been obtained; low activation energies [0.16 to 0.34 eV] were found for DLD[50,51] and Dark-Spot Defects [DSD],[51] while activation energies of 0.55 and >1 eV have been found[52,53] for an initial saturable mode of degradation.)

It is likely, then, that had the failure data[20] of Fig. 3 been recorded during operation at room temperature, there would have been considerably less overlap in the room-temperature analog of Fig. 4, since the wear-out mode had an activation energy $(E_a)_w = 0.7$ eV.[54] Suppose, for example, that $(E_a)_f \approx 0.2$ eV. Then the $\tau_f(70°C) = 1$ hour (see Fig. 4) would become $\tau_f(30°C) = 2.4$ hours; in contrast, $\tau_w(70°C) = 680$ hours (see Fig. 4) would become $\tau_w(30°C) = 15{,}368$ hours using $(E_a)_w = 0.7$ eV. Thus, the evidence from studies of specific failure mechanisms would suggest that the so-called freak population of Fig. 4 is just a low activation energy infant population whose hazard rate curve is essentially independent of temperature. Note, however, that even though there is considerably more temporal separation between the hazard rate curves for the infant and wear-out distributions at room temperature, the median wear-out lifetime is only ~2 years. Thus one would still expect a significant number of wear-out failures during some normal system lifetime (e.g., 10 years). The conventional bathtub strategy is not then anticipated to be applicable in our high-reliability use.

3.3 The problem of the low thermal activation energy mode that is not detected in accelerated thermal aging

Low thermal activation energy degradation mechanisms appear to play important roles in early failures of AlGaAs/GaAs and InGaAsP/InP lasers and LEDs.[27,48-51] Since temperature is not an important accelerant for these mechanisms, the possibility exists that the occurrence of these mechanisms in some fraction of a device population would be invulnerable to detection regardless of the temperatures at which accelerated aging was performed. The finally acquired bathtub data might then encourage misleadingly optimistic predictions about performance at a lower use-temperature, when, in fact, behavior was significantly controlled in midlife by an unapprehended mode.

To give one example[2] of what is perhaps the major nightmare of any reliability assurance scheme, consider the schematic distributions in Fig. 5a, which exist in an elevated temperature accelerated aging regime. Two lognormal distributions (dashed lines) are shown. Low activation energy infant failures occur in only 10 percent of the population. This infant distribution is assumed to have a dispersion larger than the long-term (wear-out) mode of degradation, and solely for the purpose of making a point, the activation energy of the infant population is assumed to be equal to zero. The times-to-failure are plotted on the ordinate. The degradation of the remainder of the population (90 percent) is controlled by a lower dispersion, thermally activated long-term mode. If the infant failure mechanism had not existed in 10 percent of the population, then the degradation of that

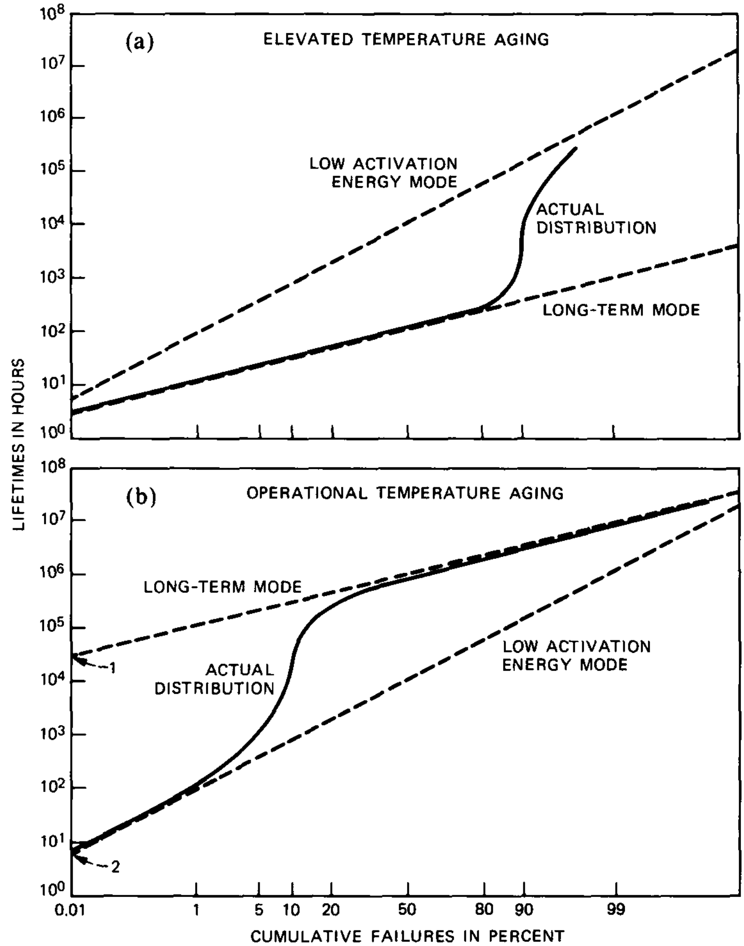

Fig. 5—(a) Hypothetical lognormal distributions (dashed lines) of failure caused by either of two mechanisms during elevated temperature aging. The short-term mode has a dispersion higher than the long-term mode, it is present in 10 percent of the devices, and it is assumed to be nonthermally activated. The failure of 90 percent of the devices is governed by the long-term, or wear-out, mode which is assumed to be thermally activated. Overall, degradation is dominated by the long-term mode beyond the 50-percent cumulative failure point. The short-term mode will remain undetected if testing is terminated at this point. The actual, or composite, distribution is shown as a solid curve. (b) At the intended use or operation temperature, the unaltered short-term mode actually dominates the earliest failures (2), even though it was predicted from (a) that the earliest failures would occur at (1).

10 percent would also have been governed by the long-term mode. A reasonable screening scenario would terminate the elevated temperature testing when the cumulative total failures had reached 50 percent or just beyond.[25] The actual distribution is shown as a solid curve. From the failure results up to the median lifetime (100 hours in the

example), it might have been concluded that a single failure mechanism, the one with the lower dispersion, was controlling.

By going to still higher temperatures, with another lot of nominally similar devices, the lower dashed curve would be downshifted to shorter times, and an activation energy would be found in the usual way.[44] The upper dashed curve would remain fixed. Extrapolation to lower operating temperatures would lead to a prediction that the first failure would occur after 10^4 hours* (see Fig. 5b). The distribution that would actually be seen if a population were aged for a long time at the intended operating temperature (solid curve in Fig. 5b) would, however, show that 10 percent of the devices failed prior to 10^4 hours.

The elevated tem)erature burn-in has been widely and traditionally used to compel a timely identification of infant failures prior to deployment.[43] Conceptually, at least, as we have noted, this has a mindless aspect in view of the knowledge that the infant failure modes of integrated circuits,[17] LEDs, and lasers[46,51] can have low thermal activation energies. The conceptual error is in the exclusive use of temperature as an accelerant for degradation or failure mechanisms that are relatively insensitive to temperature. Our discussion of Fig. 5 has shown that an infant mode may not be detected in an elevated temperature burn-in. Figure 5a also reveals that an elevated temperature burn-in, used as a screening procedure, may produce a complete disaster, even though it was a success in another case (e.g., the GaAlAs/GaAs data of Fig. 3). Thus, if the hypothetical lasers of Fig. 5a had been burned-in until the point of inflection of the S-shaped curve, with the view that all failed lasers were infants and all survivors were long-lived, the actual result would have been that all long-lived lasers would have been eliminated by the screen and all of the survivors would have been infant failures at the lower operation temperature.

An alternative view of the low activation mode problem appears in an Arrhenius plot. It is common[2,18] to assume that the Arrhenius law governs the dependence of median device lifetimes upon temperature (other degradation accelerants being held constant) through

$$\tau = C \exp\left(\frac{E_a}{kT}\right), \tag{3}$$

where E_a is the activation energy, k is Boltzmann's constant, and C is a constant independent of temperature.† Rearranging (3) gives

* This statement is true if the initial population contained 10,000 lasers, so that the first failure would occur at the 0.01-percent cumulative failure point.

† The Eyring equation in which C is a function of temperature and which includes other multiplicative exponential functions that take into account the role of other accelerants, and their interaction with temperature, has been used (see Refs. 41 and 55). For our purposes, (3) is adequate.

$$\frac{1}{T} = \frac{k}{E_a} \ln \tau - \frac{k}{E_a} \ln C, \tag{4}$$

which when plotted as $(1/T)$ versus $\ln\tau$ gives a straight line. For convenience, we desire to retain the straight line plot, but choose to label the ordinate in T; in this case the ordinate is *not* linear in temperature. A schematic plot of the median lifetime of the wear-out mechanism as a function of T is shown as a solid line in Fig. 6a. The darker portion, bounded by the two test temperatures (T_1, T_2), chosen so that failures occur in laboratory-convenient times, is the part actually determined experimentally; all else is extrapolation. Were wear-out the only degradation mechanism present, lifetimes of 10^6 hours at 10°C would be hypothetically possible.

Suppose, however, that in certain devices there also exists a degradation mechanism that has a low activation energy for thermal acceleration. Three examples of degradation behavior (differing potency) associated with such a low activation energy mechanism are shown as nearly vertical lines in Fig. 6b. These lines are not composite curves, but are drawn as though each low activation energy mechanism was the only degradation mechanism present. The presence of a mode whose potency is represented by curve a interferes with the determination of the activation energy of the wear-out mode. The determination of the activation energy for curve a would lead to predictions of unacceptably short lifetimes at the intended use-temperature. The existence of some devices, not affected by a curve a, indicates that a mode represented by curve a is not fundamental. Curve b does not prevent the determination of E_a, but the undetected presence of curve b would mean that lifetime predictions based upon extrapolation of the wear-out curve to 10°C would be too optimistic. Curve c is benign.

It is important to emphasize the fact that the presence of curve b in Fig. 9b is invulnerable to detection by any accelerated thermal aging strategy. For the temperature range in which it could be the controlling degradation mechanism, aging takes an impractically long period of time, and it consumes a considerable fraction of the useful lifetimes of good devices. Insidious low activation energy modes that can invalidate lifetime predictions based on high temperature aging have been contemplated[2,24,56] and observed.[25,57,58] In laser reliability studies, evidence[33] suggests that different degradation mechanisms can dominate at different temperatures.

For completeness, the situation pictured in Fig. 6c should be anticipated, i.e., nonfundamental modes whose activation energy exceeds that of the wear-out mode. A family of such modes with varying potency is shown as curves d, e, and f. Curve d is an early failure, curve f is without consequence, and curve e would obscure any deter-

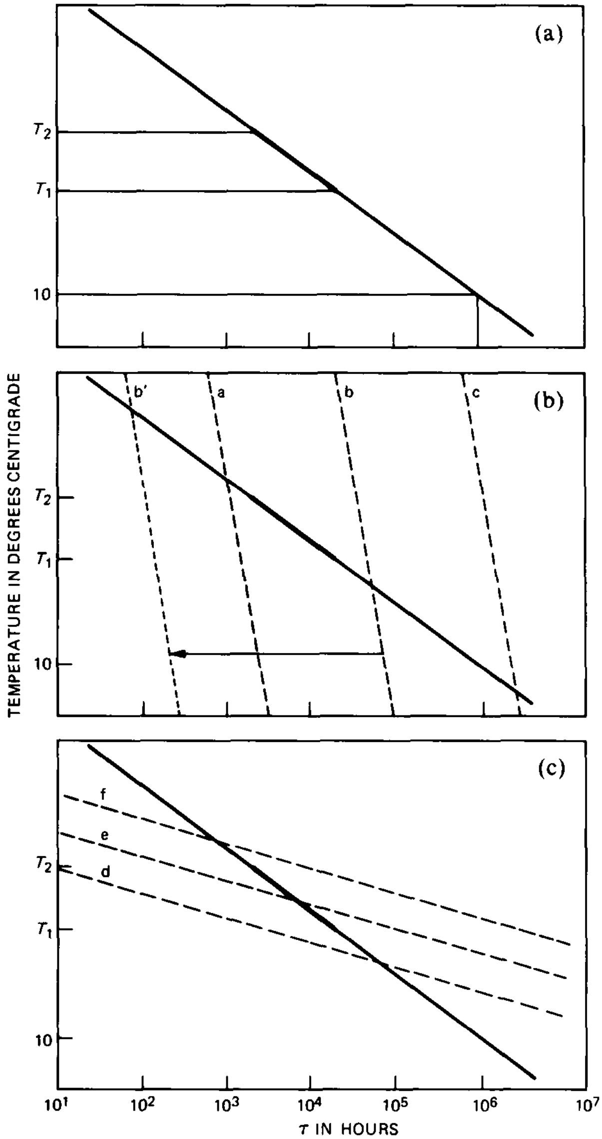

Fig. 6—(a) Hypothetical plot of the median lifetimes of wear-out-mode-dominated failures as a function of the device temperature. The actual lifetime measurements to determine the activation energy (E_a) for this thermally activated long-term mode are conducted in the temperature range $T_2 - T_1$. The remainder of the curve is established by a straight-line extrapolation of the measurements obtained between T_1 and T_2. (b) Superimposed upon (a) is a family of lifetime curves corresponding to varying degrees of potency of a weakly thermally activated nonfundamental mode, i.e., one whose activation energy is less than that of the wear-out mode of (a). The shift of line b to b′ represents a selective acceleration of the nonfundamental mode relative to the wear-out mode. (c) The same as (b), except that a higher activation energy nonfundamental mode with varying degrees of potency is superimposed on the wear-out mode failure curve a.

mination of the activation energy for the long-term mode. The obvious way to prevent curve e from introducing ambiguity in establishing the activation energy is to increase the test temperature range, so that curve e becomes like curve d.

The solution to the undetected low activation energy mode problem is to find its appropriate accelerant. This will be examined later in Section 5.2.

3.4 The problem of the single lognormal wear-out failure distribution with large dispersion

Let us imagine that a mixed population has been successfully purged of the infant subpopulation, which includes all low activation energy modes that might have caused premature failure in field use. The survivors will then presumptively follow a single lognormal wear-out failure distribution. In this section we show that the consequence of a large dispersion for wear-out is that there is no period of useful lifetime for a population, the hazard rate is a decreasing function of time for all times of interest,* and hence the traditional bathtub strategy will be inapplicable.

Consider, for example, the AlGaAs/GaAs laser failure time data[29] recorded at 55°C. The dispersion was found to be $\sigma = 3.0$. It has been assumed[2,20,22,28,30,41] and experimentally[18,32] determined that σ is not a function of temperature when a single mode of degradation is controlling. Using $E_a = 0.7$ eV[29,54] for the wear-out data, the hazard rate curve at 70°C may be calculated and it appears in Fig. 7. This plot should be compared with Fig. 4, where the scales are exactly the same; Fig. 7 represents what might have been recorded if the thermal acceleration had occurred at 70°C. Although the median wear-out lifetime is larger in Fig. 7, as compared with Fig. 4, there is no practical duration of time in Fig. 7 in which the hazard rate is relatively constant, i.e., there is no period of useful life as contemplated by the flat-bottomed portion of Fig. 1e.

The operational significance of using accelerated thermal aging as a first step in a screening technique is that it enables a reasonably prompt location of the first time at which a minimum hazard rate is nearly reached. In an ideal case, this corresponds to the left-hand edge of the bottom of the bathtub, Fig. 1e. At this point, it is hoped, most of the devices are still operative and are ready for installation. By contrast, accelerated aging for the devices represented in Fig. 7 has shown that an acceptably low hazard rate may only occur when all but a few of the devices have failed.† The situation would have been

* A hypothetical example of this has already been given in connection with Fig. 2.

† As noted in Section 2.6, this observation suggests an alternative strategy, which will be considered at the end of Section IV.

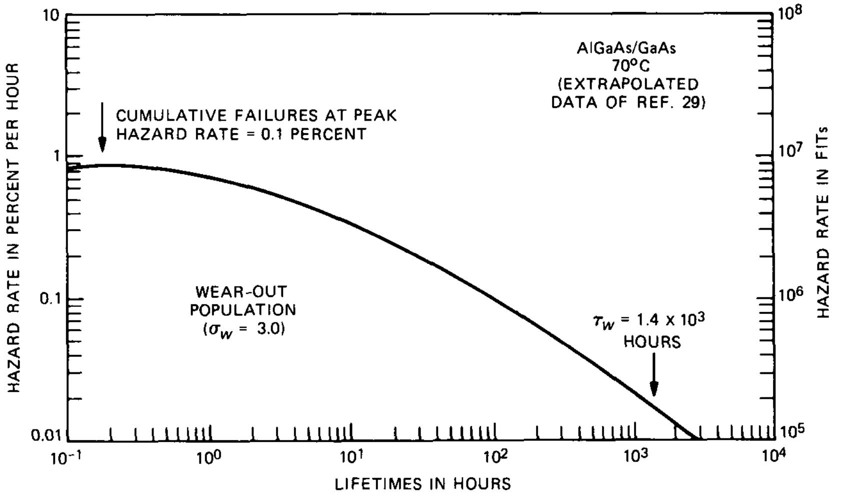

Fig. 7—In Ref. 29 an experimental lognormal failure distribution was established in a 55°C ambient for AlGaAs/GaAs lasers. If those data are extrapolated to 70°C using an activation energy equal to 0.7 eV, the corresponding hazard rate curve (shown above) may be calculated from the deduced values of median lifetime and dispersion.

no better, if instead, the failure data[29] had been accumulated at some reasonable operational temperature, e.g., 30°C, as may be seen in Fig. 8. If, for example, a failure rate of 100 FITs* is acceptable in some 30°C application, then the population represented in Fig. 8 would have to be aged for ten times the median lifetime ($\tau_w = 3 \times 10^5$ hours), at which point 75 percent of the devices would have already failed.

It seems, however, that such data[29] are indicative of poor control in growth and processing. It is much more usual to find[2,18,20-23,30] that the earliest lifetimes appearing in single lognormal failure distributions produced by accelerated thermal aging lie on the increasing portion of the hazard rate curve, e.g., Fig. 4. A particularly good example of the right-hand side of the bathtub curve comes from the 30°C aging of AlGaAs/GaAs lasers. We randomly selected nine lasers from a "tested" wafer and aged them at 3 mW per facet without any prior burn-in. Seven lasers failed, the first at 9365 hours and the seventh at 34,929 hours. Two remained operative at 45,484 hours when the aging was terminated. The lifetimes were lognormally distributed with a σ = 0.54. All seven data points appear on the increasing portion of the hazard rate curve in Fig. 9. Figures 8 and 9, which have identical

* One FIT is one failure in 10^9 device hours, so that one failure in a population of 10^4 devices operated for 1 year (8760 hours) is a failure rate equal to 11 FITs. For purposes of comparison, 10 FITs = 0.001 percent/kh.

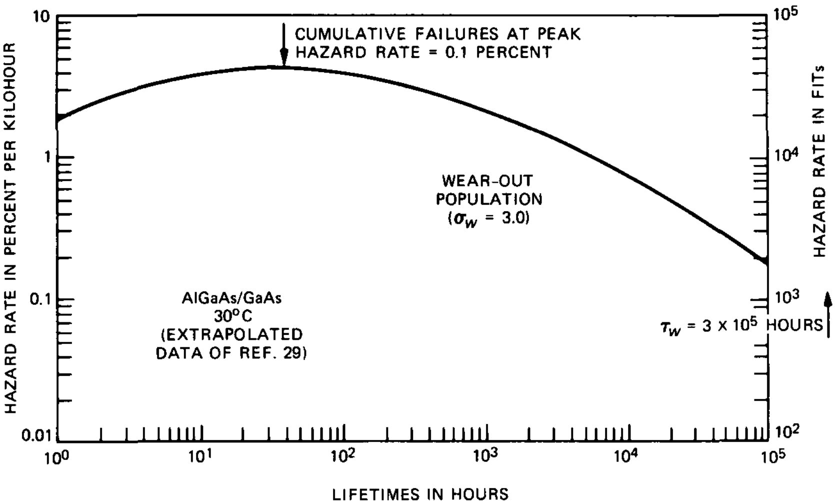

Fig. 8—If the 55°C lognormal failure data of Ref. 29 are extrapolated to 30°C using an activation energy equal to 0.7 eV, the corresponding hazard rate curve (shown above) may be calculated from the deduced values of median lifetime and dispersion.

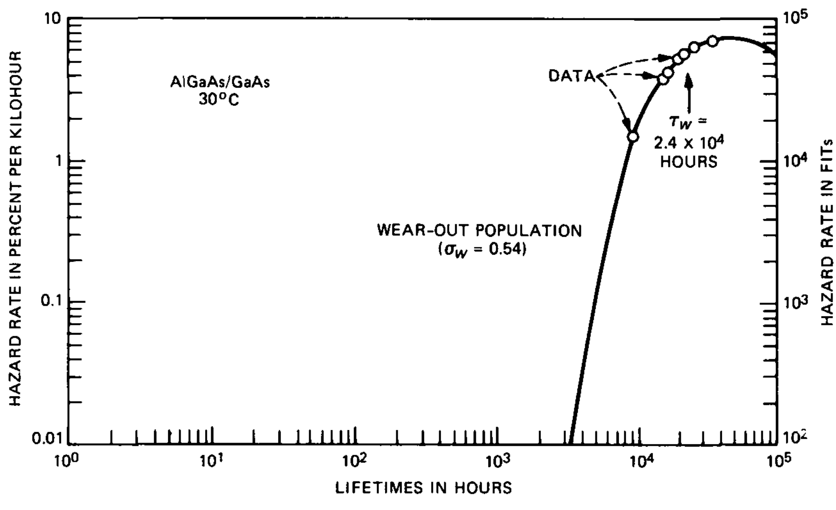

Fig. 9—Failure data (points shown) of AlGaAs/GaAs lasers at 30°C lead to a median lifetime and dispersion which enable the above shown hazard rate curve to be calculated.

scales, should be compared. Thus, there is likely to be at least some period of useful life, prior to the manifestation of wear-out, if σ is sufficiently small; the period of useful life would not appear to be large enough, in general, however, to permit an uncritical application of the bathtub strategy.

3.5 Summary of possible difficulties in the use of a bathtub strategy with an elevated temperature burn-in to censor infant failures

The potential or actual deficiencies of the conventional burn-in approach to using the bathtub strategy for reliability assurance are the following:

1. A reasonably complete distribution of the times-to-failure is required for a large population of devices so that all significant failure modes have some representation in the distribution. This might be impractical to obtain because of the long periods of aging required, even at elevated temperatures.

2. Assuming that it is possible to obtain such a distribution, it will very likely be of no use if it is not bimodal (see, e.g., Fig. 3). Not only will no obvious screening strategy be apparent if the distribution is multimodal, but it will also appear impossible to achieve a segregation of those devices whose lifetimes are governed solely by wear-out. Segregation is needed in order to determine the thermal activation energy for wear-out, which would then permit lifetime predictions at lower temperatures to be made.

3. Assuming that it is tractably bimodal, distributions at other aging temperatures are required in order to avoid the possibility that an elevated temperature burn-in will promote failure in the population that would be long-lived at a lower use-temperature, while retaining for deployment an infant failure group (see Section 3.3 and Fig. 5). If distributions of times-to-failure at low use-temperatures are required, deficiency (1) becomes even more of a problem.

4. Again, assuming tractable bimodality, any overlap of the infant and wear-out hazard rate curves at the use-temperature would mean that some wear-out failures would occur during system life.

5. Assuming that it has been shown that a burn-in at a specified temperature and for a specified duration can remove infant failures in an early-manufactured population, it will remain uncertain whether the burn-in will remain as effective when applied to subsequently manufactured populations. New device technologies evolve rapidly and changes in processing often have a significant impact on reliability.

6. Assuming that the first-established burn-in conditions remain temporally "correct," the time-to-effect may prove to be excessive. In an actual example,[52,53] ~1 kh of elevated temperature burn-in aging was required to establish a stable surviving population. Long screening times might require maintenance of a considerable number of expensive aging sockets. If the duration of the burn-in were to be arbitrarily restricted to a short duration (~10^2 hours), disaster might follow.[52,53]

7. Finally, it is noted that long-term wear-out modes of degradation have high thermal activation energies, while infant mechanisms have lower activation energies. The only degradation accelerant available

in the conventional burn-in procedure is temperature. However important it may be to compress the life-spans of devices into laboratory-convenient time periods by using an elevated temperature burn-in, it is possible that the wear-out mechanism becomes relatively more accelerated so that a considerable fraction of the useful life of a good device, one free of aberrant failure mechanisms, may be consumed by the usual burn-in screen.

IV. THE LOGNORMAL MODEL OF WEAR-OUT—AN IMPROVEMENT OF THE BATHTUB STRATEGY

As we have seen, wear-out failures are very likely to occur during a laser system lifetime. Consequently, a conventional bathtub strategy will not be applicable, even if it was successful in eliminating the infant subpopulation. A considerable improvement in establishing long-term reliability is provided by the lognormal statistical approach[24–26,34,59,60] that seeks to characterize the wear-out failure distribution by a lognormal failure law. No procedure is provided in this scheme[24–26,34,59,60] for purging the population of those devices whose degradation or failure is controlled by non-wear-out modes. An important assumption is, therefore, that an infant subgroup either did not exist, or was eliminated by a burn-in or some other means. The starting point of this analysis is similar to the case of Section 3.4, where the conclusion was that if σ (wear-out) was too large, the simple bathtub strategy failed. The natural extension provided by the lognormal statistical approach uses the hazard rate curve computed from experimentally measured values of the median lifetime τ and σ. The confidence level of the hazard rate predictions made for early time failures, where no data may exist, follows from the application of known statistical theory.[60] One can then construct replacement or redundancy schemes that allow for significant numbers of early failures if σ is large. The uncertainties of the predictions based upon inadequately large test populations have been noted.[59]

In the event that early wear-out failures are significant enough to thwart the use of the traditional bathtub approach, as assumed in this section, then the bathtub and lognormal approaches could conceivably be combined as follows. Burn in the starting population under ordinary operating conditions until the hazard rate becomes too low to measure. Presumably at this point, the infants have been eliminated. By means of accelerated thermal aging, operate the remainder of the devices until they all fail, and characterize the presumptively single lognormal wear-out failure distribution (i.e., determine E_a, σ, and τ). Then calculate the hazard rate curve for the use-temperature and extend it back in time to make statistical predictions about early wear-out failures, i.e., the failures that would have occurred soon after installation if there had been no infants.

The traditional bathtub strategy, which relied upon a useful lifespan in which no wear-out failures occurred, does not require that the wear-out failures be characterized. When early wear-out failures can cause a problem, some kind of statistical characterization of the wear-out population becomes imperative, so that compensating redundancy or replacement schemes can be instituted. This characterization approach is the only useful procedure in the event that wear-out failures are of the sudden or catastrophic kind,[24-26] because of the absence of any warning, i.e., any recognizable precursor degradation. (In the next section we shall see that laser wear-out failures are *not* of this type and thus an alternative strategy may be employed.) The lognormal approach[24-26,34,59,60] to characterization that we have highlighted is only one in a class of sampled-population or population-inferential approaches, any one (e.g., Weibull) of which might do equally well.

Despite the improvement in reliability assurance provided by the addition of the lognormal approach, or any similar population-inferential scheme, there remain several shortcomings thereof, which we hope to circumvent by our proposal (see Section V). These drawbacks are (1) the hazard rate curve for the wear-out mode is extrapolated to times less than those at which the first actual failures of the test population would have occurred at lower use-temperatures, thus demanding a complete prior elimination of infants if the resulting predictions are to be credible; (2) the failure predictions are strictly valid, if even then, for only the aged-to-failure population, and their value will be more uncertain for a subsequent population scheduled for installation, which although nominally similar, may be quite different because of inadvertent processing variations; (3) test populations that are small by necessity reduce the confidence of the lognormal approach, partly because infant mortalities will not be apparent in the failure distribution, and partly because uncertainties in the determinations of τ and σ can have a large effect on calculated hazard rates;[59] and (4) in the event that the σ (wear-out) is large, and τ at the use-temperature is comparable to the desired system lifetime, the inevitable statistical predictions of a large early wear-out failure rate could well prevent, perhaps unnecessarily, the system installation.

Objection (4) above relates directly to Figs. 7 and 8, in which the hazard rates might be excessive for all periods of time shown. A system could be viewed as impractical because of the expenses associated with redundancy or replacement. As noted at the end of Section 2.6 and in Section 3.4, this objection to system installation might be avoided by adopting the following procedure. If characterization of wear-out failures is performed until all devices in an adequately large population suffer failure, and if the failure law is credibly lognormal for all times, then it is always possible to increase the mean life to any extent desired by continuing to age the population until a sufficiently large number of the devices have failed.[61] Beyond the time at which the wear-out hazard rate is a maximum—in, for example, Fig. 8—the hazard rate decreases indefinitely. Thus, if one is willing to age devices for long periods of time prior to deployment, suffer failure of a major fraction of the initial population, and believe that the lognormal law is accurate in the extreme long-time tail of the distribution, any or all of which could be viewed as unacceptable, then the survivors would have a negligibly small hazard rate and be suitable for installation. As a practical matter, this would appear to be an economically unattractive way to circumvent objection (4).

V. THE PROPOSED APPROACH—STP (STABILIZATION, TRUNCATION, PURGE)

Using ideas that were anticipated by efforts on previous submarine cables,[62–65] we want to establish a plan that will permit the confident prediction about which particular lasers in a given initial population will possess lifetimes which exceed that of a system in which there is no redundancy and where replacement is impossible. It should be applicable to small initially available populations and make optimum use of existing knowledge of failure modes. Except where the scheme detects flaws whose elimination by alterations in the device fabrication is feasible in a timely fashion, the plan is of the black-box type, in that it should not rely on building absolute reliability into the device (an impossible task in any event), via the physics-of-failure approach, but should instead screen all devices intended for installation, in a selective and tailored manner, to eliminate all premature failures; i.e., all non-wear-out and all early wear-out failures. A road map of the proposed scheme is given in Table I.

5.1 Stabilization

We recognize that for lasers, certain modes of degradation (annealing) are transient in nature,[53,66] i.e., after some period of aging the degrading centers may become spatially homogenized (the defects may be gettered), and these modes may exist to some extent in all lasers. We understand that one purpose of a reliability assurance scheme is the stabilization[53,66] of transient modes. In general, devices may not be eliminated from the population by the process of stabilization. However, if in the course of stabilization, which may involve varying degrees of a temporally finite amount of degradation, certain initial properties, e.g., threshold, are caused to exceed specification, those devices will be eliminated, even though thereafter they may be capable

Table I—The stages, purposes, and procedures by which laser reliability assurance is established. The overall goal is uniformity of behavior, i.e., to obtain a population whose degradation, in the simplest case, is governed by a long-term wear-out-type mode, and which contains only the tolerably slow-acting manifestations thereof

Regimes for Establishing Laser Reliability Assurance, and Uniformity of Aging Behavior		Purposes	Procedures
Stabilization		To stabilize, in a timely manner, transient modes of annealing and degradation, which may be present to some extent in all lasers	Overstress aging (current, temperature, optical power, reverse bias, etc.)
Rejection	Purge	To eliminate, in a timely manner, potential premature failures due to all modes of degradation, other than the long-term wear-out mode.	Cosmetic inspection Initial and final L-I-V characterization Overstress aging (etc.)
	Truncation	To eliminate premature failures of the wear-out type	Rate monitoring

of adequate longevity. A procedure to compel rapid stabilization is important[53,66] so that unambiguous determinations may be made of the activation energy and degradation rates of the longer-term wear-out mode. The procedure for forcing a timely stabilization of some mode is similar to one regime of the purge, which we consider next.

5.2 Purge

We understand the purge[53,66] regime to include any method (cosmetic inspection, light-current characterization, overstress aging, etc.) by which actual or potential non-wear-out failures (flawed devices) are eliminated from the population. Emphasis is placed upon overstress aging,[53,66] and in particular upon three component concepts—harshness, selectivity, and tailoring.

The first is that of harshness. There is often present in the testing of new, initially expensive, and perhaps few in number, semiconductor devices, the "don't hurt my baby" syndrome. In our view, such an attitude is inconsistent with providing reliability assurance in a timely fashion. Well-made lasers can withstand operation at currents equal to 50 times threshold without adverse consequences.[53] Excessive stress levels, however, will destroy both potentially good and bad devices. The problem is to find the critical level. When found, stressing can be performed at slightly reduced levels. This procedure should produce the most robust class of survivors, in the shortest time. While temperature overstressing is well known,[18,24,25,27,41,48,51] current overstressing in LEDs,[27] current (or optical power) overstressing in lasers,[50,51,67] and voltage overstressing in photodiodes[68,69] have also been used with success in establishing reliability assurance by eliminating early failures. Overstressing using temperature cycling, thermal and mechanical shock, and humidity is well established in the semiconductor industry[8,41] as a means of producing a robust population.

Another concept is that of selectivity.[56,66] The physical basis for any reliability prediction is related to the fact that the failure of a semiconductor device is caused by interactions between the device and its electrical, thermal, and mechanical environment. Unbiased devices in hermetic packages generally last for an extremely long time at room temperature in a tranquil setting. The concept of a mode of degradation that is not accelerable by some means is therefore physically untenable. Thus, one function of a reliability assurance scheme is to discover degradation or failure driving mechanisms or accelerants, other than temperature, that will cause, if possible, a low thermal activation energy mode (curve b in Fig. 6b), which may be undetectable in a pure thermal acceleration regime, to become the dominant mode of degradation[56] (curve b′ in Fig. 6b). The notion of *selectivity** refers to

* Selectivity also comprises the idea that particular attention should be paid to those stresses that are likely to be present in the operating environment.

the fact that the chosen stress will accelerate the low activation mechanism, but not accelerate the long-term, higher-activation energy mode as much. This expectation is reasonable for AlGaAs/GaAs lasers because, for example, the low thermal activation energy mechanisms are strongly current-density dependent,[27,48] while the long-term, gradual wear-out mode is very weakly current-density dependent.[48,70] Similar strong current-density dependence of DLD and DSD formations have been found for InGaAsP/InP devices;[51,71] it is reasonable to suppose that the wear-out mode may be analogously weakly current-density dependent.

A third concept involves the idea of tailoring the stresses. The approach that should be taken to identify the premature non-wear-out failure population cannot be a straightforward elevation of every thermal, environmental, electrical, and mechanical stress ever devised. An example of the conceptual inadequacy of such a procedure is the high-current, high-temperature operation of the etched planar mesa, buried heterostructure,[72] in which the current flows around the active region in a shunt path.[53] In order to overstress the active region, lower temperatures and currents must be used.[53]

The purge regime contemplates an application of harsh, selective, tailored stresses (ideally, used at levels just shy of those at which good devices would be annihilated) for only short periods of time ($\sim 10^2$ hours) in order to avoid consuming too much of the useful lifetime of devices. This regime should produce both stabilization of transient modes of annealing and degradation, and the failure of devices whose mortality is controlled by the aberrant non-wear-out mechanisms. Stabilization is a product of the overstress regime of the purge, and may occur in all lasers;[53] it was distinguished in Table I in order to make clear the difference between it and the process of rejecting failures (or selecting the survivors), which is the goal of the purge and the truncation stages.

Unlike the bathtub strategy, the purge does not require that a distribution of times-to-failure be established at an elevated, or any other, temperature. Were a lifetime distribution established, it would be irrelevant, in principle, whether the distribution were tractably bimodal or intractably multimodal. The purge strategy, which anticipates the worst, is to fashion a screening regime so strenuous that it does not require alteration as the processing technology evolves. By employing harsh levels of potential degradation accelerants, in addition to and including temperature, the duration of the purge may be kept short.

5.3 Truncation

A fact which is of great significance, and the basis for the truncation selection scheme, is that the wear-out modes in lasers are characterized by gradual degradation rather than sudden failure, as is the case with other semiconductor devices.[24–26] This frees us from the limitations of the lognormal statistical approach, and permits the use of the rate monitoring[20,31,33,45,55,73–75] of degradation (e.g., a recording of the increase in a laser's operating current to maintain a constant output power) to estimate lifetimes of individual lasers.* When the end of life is not accessible even in accelerated thermal aging, rate monitoring is crucial. An obvious advantage of the rate-monitoring scheme exists for the case in which the wear-out lognormal distribution has a large dispersion (see Figs. 7 and 8) and only a moderate median lifetime. Rate monitoring attempts to determine as accurately as possible which survivors of the purge and stabilization procedures will be early wear-out failures and which will outlive the intended life of the system. Provided that yield was not a problem, the individual-interference approach could legitimately favor the installation of a system which the lognormal approach would have condemned. The individual-inference approach is not concerned with determining the τ and σ of the lognormal failure distribution nor with the calculation of hazard rates. It is only necessary that a sufficient number of lasers have individually projected lifetimes which exceed the reliability requirements of the system.†

Figure 10 is a simplified schematic in a population-inferential format of what an individual-inference reliability strategy based upon truncation hopes to accomplish. The initial population is assumed to be composed of both infant and wear-out-dominated failure modes. Accelerated aging to failure at elevated temperatures would produce the bimodal distribution, shown as the lower curve. If the same population had been subjected to purge overstressing to remove the infant population prior to aging, the middle curve, a single lognormal distribution, would have resulted. But, if in addition to overstressing for short time periods, a longer time degradation rate monitoring was subsequently imposed to remove the early wear-out failures, then the upper curve, now no longer lognormal, would have resulted. Note that while the median lifetime has not been substantially increased, there is a great

* Quantitative procedures for making lifetime estimates based upon observations of early wear-out degradation rates are discussed in Refs. 44 and 76.

† For applications with less than the most demanding reliability requirements (e.g., because replacement and/or redundancy are economically feasible), it may become important to minimize the cost of the device. Elaborate screening is then too expensive.[43] The focus may shift back toward the physics-of-failure approach and procedures to increase yield. The lognormal statistical approach is used even though it is possible, in principle, to monitor the gradual degradation of devices in order to be able to estimate individual failure times.

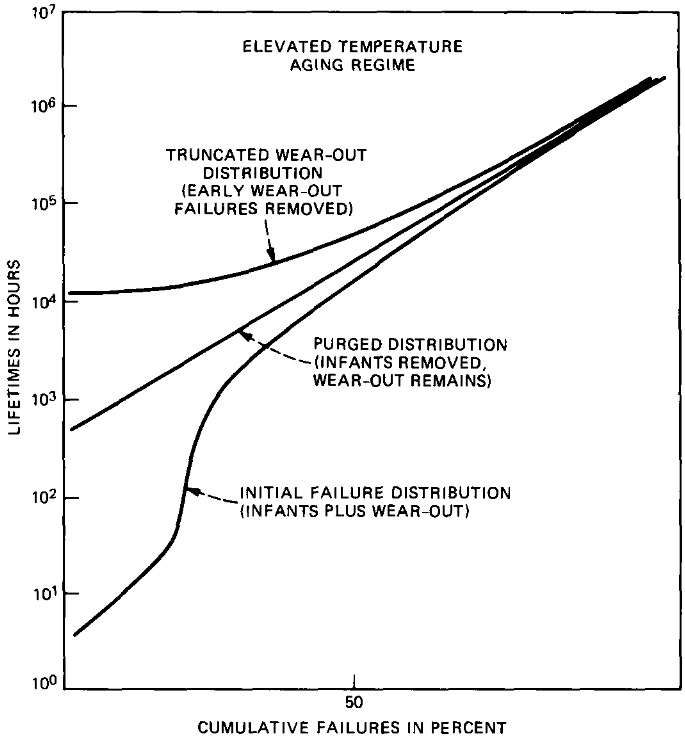

Fig. 10—The S-shaped bimodal distribution (bottom curve) contains both infant (or freak) and wear-out failures. With the infant failures removed as a result of harsh overstress testing, the single lognormal wear-out distribution (middle curve) is expected to occur. Using rate monitoring, to eliminate the early wear-out failures, it should be possible to produce the screened wear-out distribution (top curve).

reduction in the number of failures prior to $\sim 10^4$ hours, the hypothetical system lifetime at the elevated temperature. Rate monitoring, for example, during a short duration elevated temperature burn-in, is not by itself adequate as a reliability assurance strategy since it might not compel a timely stabilization of transient modes, nor might it detect sudden low thermal activation energy infant failures.

VI. SUMMARY AND CONCLUSION

It is understood that the spot checking of lasers is unacceptable as a means to assure reliability. Every laser of initial subcable quality must pass through a screening procedure that schematically consists of two active aging regimes. In the first, the various degradation accelerants (e.g., optical power, current, temperature, humidity, etc.) are elevated, whether singly or in concert, to high levels, for relatively short periods of time ($\sim 10^4$ hours). In this overstress regime, saturable modes of annealing or degradation should become stabilized and non-wear-out controlled devices should be identifiable by their large degradation rates or actual failures. In the second regime, longer-term aging is performed, with sufficient thermal acceleration so that clear degradation (beyond experimental error) occurs in times $\sim 10^3$ hours. This rate-monitoring stage serves three purposes: (1) It enables the rates of degradation of the survivors of the harsh overstress stage to be compared with the rates of degradation of nonoverstressed lasers so it can be demonstrated that the purge did not introduce a mode of degradation which would not otherwise have been present. (2) It permits determination of rates of long-term degradation due to the wear-out mode, which is present in each laser in different degrees, and hence the extrapolated times-to-failure can be estimated, which permits early failures in the wear-out population to be identified. (3) It may serve to expose any mode of degradation with a long incubation period which is not eliminated by the purge, thus compelling the purge regime to be suitably modified where possible, so that the undetected mode becomes susceptible to identification (corrective feedback). Overall, these regimes will serve to reject some class of lasers and to sort the stabilized survivors.

We desire that the harsh overstress tests be necessary (every aspect of that test should be directed at some mode), sufficient (all possible modes are detected), and restrained (good lasers are not eliminated). We assume that all modes, except for random failures, are accelerable by some means, and hence identifiable. The challenge is to find the correct strategy for the particular device. The questions that should be addressed in any first practical implementation are (1) Does the screening eliminate all weak lasers and those prey to premature failure? (2) If the answer is yes, is too much useful life of good lasers consumed during the screening? (3) Does the purge introduce degradation not otherwise present at the intended operating temperature? (4) Does the purge eliminate lasers which would have been adequately reliable at the intended use-temperature?

The motivation for, and the assumptions of, the STP (Stabilization, Truncation, Purge) strategy may be summarized as follows:

1. The long-term reliability of all lasers is controlled by some wear-out mechanism, which may be processing- or structure-related, as opposed to material-related, and which is tolerable for the intended use even if the wear-out mechanism is unknown, or known, but not susceptible to any reduction in its potency.

2. Some lasers, not identifiable after initial characterization, are additionally afflicted with initial drifting of operating parameters and/or one or more other modes of degradation, which are randomly distributed within the population of as-received lasers, and which cause premature failure in the desired regime of operation.

3. The yield of lasers whose degradation is governed exclusively by the wear-out mechanism is inadequately large so that indiscriminate selection is not possible, especially in view of the replacement costs associated with failed devices.

4. Extensive redundancy is, however, not entirely adequate as a reliability assurance strategy because of the requisite increase in complexity and cost which is associated with sparing.

5. The operating stresses, in particular current, or current and optical power, in concert with temperature cause device failure; the shelf life is long.

6. Elevating the operating stresses will accelerate degradation, and elevation to levels just shy of those which annihilate good devices will do the best possible job of producing a robust population of survivors.

7. The duration of the overstress aging should be sufficiently long to stabilize saturable modes of degradation and annealing and to detect premature non-wear-out failure modes (especially those which have incubation periods), but sufficiently short so that enough useful lifetime remains in the survivors to meet the system requirement.

8. The goal in the preceding motivation can be optimally achieved if the modes of premature failure and stabilizable drift have dependences upon the accelerants (optical power, current, temperature, etc.) which are different from those of the wear-out mechanism, and which would permit selective acceleration.

9. Even if the nonfundamental modes of premature failure have the same dependences on the accelerants as does the wear-out mode, acceleration of aging to identify early failures is still essential, especially if they are of the sudden type.

The goals of the STP strategy are then:

1. To identify, where possible, repairable design, growth, and processing flaws that cause, or might cause, failure during the purge, thereby leading ultimately to an increase in the number of survivors.

2. To stabilize initially existing drifts in operating parameters (saturable degradation and annealing mechanisms) that give misleading indications about long-term reliability.

3. To identify and eliminate devices whose lifetimes are dominated by premature failure mechanisms, whether denominated infant, early wear-out, or otherwise.

4. To produce a class of survivors:

(a) whose degradation is governed exclusively by a tolerably slow acting long-term mechanism;

(b) whose degradation is thermally accelerable and whose thermal activation energy determination is unobscured by the presence of nonfundamental modes of degradation or drift;

(c) which possesses no mode of degradation which was introduced solely by the purge and which would not have been present otherwise; and

(d) for which degradation rates can be unambiguously established so that reasonable predictions about the time-to-failure in the operating environment can be made for every single member of the long-lived class, unobscured by the presence of nonfundamental modes of degradation or drift.

VII. ACKNOWLEDGMENTS

The authors wish to thank the following of their colleagues for helpful discussions: W. H. Becker, A. T. English, J. C. Irvin, and C. M. Melliar-Smith. S. Amster, A. B. Godfrey, J. H. Hooper, and especially M. Tortorella provided critical readings of the manuscript, which contributed significantly to its improvement.

REFERENCES

1. I. Bazovsky, *Reliability Theory and Practice*, Englewood Cliffs, N.J.: Prentice-Hall, 1961, Chap. 5, pp. 32–5.
2. F. H. Reynolds, "Thermally Accelerated Aging of Semiconductor Components," IEEE Proc., *62*, No. 2 (February 1974), pp. 212–22.
3. R. E. Barlow and F. Proschan, *Statistical Theory of Reliability and Life Testing: Probability Models*, New York: Holt, Rinehart, and Winston, 1975, Chap. 3, pp. 52–80.
4. E. J. Henley and H. Kumamoto, *Reliability Engineering and Risk Assessment*, Englewood Cliffs, N.J.: Prentice-Hall, 1981, Sect. 4.2.1, pp. 161–9.
5. R. R. Carhart, "A Survey of the Current Status of the Electronic Reliability Problem," Rand Corp. Res. Memo. RM-1131 (August 1953), pp. 1–134.
6. D. R. Earles and M. F. Eddins, "Reliability Physics (The Physics of Failure)," 9th Nat. Symp. Reliab. Qual. Cont. (January 1963), pp. 43–57.
7. M. L. Shooman, *Probabilistic Reliability: An Engineering Approach*, New York: McGraw-Hill, 1968, Chap. 4, pp. 159–201.
8. C. G. Peattie et al., "Elements of Semiconductor-Device Reliability," IEEE Proc., *62*, No. 2 (February 1974), pp. 149–68.
9. J. Wood, *Reliability and Degradation: Semiconductor Devices and Circuits*, M. J. Howes and D. V. Morgan, eds. New York: Wiley, 1981, Chap. 4, pp. 191–236.
10. W. Nelson, *Applied Life Data Analysis*, New York: Wiley, 1982, Chap. 2, pp. 15–102.
11. J. H. K. Kao, "A Summary of Some New Techniques on Failure Analysis," 6th Nat. Symp. Reliab. Qual. Cont. (January 1960), pp. 190–201.
12. H. P. Goode and J. H. K. Kao, "Sampling Procedures and Tables for Life and Reliability Testing Based on the Weibull Distribution (Hazard Rate Criterion)," 8th Nat. Symp. Reliab. Qual. Cont. (January 1962), pp. 37–51.
13. A. L. Floyd, "Reliability Testing," 7th Nat. Symp. Reliab. Qual. Cont. (January 1961), pp. 134–43.
14. D. Mannheim and R. Russell, "Accelerated Tests and Predicted Capacitor Life," 7th Nat. Symp. Reliab. Qual. Cont. (January 1961), pp. 253–61.
15. F. M. Gryna, "Basic Statistical Concepts of Reliability," 7th Nat. Symp. Reliab. Qual. Cont. (January 1961), pp. 41–50.
16. L. A. Weaver and M. P. Smith, "The Life Distribution and Reliability of Electromechanical Parts of an Inertial Guidance System," 8th Nat. Symp. Reliab. Qual. Cont. (January 1962), pp. 148–55.
17. D. S. Peck, "New Concerns About Integrated Circuit Reliability," IEEE Trans. Electron Dev., *ED-26*, No. 1 (January 1979), pp. 38–43.
18. D. S. Peck and C. H. Zierdt, "The Reliability of Semiconductor Devices in the Bell System," IEEE Proc., *62*, No. 2 (February 1974), pp. 185–211.
19. A. T. English and C. M. Melliar-Smith, "Reliability and Failure Mechanisms of

Electronic Materials," Ann. Rev. Mater. Sci., *8* (1978), pp. 459–95.
20. W. B. Joyce, R. W. Dixon, and R. L. Hartman, "Statistical Characterization of the Lifetimes of Continuously Operated (Al,Ga)As Double-Heterostructure Lasers," Appl. Phys. Lett., *28*, No. 11 (June 1976), pp. 684–6.
21. L. R. Goldthwaite, "Failure Rate Study for the Lognormal Lifetime Model," 7th Nat. Symp. Reliab. Qual. Cont. (January 1961), pp. 208–13.
22. G. A. Dodson and B. T. Howard, "High Stress Aging to Failure of Semiconductor Devices," 7th Nat. Symp. Reliab. Qual. Cont. (January 1961), pp. 262–72.
23. D. S. Peck, "The Analysis of Data From Accelerated Stress Tests," 9th Ann. Proc. Reliab. Phys. Symp. (March 1971), pp. 69–78.
24. J. C. Irvin and A. Loya, "Failure Mechanisms and Reliability of Low-Noise GaAs FETs," B.S.T.J., *57*, No. 8 (October 1978), pp. 2823–46.
25. A. S. Jordan, J. C. Irvin, and W. O. Schlosser, "A Large Scale Reliability Study of Burnout Failure in GaAs Power FETs," 18th Ann. Proc. Reliab. Phys. Symp. (April 1980), pp. 123–33.
26. A. S. Jordan et al., "The Reliability of 302A Numerics," B.S.T.J., *57*, No. 8 (October 1978), pp. 2983–99.
27. C. L. Zipfel et al., "Reliability of DH $Ga_{1-x}Al_xAs$ LEDs for Lightwave Communications," 19th Ann. Proc. Reliab. Phys. Symp. (April 1981), pp. 124–9.
28. R. L. Hartman, N. E. Schumaker, and R. W. Dixon, "Continuously Operated (Al,Ga)As Double-Heterostructure Lasers With 70°C Lifetimes as Long as Two Years," Appl. Phys. Lett., *31*, No. 11 (December 1977), pp. 756–9.
29. A. Thompson, "The Reliability of a Practical $Ga_{1-x}Al_xAs$ Laser Device," IEEE J. Quantum Electron., *QE-15*, No. 1 (January 1979) pp. 11–13.
30. M. Ettenberg, "A Statistical Study of the Reliability of Oxide-Defined Stripe CW Lasers of (Al,Ga)As," J. Appl. Phys., *50*, No. 3 (March 1979), pp. 1195–202.
31. A. R. Goodwin et al., "Enhanced Degradation Rates in Temperature-Sensitive $Ga_{1-x}Al_xAs$ Lasers," IEEE J. Quantum Electron., *QE-13*, No. 8 (August 1977), pp. 696–9.
32. M. Ettenberg and H. Kressel, "The Reliability of (AlGa)As CW Laser Diodes," IEEE J. Quantum Electron., *QE-16*, No. 2 (February 1980), pp. 186–96.
33. S. Ritchie et al., "The Temperature Dependence of Degradation Mechanisms in Long-Lived (GaAl)As DH Lasers," J. App. Phys., *49*, No. 6 (June 1978), pp. 3127–32.
34. A. Jordan, "A Comprehensive Review of the Lognormal Failure Distribution With Application to LED Reliability," Microelectron. Reliab., *18*, No. 3 (1978), pp. 267–79.
35. J. H. K. Kao, "Statistical Models in Mechanical Reliability," 11th Nat. Symp. Reliab. Qual. Cont. (January 1965), pp. 240–7. (See also N. R. Mann, R. E. Schafer, and N. D. Singpurwalla, *Methods for Statistical Analysis of Reliability and Life Data*, New York: Wiley, 1974, Sect. 4.6.)
36. W. Weibull, "A Statistical Distribution Function of Wide Applicability," J. Appl. Mech., *18*, No. 3 (September 1951), pp. 293–7.
37. F. Jensen and N. E. Peterson, *Burn-in*, New York: Wiley, 1982, Chap. 2, pp. 3–19.
38. R. E. Barlow and F. Proschan, *Mathematical Theory of Reliability*, New York: Wiley, 1965, Chap. 2, pp. 9–45.
39. B. Unger, "Electrostatic Discharge Failure of Semiconductor Devices," 19th Ann. Proc. Reliab. Phys. Symp. (April 1981), pp. 193–9.
40. E. B. Fowlkes, "Some Methods for Studying the Mixture of Two Normal (Lognormal) Distributions," J. Amer. Statist. Ass., *74*, No. 367 (September 1979), pp. 561–75.
41. M. Stitch et al., "Microcircuit Accelerated Testing Using High Temperature Operating Tests," IEEE Trans. Reliab., *R-24*, No. 4 (October 1975), pp. 238–50.
42. Linear IC Reliability, Western Electric Data Sheet Supplement, *LIC 1-1* (September 1981), pp. 1–5.
43. W. Kuo and Y. Kuo, "Facing the Headaches of Early Failures: A State-of-the-Art Review of Burn-in Decisions," IEEE Proc., *71*, No. 11 (November 1983), pp. 1257–66.
44. W. B. Joyce et al., "Methodology of Accelerated Aging," AT&T Tech. J., this issue.
45. H. Ishikawa et al., "Accelerated Aging Test of $Ga_{1-x}Al_xAs$ DH Lasers," J. Appl. Phys., *50*, No. 4 (April 1979), pp. 2518–22.
46. B. C. DeLoach et al., "Degradation of CW GaAs Double-Heterostructure Lasers at 300K," IEEE Proc., *61*, No. 7 (July 1973), pp. 1042–4.
47. P. Petroff and R. L. Hartman, "Defect Structure Introduced During Operation of Heterojunction GaAs Lasers," Appl. Phys. Lett., *23*, No. 8 (October 1973), pp. 469–71.

48. S. Yamakoshi et al., "Growth Mechanism of <100> Dark-Line Defects in High Radiance GaAlAs LEDs," IEDM Tech. Dig. (December 1978), pp. 642–5.
49. H. Imai et al., "Degradations of Optically-Pumped GaAlAs Double Heterostructures at Elevated Temperatures," Jap. J. Appl. Phys., *18,* No. 3 (March 1979), pp. 589–95.
50. K. Endo et al., "Rapid Degradation of InGaAsP/InP Double Heterostructure Lasers Due to <110> Dark Line Formation," Appl. Phys. Lett., *40,* No. 11 (June 1982), pp. 921–3.
51. M. Fukuda, K. Wakita, and G. Iwane, "Dark Defects in InGaAsP/InP Double Heterostructure Lasers Under Accelerated Aging," J. Appl. Phys., *54,* No. 3 (March 1983), pp. 1246–50.
52. B. W. Hakki, P. E. Fraley, and T. F. Eltringham, "1.3-μm Laser Reliability Determination for Submarine Cable Systems," AT&T Tech. J., this issue.
53. F. R. Nash et al., "Implementation of the Proposed Reliability Assurance Strategy for an InGaAsP/InP, Planar Mesa, Buried Heterostructure Laser Operating at 1.3 μm for Use in a Submarine Cable," AT&T Tech. J., this issue.
54. R. L. Hartman and R. W. Dixon, "Reliability of DH GaAs Lasers at Elevated Temperatures," Appl. Phys. Lett., *26,* No. 5 (March 1975), pp. 239–42.
55. K. Mizuishi et al., "Acceleration of the Gradual Degradation in (GaAl)As Double-Heterostructure Lasers as an Exponent of the Value of the Driving Current," J. Appl. Phys., *50,* No. 11 (November 1979), pp. 6668–74.
56. F. R. Nash and R. L. Hartman, "Accelerated Facet Erosion Formation and Degradation of (Al,Ga)As Double-Heterostructure Lasers," IEEE J. Quantum Electron., *QE-16,* No. 10 (October 1980), pp. 1022–33.
57. W. C. Ballamy and L. C. Kimerling, 'Premature Failure in Pt-GaAs IMPATTS—Recombination-Assisted Diffusion as a Failure Mechanism," IEEE Trans. Electron Dev., *ED-25,* No. 6 (June 1978), pp. 746–52.
58. A. T. English and E. Kinsbron, "Electromigration-Induced Failure by Edge Displacement in Fine-Line Aluminum—0.5% Copper Thin Film Conductors," J. Appl. Phys., *54,* No. 1 (January 1983), pp. 268–74.
59. J. C. Irvin, *GaAs FET Principles and Technology,* J. V. DiLorenzo, ed., Dedham, MA: Artech House, 1982, pp. 353–400.
60. A. S. Jordan, "Confidence Limits on the Failure Rate and a Rapid Projection Nomogram for the Lognormal Distribution," Microelectron. Reliab., *24,* No. 1 (1984), pp. 101–24.
61. G. S. Watson and W. T. Wells, "On the Possibility of Improving the Mean Useful Life of Items by Eliminating Those With Short Lives," Technometrics, *3,* No. 2 (May 1961), pp. 281–98.
62. A. J. Wahl et al., "SF Submarine Cable System: Transistors, Diodes, and Components," B.S.T.J., *49,* No. 5 (May–June 1970), pp. 683–98.
63. I. G. Abrahamson, E. T. Lee, and E. B. Slutsky, "Procedures for Selection of Semiconductor Diodes for Use in Undersea Cable Systems," IEEE Trans. Reliab., *R-21,* No. 4 (November 1972), pp. 200–6.
64. L. E. Miller, 'Reliability of Semiconductor Devices for Submarine-Cable Systems," IEEE Proc., *62,* No. 2 (February 1974), pp. 230–44.
65. W. M. Fox et al., "SG Undersea Cable System: Semiconductor Devices and Passive Components," B.S.T.J., *57,* No. 7 (September 1978), pp. 2405–34.
66. E. I. Gordon, F. R. Nash, and R. L. Hartman, "Reliability Assurance for New Technology Semiconductor Devices," IEEE Electron Dev. Lett., *EDL-4,* No. 12 (December 1983), pp. 465–6.
67. R. T. Lynch, "Effect of Screening Tests on the Lifetime Statistics of Injection Lasers," IEEE J. Quantum Electron., *QE-16,* No. 11 (November 1980), pp. 1244–7.
68. R. H. Saul, F. S. Chen, and P. W. Shumate, "Reliability of InGaAs Photodiodes for SL Applications," AT&T Tech. J., this issue.
69. R. H. Saul and F. S. Chen, "Reliability Assurance for Devices with Sudden-Failure Characteristic," IEEE Electron Dev. Lett., *EDL-4,* No. 12 (December 1983), pp. 467–8.
70. S. Yamakoshi et al., "Reliability of High Radiance InGaAsP/InP LEDs Operating in the 1.2–1.3 μm Wavelength," IEEE J. Quantum Electron., *QE-17,* No. 2 (February 1981), pp. 167–73.
71. S. Yamakoshi et al., "Degradation of High Radiance InGaAsP/InP LEDs at 1.2–1.3 μm Wavelength," IEDM Tech. Dig. (December 1979), pp. 122–5.
72. M. Hirao et al., "Fabrication and Characterization of Narrow Stripe InGaAsP/InP Buried Heterostructure Lasers," J. Appl. Phys., *51,* No. 8 (August 1980), pp. 4539–40.

73. H. Kressel, M. Ettenberg, and I. Ladany, "Accelerated Step-Temperature Aging of $Al_xGa_{1-x}As$ Heterojunction Laser Diodes," Appl. Phys. Lett., *32*, No. 5 (March 1978), pp. 305–8.
74. Y. Furukawa et al., "Accelerated Life Test of AlGaAs-GaAs DH Lasers," Jap. J. Appl. Phys., *16*, No. 8 (August 1977), pp. 1495–6.
75. H. Imai et al., "Activation Energy of Degradation in GaAlAs Double Heterostructure Laser Diodes," J. Appl. Phys., *52*, No. 5 (May 1981), pp. 3167–71.
76. A. R. Eckler, "A Statistical Approach to Laser Certification," AT&T Tech. J., this issue.

APPENDIX

Glossary of Some Reliability Terminology

A.1 When failures occur

Wear-out* failure—This failure eventually terminates the life of all† devices, provided that they have not failed previously from non-wear-out causes. In connection with the failure of, for example, brake linings, wear-out can be understood as using up some property until little or nothing of it remains. In semiconductor devices, where the wear-out concept has no comparably simple meaning, all wear-out failures appear to be lognormally distributed. The dispersion or variance of the lognormal wear-out distribution occurs because all devices are initially inequivalent and hence they proceed to demise at different rates. The thermal activation energies for semiconductor wear-out failures lie in the range 0.5 to 1.5 eV.

Midlife* failure—This is an accidental failure of a device during its useful life, prior to wear-out. The accidental aspect refers to devices being at the mercy of inherently unpredictable external events. Even though all of the devices are initially equivalent in all respects and contain no built-in weaknesses, some devices will fail because of random accumulations of operating stresses, e.g., current surges, due to electrostatic discharges, which exceed normal specifications. Although these failures are usually infrequent, they would be expected to be exponentially distributed in time, by analogy with radioactive decay in which all nuclei are viewed as initially equivalent. The corresponding hazard rate is a constant.

Infant* failure—A failure from a variety of possible causes, all of which constitute an initial weakness in the as-made device, which terminates the operation of a device early in its intended use. Thermal activation energies are generally low ($<$0.4 eV). Since the weakness is in the device, screening techniques employing overstresses are useful

---

* Wear-out, midlife, and infant are defined in the context of the bathtub curve, Fig. 1a.

† Window glass, whose purpose is to let in some light and keep out rain, never wears out. For the purposes of a discussion of semiconductor reliability, it is assumed that all biased devices will eventually fail because of diffusion of nonradiative defects to the active region, if for no other reason.

in detecting the fraction of the population potentially susceptible to such early failures. Infant failures in an unscreened population may not be recognized if the sample size is too small.

Freak failure—An early failure group observed during accelerated thermal aging which is manifested as an S-shaped distortion in the early time portion of the lognormal wear-out failure distribution. Only if the accelerated aging is performed at another elevated temperature is it possible to tell if the freak distribution is associated with the presence of a high-activation (wear-out-type) or low-activation (infant-type) mechanism.

Premature failure—A failure which occurs prior to some desired time. It can comprehend the aggregate of the infant, midlife, and the early portion of the wear-out failures.

A.2 Why failures occur

Random or chance failure—The previously defined midlife failure is a random or chance failure. The failure is event dependent. It is an external randomly occurring event, rather than a device deficiency, that causes the failure. The failure (or start of degradation) cannot be anticipated because prior to the event, no conceivably measurable device parameter would have exhibited any precursor degradation. The failure (or start thereof) is sudden. No preinstallation screen, burn-in, accelerated aging, or purge will identify devices potentially susceptible to chance failure, since all devices are equally prey. This use of the word random is rather specific, and the use in other contexts can cause confusion. Thus, wear-out failures are never random failures despite the fact that the initial inequivalence in the devices, which is responsible for the dispersion in the wear-out failure distribution, was promoted by some randomness in growth or processing. The location of devices, in a box, subject only to wear-out failure, may also be said to be random.

Preordained failure—Wear-out failures and some infant failures are examples. A preordained failure is one in which some internal mechanism causes a device to fail deterministically at a rate preordained at the moment of birth. Preordained failures are, therefore, time dependent.

A.3 How failures occur

Gradual failure—A failure that is characterized by a graceful degradation, a gradual change with time in one or more critical parameters. The essence of the failure is the exceeding of some specification; the device is often still operative (e.g., wear-out of a pair of shoes).

Sudden or catastrophic failure—A failure that gives no warning of its imminence. It may actually occur rapidly (e.g., light bulb failure),

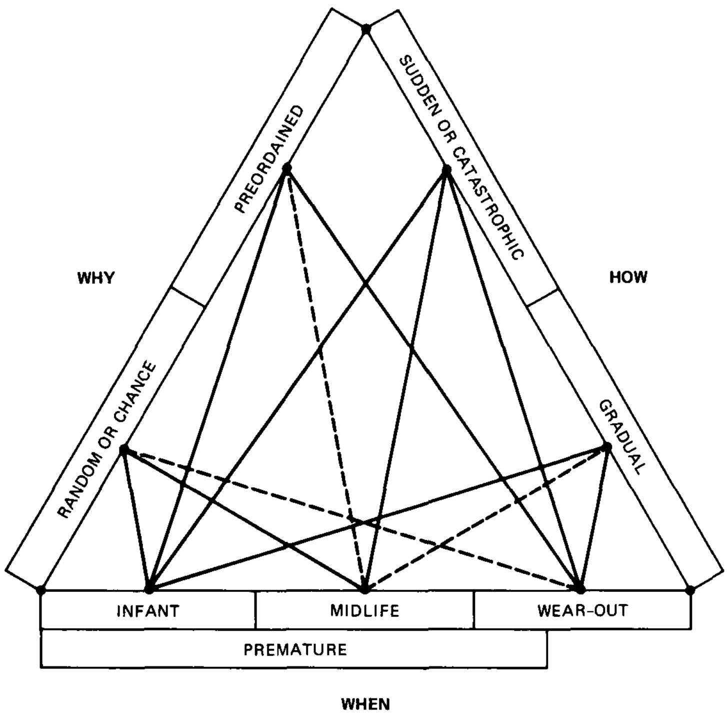

Fig. 11—Depiction of connections among terms defined in the Appendix. Solid lines denote the connections; e.g., wear-out failures can be sudden or gradual. Dashed lines denote an absence of connection; e.g., wear-out failures are not randomly provoked. The classification "premature" is a system-lifetime term; all other terms are referenced to device lifetimes.

or in the period in between times of observation. A preordained failure may appear sudden because it is impossible or inconvenient to monitor the parameters that would have enabled failure to be anticipated. Even had monitoring occurred, the changes might have been too small to detect if there was an incubation period prior to reaching a failure threshold, e.g., thermal runaway. The other aspect to the catastrophic failure is the complete inoperativeness of the device under any circumstances, subsequent to the failure, e.g., the opening of an electrical contact due to corrosion. Both random and preordained failures may be sudden.

A summary of the connections among these terms appears in Fig. 11.

AUTHORS

Richard W. Dixon, A.B. (Engineering and Applied Physics), 1958, Harvard College; M.S. and Ph.D. (Engineering and Applied Physics), Harvard Univer-

sity in 1960 and 1964, respectively; AT&T Bell Laboratories, 1965—. Mr. Dixon is currently the Director of the Lightwave Devices Laboratory of AT&T Bell Laboratories.

Eugene I. Gordon, B.S. (Physics), 1952, City College of New York; Ph.D. (Physics), 1957, The Massachusetts Institute of Technology; AT&T Bell Laboratories, 1957–83; Lytel, Inc., 1983—. Mr. Gordon has worked in the field of gas-discharge physics; microwave traveling-wave tubes; gas and GaAs injection lasers; acousto-optic modulation and deflection devices; image and display devices for the *Picturephone®* visual telephone service; and photo and electron beam lithography and packaging of silicon integrated circuits. Most recently, Mr. Gordon was Director of the Lightwave Devices Laboratory. He is the author of 43 published articles and holds 23 patents. Mr. Gordon received the IEEE 1975 Vladimir K. Zworykin Award and was elected to the National Academy of Engineering in 1979. He is past chairman of the Working Group on Imaging and Display Devices for the Office of the Director of Defense Research and Engineering and chairman of the Publications Committee for the Administrative Committee of the IEEE Electron Devices Society. Fellow, IEEE. Member, Advisory Group on Electron Devices, IEEE Publications Board, American Physical Society, Phi Beta Kappa, Sigma Xi.

Robert L. Hartman, B.A. (Mathematics), 1958, Columbia College; B.S. (Engineering Physics), M.S. and Ph.D. (Physics), Columbia University, in 1959, 1961, and 1968, respectively; AT&T Bell Laboratories, 1968—. At AT&T Bell Laboratories, Mr. Hartman is presently engaged in studying the physics of III-V compound semiconductor devices, particularly InGaAsP/InP and (Al,Ga)As-GaAs double-heterostructure injection lasers. He has contributed extensively to the understanding of the degradation mechanisms in these devices and authored or coauthored several key papers in this field, especially those involving the discovery and analysis of dark-line defects; the discovery of strain effects upon laser reliability; the attainment of long operating lifetimes in LEDs and injection lasers; the first measurement of the thermal acceleration of injection laser operating lifetime; and effects of surfaces on LED and laser reliability. In 1980 Mr. Hartman was appointed Supervisor of the Semiconductor Lightwave Lasers Group. Member, Sigma Xi, American Physical Society.

William B. Joyce, B.E.P. (Engineering Physics), 1955, Cornell University; Ph.D. (Theoretical Physics), 1966, Ohio State University; AT&T Bell Laboratories, 1966—. Mr. Joyce publishes primarily in the areas of architectural acoustics, classical mechanics and optics, history of science, semiconductor optical devices, research and development organization, and reliability. Member, American Physical Society.

Franklin R. Nash, B.S. (Physics), 1955, Polytechnic Institute of New York; Ph.D. (Physics), 1962, Columbia University; AT&T Bell Laboratories, 1963—. At AT&T Bell Laboratories, Mr. Nash has been involved in research on gas lasers (modulation and mode locking), detection of subnanosecond pulses using photodiodes, nonlinear optical materials, parametric oscillators, neoclassical theory of radiation, and semiconductor lasers (mode guidance, material defects, electro-optic anomalies, and degradation mechanisms). Currently, he is engaged in the reliability assessment of long-wavelength semiconductor lasers. Member, American Physical Society, IEEE.