IMPROVING THE PERFORMANCE OF AN INTEGRATED CIRCUIT MANUFACTURING LINE

Jeffry M. Neve, F. Duane Ray, and Joseph P. Sitarik

Jeffry M. Neve *is with AT&T Components and Electronic Systems in Kansas City.* ***F. Duane Ray*** *is with the AT&T Engineering Research Center at Allentown, Pennsylvania.* ***Joseph P. Sitarik*** *is with the Engineering Research Center currently on assignment with the Sematech industry-government consortium. Mr. Neve is a department chief responsible for ceramic multilayer board engineering and manufacturing. Since joining AT&T in 1971, he has held many positions in research, engineering, and manufacturing. He received a B.S. in metallurgical engineering from Michigan State University, an M.S. in ceramics engineering from Massachusetts Institute of Technology, and an M.B.A. from Rockhurst College. Mr. Ray is a member of technical staff in the Manufacturing Systems Engineering* (continued on page 48)

In 1986, two years after startup, a state-of-the-art integrated circuit (IC) clean room facility at AT&T was not performing as expected. It was experiencing a high inventory, a correspondingly long manufacturing interval, and an undesirably low production rate. If more material was started into the line, the production rate changed very little and the inventory and interval increased. In midyear, a management team was formed to address these problems. By the end of 1986, as a consequence of the team's efforts, both the manufacturing interval and the inventory had decreased by 40 percent. These improvements were obtained while the production rate increased by 10 percent and without expenditure of significant capital or additional human resources. Here, a description of the techniques used to achieve these results is presented.

Introduction

The clean room at AT&T Technologies—Kansas City was completed in 1984 for the production of metal-oxide-semiconductor (MOS) ICs with feature sizes down to 1.25 μm. The manufacturing line was producing basically one code but with some variations. After ramp-up during 1985, the manufacturing line exhibited intervals which were five times the "butt-to-butt" time (defined as the sum of all loading, processing, and inspection times) and correspondingly high inventories. Moreover, an increase in production rate was required to meet the projected demand. Although reliable information was scant, our perception was that the very best of similar lines worldwide operated at 2.2 to 2.5 times the butt-to-butt time. In addition, yields were low and rework was averaging 12 percent of production activity. Since it is generally accepted that reducing the manufacturing interval improves yields by providing more rapid feedback on the process, one goal was to reduce the manufacturing interval to 2.5 times the butt-to-butt time. However, it could not be determined whether the anticipated improvement in yield

would lead to a sufficient increase in the production rate of good chips.

Representatives of several organizations were formed into a task force composed of three teams: Capacity/Interval Management, Defect Analysis, and Electrical Yield. This article describes the work of the first of these, itself made up of representatives from three groups: Process Engineering, Production, and Industrial Engineering. Process engineering had primary responsibility for equipment and process-related issues. Production controlled the operating philosophy, including product priority. Industrial Engineering included engineers responsible for simulation modeling as well as resources from the R&D community, which provided tools and methods for use in examining problems.

Background

In a modern IC manufacturing facility, a wafer of silicon approximately 150 mm in diameter and 1 mm thick is passed through a sequence of some 150 to 200 steps, which create on the surface hundreds of replicas of a device or "chip." Each chip consists of a network of thousands of active and passive circuit elements. Many steps must provide physical accuracies of a fraction of a micrometer, and many other steps must result in carefully controlled chemical additions to specifically defined areas of the silicon. These stringent requirements necessitate the use of a clean room containing a variety of very complex and expensive equipment.

To utilize the equipment to the fullest, the rooms are operated 7 days a week, 24 hours a day. The wafers are normally carried through the manufacturing process in lots of 50, but at some steps it is efficient to batch several lots together before processing. Also, because of the expense of the material and the complex processing, wafers or lots that are processed incorrectly are reworked whenever possible instead of being junked. Rework is often performed with wafer quantities less than a full lot. These smaller groups of wafers may be held until a full (reconstituted) lot is formed before being put back into the normal process stream. At the end of the manufacturing process, the chips are sawed apart from one another and packaged separately.

Two basic process sequences are used to create the circuit elements and their interconnections. The first, a photolithographic sequence outlined in Figure 1, is used to define the areas in surface layers where material for the elements or interconnections will be incorporated. Each wafer must cycle through this photolithographic sequence 8 to 12 times during its manufacture. A device is defined by the specific set of photolithographic masks that is used. Proper alignment of the photolithographic features of a given step with those defined in previous steps is critical. The second basic process sequence, also shown in Figure 1, is the sequence that uses ion implantation to add the minute quantities of material required to give the desired electrical properties. This sequence uses either patterned photoresist or an etched surface layer to define the implanted areas. The amount and composition of implanted material must be carefully controlled.

Each lot entering the room is identified by a number written in bar code form. Each time a step is completed, information about lot status is entered into the information system by operators using either keyboard input or a bar code wand.

A consistent process is required to reduce rework and junk. Equipment must be monitored and maintained, records must be kept, and operators must be aware of proper procedures to achieve both high quality and smooth product flow through the room. Also, management needs summary information about the availability and status of equipment, the product in the line, and yields.

Analysis

The initial focus was to compare the existing input rate to the capacity of the line. The Deterministic Capacity Model (DCM) developed by D. Y. Burman and S. Sathaye of the Manufacturing Systems Engineering Department was used, since it contains a simple but powerful deterministic model of a process line. Even a rough estimate of

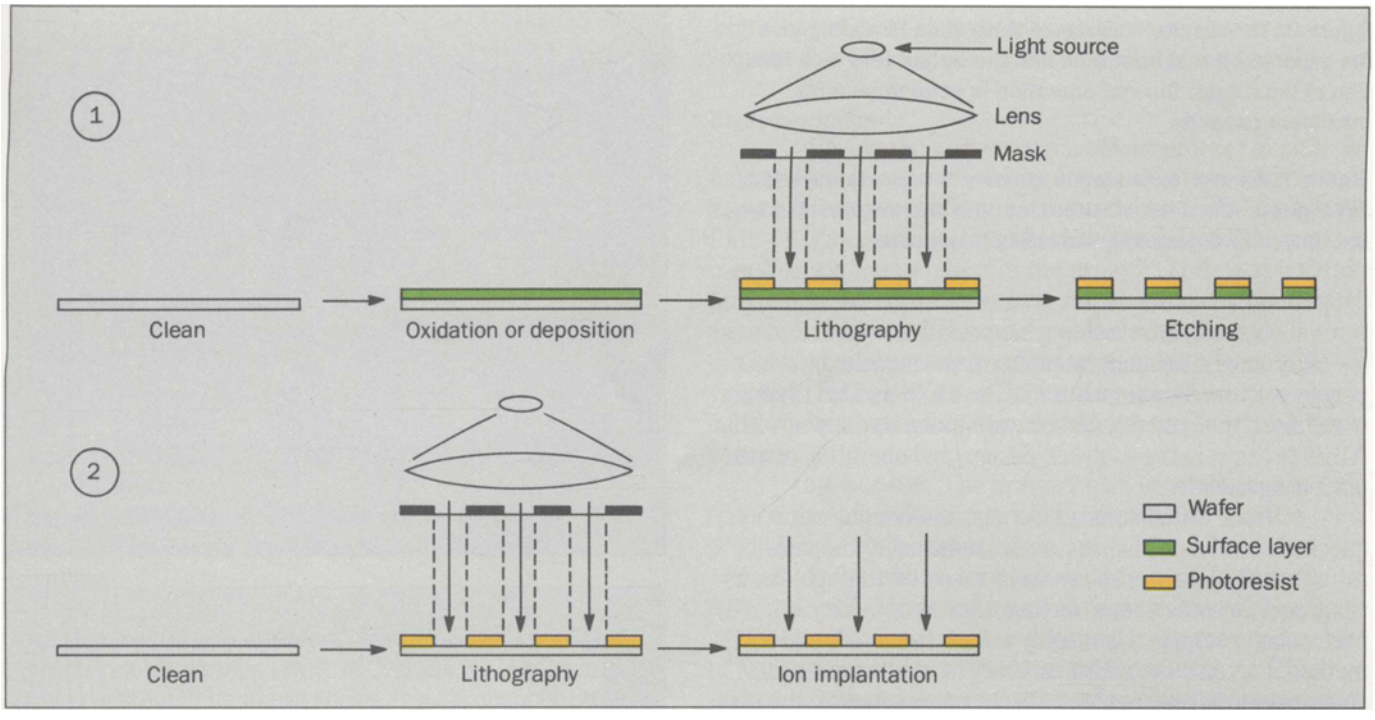


Figure 1. The two major process sequences in integrated circuit manufacturing. The sequences are combined to form the process recipe, which may contain 100 to 200 individual steps.

capacity requires some data, or reasonable estimates, of the routing, process times, quantity of equipment, down times and repair times, rework, yields, batching and setups. These data not only allow a quantitative capacity analysis, but also have other benefits. First, a more complete analysis using a discrete event simulator could eventually utilize all of the data. Second, the collected data create a summary of information regarding the important product flow parameters.

The results of the deterministic modeling showed that the input to the line was very close to, or perhaps slightly above, the capacity of the line. In addition, observations of the line and discussions with the engineering and operating personnel suggested that the product flow was highly variable. This combination of high variability and high utilization is where deterministic tools are least accurate. Consequently, to better estimate the relationship between input rate and interval, a discrete event simulation model was constructed.[1-3]

The simulation modeling tool used was the Generic Simulation Tool (GST) developed by F. Javier Gurrola-Gal of the Manufacturing Systems Engineering Department. The GST serves as a very friendly and efficient interface to a popular simulation language, SIMAN (trademark of Systems Modeling Corporation). It uses a subset of the full SIMAN language and creates generic modeling structures automatically from files of input data. In addition, it checks the input

Figure 2. The results of discrete simulation modeling give the expected throughput time and the output rate as a function of input rate. Current operation is seen to be near maximum capacity.

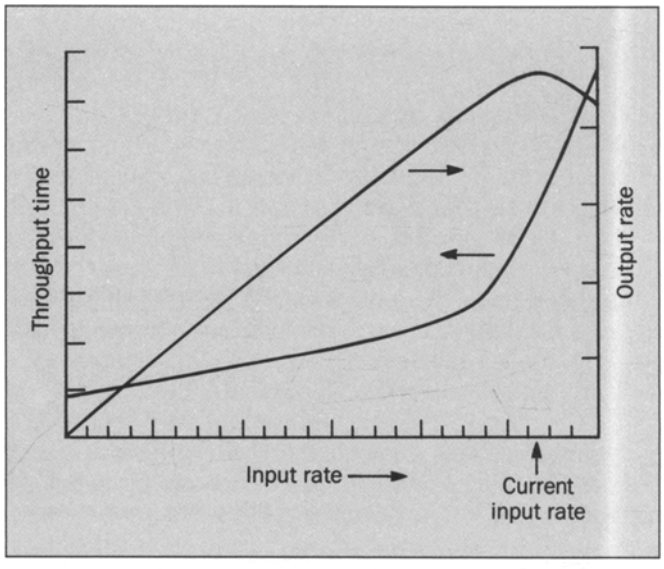
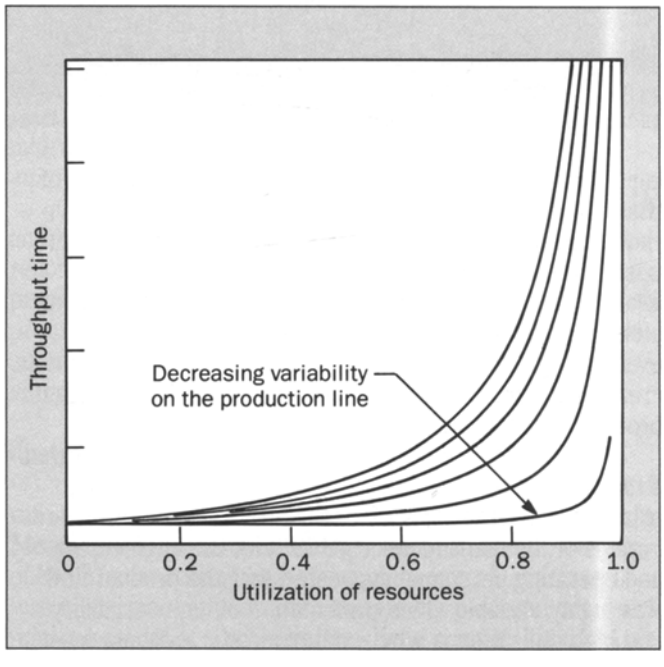


Figure 3. Results from simple queuing theory are depicted. For a given utilization of resources, the decrease in throughput time with decreasing variability is apparent.



data and presents the output in a usable form. These features allow rapid, easy building of simulation models and greatly enhance the understanding of the modeling work by people not versed in simulation. (The DCM and GST belong to a collection of roughly 60 software tools, developed within AT&T to assist in the analysis, design, and operation of manufacturing facilities.)

After initial simulation runs, the results were checked to validate that the model reasonably tracked actual results. Discrepancies were resolved through discussions and, in some cases, further studies of facility performance. Again, this highly interactive, back-and-forth methodology had the additional benefit of enhancing the communication between and among the engineers, the production people and the modelers. Potential problems were already being identified and attacked even before any simulation results were available.

The results of the simulation modeling are shown in Figure 2. The input rate in use at the time of the study is marked on the figure. The results confirmed the estimate obtained from the DCM that the line was being run at its maximum capacity. Figure 2 also suggests that the only means for reducing the manufacturing interval is to reduce the wafer input rate. According to the figure, however, this could reduce the wafer output rate. To achieve an increased production rate of good chips would thus require a significant improvement in yields—an improvement that could not be guaranteed in the time available.

A consideration of simple queuing analysis pointed to the resolution of the apparent dilemma. The concept, shown in Figure 3, depicts the relationship between

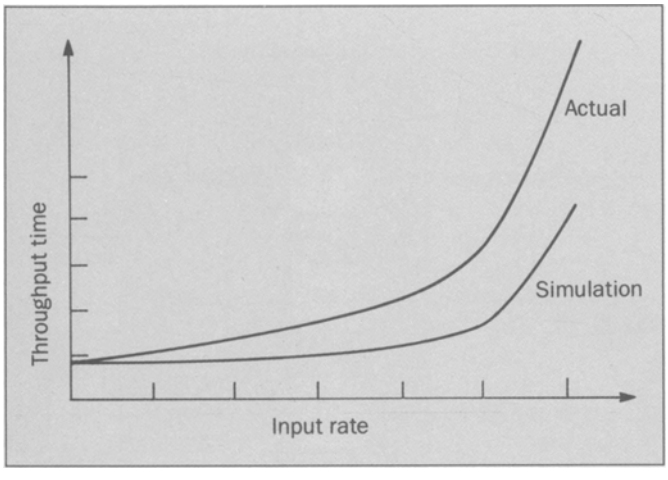


Figure 4. Simulation results show the benefit to be gained by decreasing the variability of equipment downtime.

throughput time (a manufacturing interval) and utilization of resources for various levels of "variability" for a simple queuing system. The figure shows that throughput time can be decreased and the production rate increased if the level of variability is reduced. For an IC manufacturing line (which is a very complex queuing system), primary causes of variability include variances in the mean time to failure and mean time to repair for equipment, variances in process times (especially in photolithography), and the batching of several lots of wafers before they are passed on to the next process step.

The model whose results are shown in Figure 2, and which agreed well with the line's actual performance, was based on parameters reflecting a high degree of variability in the line's operations. This observation, combined with the insights gained from queuing theory, suggested that it should be possible to reduce the manufacturing interval while increasing the production rate by reducing the level of variability in the line. Indeed, a simulation carried out with more typical parameters (see Figure 4) showed that considerable improvement could be expected.

Implementation

Having the proper management process in place is fundamental to a successful implementation, and this aspect of a project is frequently overlooked. The Capacity/Interval Management Team held regular biweekly meetings to track the progress of the project. The goals and the progress to date were regularly distributed to upper management. Daily metric reports were developed and distributed to everyone involved to give the project high visibility. These reports were used by supervisory personnel associated with the clean room to determine priorities of work.

Demand-Pull. The primary plan of attack was to reduce the unnecessary variability introduced into the product flow by operational policies.

In the past several years, U.S. manufacturers have reported significant improvements in the performance of production lines due in large part to the application of just-in-time (JIT) or "demand-pull" practices.[4-6] These techniques were viewed as an excellent way of improving the interarea communication and smoothing the product flow in the line. The key to demand-pull philosophy is shown in Figure 5. In complicated processes where certain facilities are common to many different process steps, a priority scheme must be developed for selecting the next lot from the queue. In "push" systems, priority is usually determined by the amount of product waiting to be processed. By choosing the product with the largest queue, the efficiency of the process step is maximized over the short run, because the setup time is minimized. Using "demand-pull," more emphasis is placed on the downstream need and the priority is determined by the inventory at downstream steps. In this manner, the efficiency of the entire line is improved although a particular step may suffer since more setup time may be needed.

Inventory Goals. A clean room employs many first-line supervisors on each shift. These supervisors are

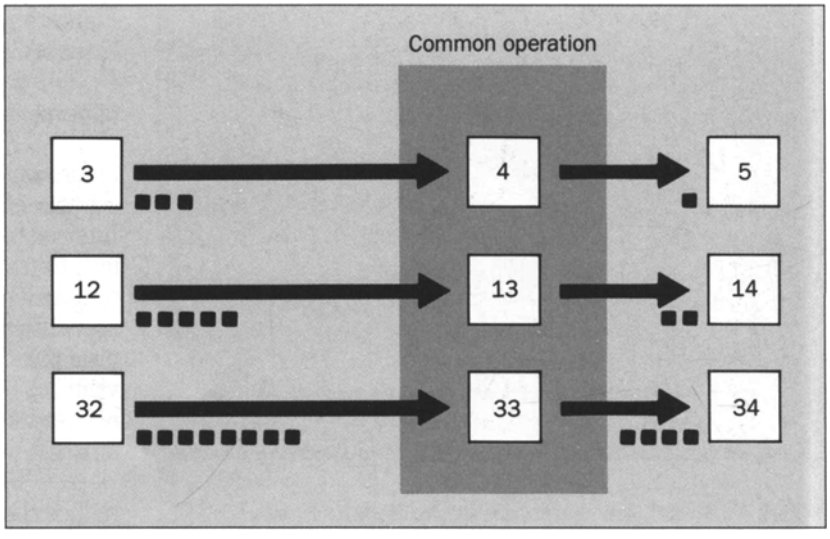


Figure 5. The demand-pull philosophy is illustrated by the dilemma of choosing work at an operation common to several process steps. Numbers in large blocks indicate the process step. Small blocks represent lots in the queue. In a "push" system, work from step 33, with the highest inventory, would be chosen. In a "pull" system, work from step 4 would be chosen, since this would assure downstream steps are kept busy.

directly responsible for the allocation of direct labor and typically make numerous decisions on which product or process should be run. To ensure a smoothly running line, these decisions must be consistent, timely, and well-informed. To support such decisions, inventory goals were established by "zone," a group of successive process steps. Twenty-six zones were identified by using the simulation model as initial input, with modifications based on discussions with engineering and production. Over time, the inventory goals for each zone were reduced as improvement became evident. To assist shop supervision, a simplified report was generated from the shop flow data base four times daily. The report showed actual inventory by zone and step, lot moves by zone and step, inventory goal by zone, and difference between actual and goal. Problem areas were easily identified by excess inventory or absence of lot moves. The report was used initially at daily meetings to establish priorities.

Where possible, demand-pull decisions were expedited by the use of shelves containing completed product at a process step. In a facility (such as an inspection station) which services several other facilities, completed product would be placed on the shelves segregated according to the area they go to next. By observing the lots or lack of lots on the shelf, an operator could identify the product needed downstream. In IC processing, there are many operations with maximum delay times between steps. At these operations, shelves could not be used, but a formalized ordering procedure was instituted across all shifts.

Information Access. One of the problems of scheduling priorities in a complex process is projecting inventory levels in the short (daily) run. Observations indicated that failure to anticipate the arrival of a product led to an unnecessary setup. To address this problem, an information system was developed by one of the authors (F. D. Ray) to project inventory levels at each process step by shift for the following two days. The software program, called the Dynamic Management Information System (DMIS), had embedded in it a deterministic model of the manufacturing line, utilized the shop flow inventory data base, and was capable of "what-if" analysis of estimated lot moves at operations, including the ability to adjust the capacity of facilities with known problems. The program was accessible through the computer terminals in the clean

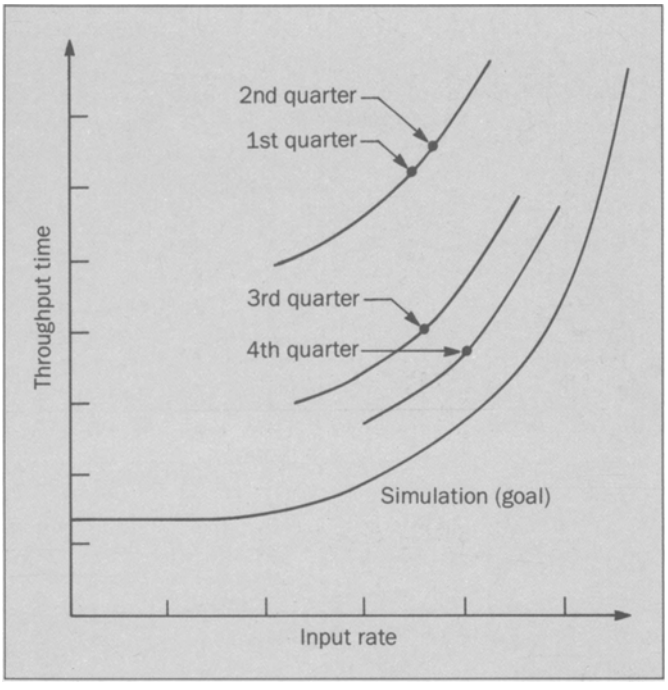


Figure 6. The reduction in throughput time with a corresponding increase in capacity demonstrates the improvement obtained by concerted efforts to reduce variability in the process.

room and could be used for real-time analysis.

Maintenance. Efforts were also focused on improving the maintenance program, especially preventive maintenance. It was anticipated that this would have the dual impact of increasing available machine uptime and of reducing the variability of uptime, both of which would increase effective capacity.

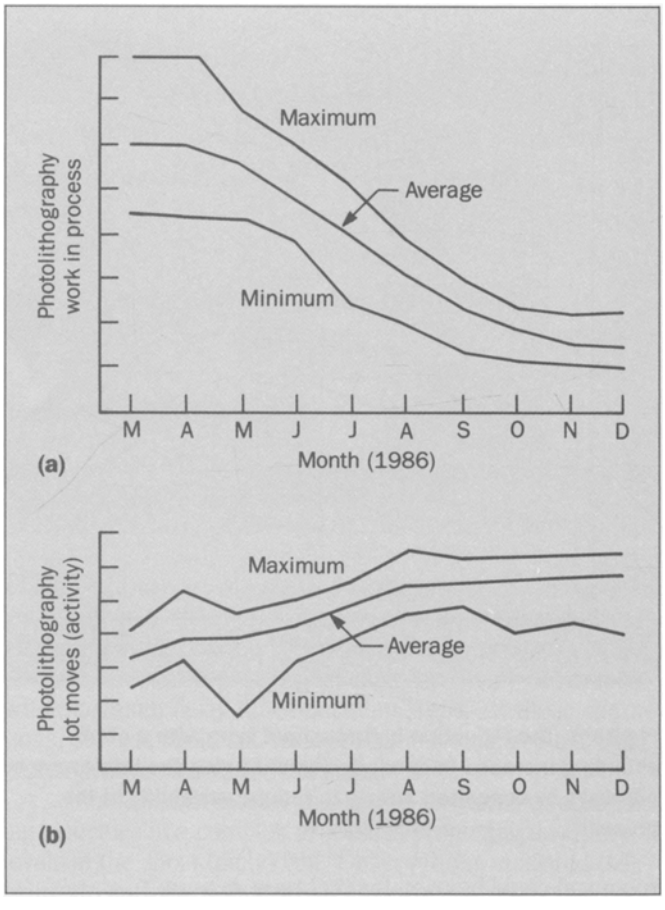
The photolithography operation, which uses very sophisticated equipment and was known to be the major bottleneck in the line, was selected as the target project. Since it was suspected that the equipment was not running at maximum throughput, a preventive maintenance program was established. In addition, repair classes were conducted for the dedicated technicians in the photolithography area. Several similar programs in other problem areas, including ion implantation and metals deposition, were initiated. Even the services, such as air, water, and fume hood venting, received improved preventive maintenance to reduce the disruptions caused by these services.

Information on machine status to direct maintenance efforts and to acquire better data for modeling efforts was needed. One method of collecting information on machine status consisted of using hand-held bar code readers with internal memory. Bar codes signifying machine number, machine running product, idle, waiting for maintenance, under repair, etc. were made available at each machine so operators could read status easily. On a daily basis, data were extracted from the readers into a personal computer. A program was written and used to identify problem machines and to establish a baseline of performance. Since variability in the line can be affected by allocation as well as effectiveness of repair effort, the highest maintenance priority was assigned to those areas identified as bottlenecks.

Education. One of the objectives of the team was to educate all personnel from the highest engineering management to the production workers concerning demand-pull philosophy and methods to reduce variability. This was accomplished in several ways. Biweekly status reports were developed and distributed. Numerous technical courses were offered in areas such as maintenance and repair techniques. Finally, the concept of demand-pull was introduced to all operators with the aid of a videotape as part of an effort to stimulate suggestions and increase operator involvement. The videotape contained a mock IC manufacturing line and visually captured the results of reducing work in process and machine variability. Improvements were evident in the form of reduced manufacturing space, reduced rework, and increased output.

Another method to reduce variability was to improve communications between shifts. A maintenance pass-down log (a record of maintenance performed on the

Figure 7. Inventory (a) and activity (b) in the photolithography stepper operation shows the dramatic improvement achieved by reducing variability, most notably with preventive maintenance and technician training.



previous shift) was kept on the shop information system for each piece of equipment.

Results

The significant improvement obtained in the overall performance of the line is demonstrated in Figure 6, which shows manufacturing interval as a function of input. The first two quarters had high variability and therefore high inventory. In the third quarter, as the improvement programs began to take effect, the variability dropped, allowing the input rate to be maintained while a reduced interval was realized. In the fourth quarter, as improvements continued, the input rate increased and the interval was further reduced. Also shown on Figure 6 is the simulation goal discussed in conjunction with Figure 4. This indicates that further improvement is possible.

One area of the line where significant improvements were realized was photolithography. Not only did the inventory in the photolithography area drop considerably (Figure 7a), but the variability in the inventory level decreased. As these improvements were occurring, the activity level in the photolithography area was actually increasing (Figure 7b). The improved performance is ascribed to preventive maintenance and technician training. In fact, during the course of the project, the major bottleneck in the line was intentionally shifted from photolithography to final test.

The manufacturing interval is shown in Figure 8. The upper curve is the total interval, including final test, and the lower curve is the total interval, minus the final test interval. In October 1986, there was an increase in the final test time because of increased testing to improve assembly yields, but the interval for the rest of the line continued to decrease. In November, the final test time was slightly decreased and the total interval began to drop again. Final test became the bottleneck. Since activity lost in a bottleneck operation cannot be recovered,[7] a safety stock inventory was purposely kept at final wafer test to prevent a stock-out condition due to upstream fluctuations. The queue selection method at this operation was changed from first-in, first-out (FIFO) to a random procedure to ensure that yield information was obtained on a timely basis on the most recent lots.

As the interval was decreasing in final test, the activity in final test was increasing (see Figure 9.) Not only was the average level of testing increasing, but the variability was significantly decreasing, demonstrating a higher,

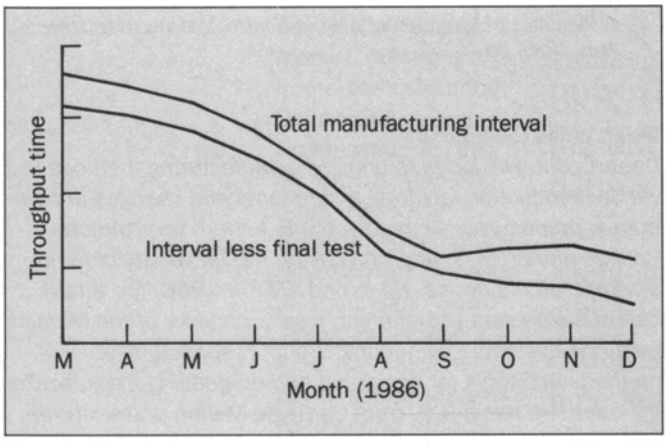


Figure 8. The manufacturing interval decreased over the study period, and a greater portion of the interval was shifted (purposely) to the final test operation.

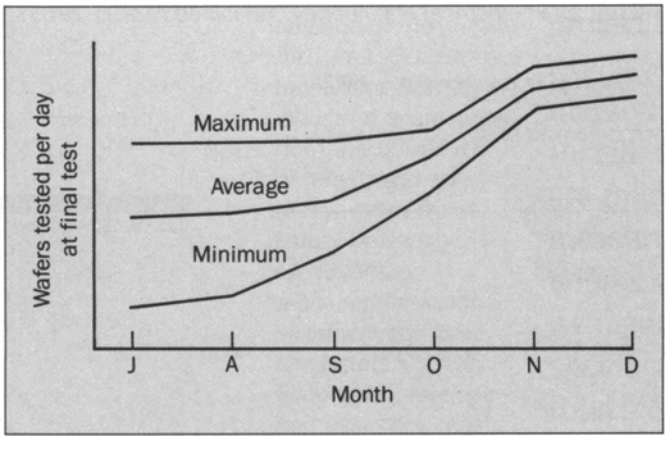


Figure 9. Wafers tested per day at final test increased, and, at the same time, the variation in the number tested decreased.

smoother flow of product from the line. In a multicode environment, lot prioritization at each process step would also be required. This would necessitate the use of prioritization schemes and decision strategies for batching and routing.[8]

Conclusion

The performance of an IC manufacturing line was improved substantially by the application of an interdisciplinary team approach. A variety of information systems and tools, varying in degree of complexity, were applied to several areas of the manufacturing process. Simulation modeling was used to determine reasonable goals for WIP and manufacturing interval, as well as to examine proposed process and operational changes. A demand-pull philosophy was implemented in production to prioritize lot moves. Preventive maintenance programs and technician training were implemented to reduce equipment downtime and variability. Supervisor meetings and information tools based on the shop information system provided timely information for decision support. No single information system or tool, but rather the application of many methods, was responsible for the substantial improvement in performance. The methods used to improve performance have general application to other manufacturing processes.

In parallel with the Capacity/Interval Management Team, responsible for the activities described in the foregoing sections, other engineering and R&D personnel addressed specific process improvements. Over the duration of the project, the rework activity decreased from 12 percent to 5 percent and the chip test yield increased by 27 percent.

It is evident that the reduction in rework rates contributed in an important way to reducing variability in the line.

Acknowledgments

It should be clear from the preceding article that the authors are more the documenters of a successful team effort than the sole agents of change. Many people worked very hard to make these improvements. We must acknowledge the support of C. M. Melliar-Smith, R. B. Koch, D. C. Krupka, and T. K. Merrell. In addition, the contributions and involvement of D. L. Engeman, L. L. Jones,

D. R. Mc Lachlan, C. W. Pierson, S. J. Salva, E. J. Thiele, K. L. Wagner, and L. L. Wesner were vital to the success of this project. Also, discussions with P. Crossland of Rockhurst College were helpful in establishing direction for parts of the project.

References

1. D. Y. Burman et al., "Performance Analysis Techniques for IC Manufacturing Lines," *AT&T Technical Journal*, Vol. 65, No. 4, July-August 1986, pp. 46-57.
2. J. J. Golovin, "A Total Framework for Semiconductor Planning and Scheduling," *Solid State Technology*, May 1986, pp. 167-170.
3. E. Lohraspbour and S. Sathage, "Simulation of VLSI Manufacturing Areas," *VLSI Design*, Vol. V, 1984, pp. 84-92.
4. L. Corey, "Just-In-Time Approach to IC Fabrication," *Solid State Technology*, May 1986, pp. 177-179.
5. "Quality Quest, The Resurrection of High Tech Manufacturing," *Electronic Business*, January 15, 1987, pp. 52-64.
6. R. R. Mayees, "A Critical Look at Kanban, Japan's Just-In-Time Inventory System," *Management Review*, December 1984, pp. 48-51.
7. E. M. Goldratt, "Cost Accounting, the Number One Enemy of Productivity," *Proceedings*, 1983 American Production and Inventory Control Society Conference, pp. 433-435.
8. S. B. Gershwin et al., "A Control Perspective on Recent Trends in Manufacturing Systems," *IEEE Control Systems Magazine*, April 1986, pp. 3-15.

Biographies (continued)

Department. His job is to analyze manufacturing methods in various production facilities and recommend changes in operational procedures. He received a B.A. from Swarthmore College and an M.S. and Ph.D. from the University of Pennsylvania, all in physics. He joined AT&T in 1966. Mr. Sitarik, before his present assignment, was supervisor of the Manufacturing Systems Engineering Group, responsible for methods and tools for designing and diagnosing manufacturing lines. He has a B.S. from Carnegie Mellon University, an M.S. from Ohio State University, and a Ph.D. from Lehigh University, all in electrical engineering. He joined AT&T in 1963.

(Manuscript received August 2, 1987)