Competitive Semiconductor Manufacturing:
Summary Report on Findings from
Benchmarking Eight-inch, sub-350nm Wafer Fabrication Lines

Robert C. Leachman

Competitive Semiconductor Manufacturing Program
Engineering Systems Research Center
3115 Etcheverry Hall
University of California at Berkeley
Berkeley, CA 94720-1750

January 31, 2002
1. Introduction

The Competitive Semiconductor Manufacturing (CSM) Program at the University of California at Berkeley has made a ten-year effort to benchmark the manufacturing performance of leading semiconductor manufacturers world-wide. The focus of the CSM Program is on the “front-end” (wafer fabrication and electrical die sort) stages of the overall manufacturing process, since these stages account for about 90% of the capital cost and 80% of the cycle time of manufacturing. The front-end also accounts for most of the technology development associated with semiconductor manufacturing.

This report summarizes findings from benchmarking ten fabrication plants processing eight-inch silicon wafers to fabricate digital devices with feature sizes of 350nm and smaller. All of these fabrication plants were constructed in the 1993 – 1996 time frame. Performance data were collected from each participant for some or all of the time frame 1996 – 2000. The individual identities of the participants are confidential, but each participant is consistently identified across various metrics using the labeling scheme M1, M2, … , M10.

This is a preliminary report on the findings. It provides technical metrics on manufacturing performance and a discussion of key practices that underlie leading performance. A more complete final report will be issued subsequently that will include additional metrics on equipment performance and additional analyses of the trade-offs between various technical metrics as well as economic interpretation of performance.

Sponsors of this phase of the CSM survey include SEMATECH, the Electronics Industry Association of Japan, the Semiconductor Research Institute of Japan, Taiwan Semiconductor Manufacturing Corp., United Microelectronics Corp., Winbond, Samsung Electronics Corp., Micrus, Inc., Cypress Semiconductor, and ST Microelectronics. The conclusions expressed herein do not necessarily express the views of any sponsor.
2. Metrics of Manufacturing Performance

To develop appropriate metrics of manufacturing performance, it is useful to consider the underlying economics of manufacturing.

Semiconductor manufacturing is capital-intensive. In the fabrication of advanced digital products, investment in processing equipment and manufacturing facilities account for about 65% of manufacturing cost, utilities account for about 15%, materials account for between 10 and 15%, indirect labor accounts for between 5 and 10%, and direct labor accounts for 5% or less.

The two most significant performance factors determining manufacturing cost are the manufacturing yield and the equipment throughput. Other important factors influencing cost include human productivity and the productivity of clean-room floor space.

Manufacturing yield measures the fraction of input to the manufacturing process that is transformed into salable product. It expresses the fraction actually produced of the total amount of product that theoretically should have been produced from the given input of blank silicon substrates (“wafers”).

Equipment productivity measures the achieved unit output rate of a particular type of equipment asset. It may be expressed as a throughput figure (units of output per machine per day), or as a dimensionless efficiency score.

In addition to product cost, manufacturing performance also affects sales revenues. Sales prices for semiconductors and the electronic products that incorporate semiconductors tend to decline rapidly with time, 25-35% per year. In this situation, the speed and reliability of manufacturing can have a very large impact on sales revenues. Those semiconductor vendors able to offer new products earlier than their competitors enjoy substantially higher average selling prices.
Time-to-market is a function of the time required to develop and qualify new manufacturing processes and products, the time to install and qualify new equipment, the time to ramp up yield and volume, and the elapsed time of the manufacturing process itself (i.e., the so-called manufacturing cycle time). By reducing these times, sales revenues can be increased.

Moreover, customer willingness to buy from a particular semiconductor vendor is influenced by the timeliness of the vendor’s product deliveries relative to the delivery dates promised to the customer (i.e., the so-called on-time delivery performance), since each customer is concerned about the time-to-market for his products. A semiconductor company offering superior delivery performance may gain market share or may be able to charge a higher price for its manufacturing service.

The CSM metrics of manufacturing performance are known in the industry as the “Berkeley metrics.” The CSM metrics are made available on the CSM web site\(^1\) and are followed closely by most major semiconductor manufacturers as well as by many suppliers to the industry. The metrics are summarized as follows.

2.1. Yield Metrics

The overall front-end manufacturing yield is the product of the wafer-level yield of the fabrication process (line yield) and the die-level yield of the electrical die sorting process (die yield).

**Line yield**

Line yield expresses the average fraction of wafers started that emerge from the fabrication process flow as completed wafers available for electrical die sorting. Higher line yields reflect more useful output per unit input and thus higher productivity. As an accounting practice, the line yield of a process flow is computed for a given time frame

---

\(^1\) [http://euler.berkeley.edu/csm/esrc](http://euler.berkeley.edu/csm/esrc)
as a ratio in which the numerator is the number of completed wafers, and the denominator is the sum of the number completed wafers and the number of wafers scrapped.

Wafers may be unintentionally broken or scratched during processing due to malfunctions of wafer handling mechanisms or operator mishandling. Line yield losses also may be the result of processing cycles that are aborted due to equipment malfunction or from wafers rejected by quality inspections that detected mis-processing. Mis-processing can result from human errors (wrong machine recipe selected, processing step repeated or skipped, etc.) as well as from out-of-control process conditions. Thus the line yield metric reflects the level of equipment reliability, the degree of process control, and the level of operator proficiency. It also may reflect the degree of focus in the factory, since a factory operating a single process flow needs to make far fewer adjustments of the equipment than one operating multiple process flows.

All other factors being equal, line yields tend to be higher in large fabs operating a low number of distinct process flows, whereby processing equipment may be dedicated to performing a single process recipe. In all fabs, improvements in line yields can result from the introduction of more sophisticated process control, the automation of recipe download, the introduction of controls preventing the processing of the wrong lot, improvements to equipment reliability, and from increasing operator understanding of processing procedures and troubleshooting instructions.

The number of layers of circuitry varies according to the complexity of the product. All things being equal, one would expect the line yield for a product with more layers to be lower. Thus a normalized line yield metric is used, expressing the line yield per twenty layers of circuitry. That is, the given line yield for a process flow is converted into a metric score for the line yield per twenty layers using the following formula:

\[ LY_{20} = LY^{(20/ML)} \]
where $ML$ is the number of mask layers in the process flow, $LY$ is the reported line yield for the process flow, and $LY20$ is the calculated line yield per twenty layers. A factory-level score is computed as the weighted-average of $LY20$ scores for the major process flows operated by the factory (weighted by the wafer starts in each flow):

$$FLY20 = \frac{\sum (WS_i) \cdot LY20_i}{\sum (WS_i)} \text{,}$$  \hspace{1cm} (2)

where $WS_i$ is the number of wafer starts per week in process flow $i$, $LY20_i$ is the line yield per twenty layers for process flow $i$, and $FLY20$ is the weighted-average fab line yield.

**Die yield**

Rarely does every integrated circuit that is printed on a completed wafer function properly. In the electrical die sorting (EDS) process, each integrated circuit (“die”) on the wafer is tested (“probed”) to see if it functions, and inoperative die are identified to be discarded later. The fraction of the total die on a wafer that pass EDS is termed the die yield of the wafer. Typically, die yield accounts for a larger loss of potential output than does line yield.

Causes of die yield loss may be classified into (1) large-area faults, and (2) losses due to contaminating particles lodged in the circuitry, the latter often referred to as simply defects. Large-area faults arise from a failure of the processing equipment to correctly perform the desired process operation (e.g., over-etching or under-etching, excessive or inadequate deposition of dopants, lack of registration of photo layers, etc.). These faults show up as wholesale or patterned areas of the wafer surface with few or no dice performing as desired, or even as entire wafers or entire lots of wafers with no working dice. On the other hand, particles are much smaller than the area of a die; a single particle may cause the circuit to have a short or an open, thereby causing the die to fail. Thus defects result in more randomly distributed patterns of failed die over the wafer surface.
In the life cycle of a typical CMOS process flow, failed-process problems are usually corrected early in the life of the flow as process and equipment control are improved and/or as the products are redesigned to better conform with the capabilities of the equipment. Failed-process problems may persist over the life of the flow if a fab deliberately utilizes low-cost, older-generation equipment that is marginally capable of performing the desired process, and/or if the product design deliberately violates one or more “design rules” governing the process. Apart from such cases, failed-process problems tend to dominate early-life of the process, while particle losses are more predominant in mature CMOS process flows.

Historically, people, the ambient clean room air and water used to clean wafers were through to be the primary sources of contaminating particles. But gowning, clean room airflow and water protection have been improved, and it is now generally believed that 80% or more of fatal defects land on the wafers while they are resident in the processing chambers of the fabrication equipment. Pressure spikes in processing chambers, leaks in vacuum chambers during evacuation, flakes given off by handling mechanisms, air bubbles in photoresist applications, contaminated liquid and gas flows, etc. are examples of particle problems. Thus overcoming particle losses also is an equipment issue.

Particle-related losses can be easily modeled as randomly distributed over the wafer surface. However, equipment operating improperly or with improper controls may spew out dense bursts of particles (sometimes characterized by certain “signature” patterns on the wafer), causing the die populating a large portion of the wafer to fail. In general, defects are not distributed uniformly over the surface of a wafer, nor are they distributed uniformly from wafer to wafer.

Thus it is not easy to sort out yield losses by process-failure and particle causes, as the pattern of failed die on a wafer could have resulted from many combinations of causes. The die yields reported by the CSM participants are simply the observed yields at EDS, reflecting both particle-related and large-area faults.
Although particle-related losses do not fully account for total die yield losses, they are nonetheless significant, and everything else held equal, a product with a larger die size may be expected to have a lower yield, since it has a higher probability of hosting a fatal particle. To compare die yields among CSM participants, it is necessary to normalize for die area. The CSM Program uses the basic Murphy defect model to convert actual die yield recorded for a major product in each process flow into a defect density score expressing the number of fatal defects per square centimeter of wafer surface area for the process flow. Specifically, the Murphy model expresses the fractional die yield (i.e., the fraction of gross die that pass the electrical tests at EDS) as

$$Y = \left(\frac{1 - e^{-AD}}{AD}\right)^2,$$

where $Y$ is the observed die yield\(^2\), $D$ is the defect density over the wafer surface and $A$ is the die area, expresses in square centimeters. The CSM Program uses this basic Murphy model to report trends in defect density among the participants. Although the Murphy model and other defect density models were designed as a means of analyzing particle-related losses, the CSM Program uses the defect density metric as a normalized measurement of total die yield loss. Low die yields (equivalently, high defect densities) suggest that parametric problems probably dominate particle-related defects, and in such cases, “defect density” is a misnomer; in any case, it is an oversimplification.

As fabs introduce process flows to fabricate circuits with finer and finer geometries, particles with smaller and smaller sizes can be fatal. Thus an improved level of particle control is necessary to achieve the same die yield for a finer geometry, and accordingly, process flows need to be classified by geometry for comparison of defects.\(^3\)

---

\(^2\) Some participants categorize overall EDS losses into sort line yield (accounting for entire wafers that get discarded in EDS) and electrical die yield (accounting for die losses from wafers not discarded). The “observed die yield” herein is the product of these two factors.

\(^3\) By the same token, there are new sources of failed-process problems and they become more difficult to mitigate as geometry is reduced.
A further classification that is necessary is to segregate process flows for making logic devices from those making memory devices. This is because of the substantial amount of redundancy built into memory circuits, whereby failed memory cells can be disconnected and replaced with spare cells included in the product design for this purpose. (This operation, known as “laser repair,” is performed using lasers in the EDS process.) For memory devices, the final die yield is called the “repaired yield,” while the die yield before the laser operation is termed the “virgin yield”.

**Integrated yield**

To obtain an overall front-end yield score for each participant, the CSM Program defines an integrated yield metric as follows. For each process flow operated by each participant, the defect density score $D$ derived using (2) from the participant’s given die yield and die size is plugged back into equation (2) along with a die area $A = 0.5 \text{ cm}^2$ to estimate a die yield $Y$ the flow would achieve if it were producing a product with a die area of 0.5 cm$^2$. This die yield is then multiplied by the line yield score $LY20$ for the flow computed using (1). Mathematically, it is expressed as

$$ IY = (LY20) \left( \frac{1 - e^{-0.5D}}{0.5D} \right)^2, $$

where $D$ is the calculated defect density for the process flow, $LY20$ is the calculated line yield, and $IY$ is the resulting integrated yield for the process flow. Like defect density, scores for this metric are classified by geometry and by logic vs. memory.

**2.2. Equipment Productivity Metrics**

**Wafer throughput**
Photolithography typically comprises the highest concentration of capital expense of all equipment types in a wafer fab and is most commonly the long-run equipment bottleneck. Thus measurement of photolithography equipment productivity is sometimes used as a proxy for measuring the throughput efficiency of a fab. The expensive machines that perform the exposure step are known as *steppers* and *scanners*.

Although photolithography usually represents the greatest concentration of capital expense for equipment in a fab, the steppers and scanners are not the bottleneck equipment types in all CSM participating fabs. Even when the equipment set was installed in the fab with the expectation that photolithography would be the capacity limiter, changes in process specifications, set-up requirements, and/or changes in demand mix since fab start-up may have shifted the bottleneck away from photolithography. Thus limits on the utilization of photo machines may be imposed by a lack of capacity available at other equipment types.

While equipment performance is often tracked in terms of percentage equipment *utilization*, this metric has its drawbacks. Many CSM participants are able to engineer significant reductions in reticle (mask) set-up times and in wafer processing times, and such productivity gains are not measurable using the utilization metric. Since equipment throughput may be increased not only by increasing equipment utilization but also by reducing mask set-up and wafer processing times, a metric of true wafer throughput of the equipment is more useful. Lacking data on the actual process times at the participants, the CSM Program utilizes simple measure of wafer throughput, explained as follows.

The various types of photolithography equipment (pre-clean benches, pre-bake ovens, photoresist coating tracks, exposure machines, develop tracks, inspection and metrology equipment) are operated in sequence to carry out the photolithography process. The exposure machine is generally the slowest and most expensive. “Steppers” derive the name from the way they work. To expose circuitry patterns in submicron geometries with sharp focus, it is not optically feasible to expose the entire wafer at once. Instead, small
groups of die are exposed in sequence, whereby the machine “steps” over the surface of the wafer performing multiple exposures in order to expose all of the die on the wafer.

The total time for a stepper to process a wafer is a complex. It is a function of the field size the stepper is configured to expose, the number of die that fit into a field, the number of die on the wafer, the alignment offsets across the surface of the wafer, and the desired exposure energy. Exposure energies (and hence exposure times) vary by layer, e.g., metal layers take longer than implant layers. Thus there is variation in the total amount of stepper processing time embodied in each product, and one must expect some variation in total wafer throughput depending on the product mix.

Some CSM participants argue that stepper and scanner throughput scores need to be conditioned based on the number of mask changes that are necessary, i.e., based on the variety of die types that are produced. A machine “set-up” involves a particular reticle (mask) to be used that must be inserted in the machine, and, in some fabs, tested before allowing repetitive use. While a stepper may accommodate up to a dozen or so masks in its internal magazine, it is nevertheless argued that a fab that must process hundreds of product types per day will of necessity experience more lost stepper time than another fab producing only a handful of die types.

However, other participants have eliminated the requirement to perform a test exposure following a mask change, and they have mostly or fully automated the mask changes (in the sense that processing is minimally interrupted by mask changes). Set-ups in these fabs require something on the order of 30-45 seconds rather than 20-30 minutes consumed in other fabs. In fact, some of the CSM participants achieving high stepper throughput scores have very high active die counts in their fabs.

The CSM Program defines the stepper throughput (scanner throughput) metric as the average number of wafer operations performed per machine per calendar day, considering only non-rework masking operations. The metric can be computed for a particular type of scanner or stepper, or for a set of types. For a given set of exposure
machines, the number of non-rework wafer operations $SWO$ for a given process flow is estimated as

$$SWO = \left(\frac{WS}{7}\right)(NS)(LY')$$

where $WS$ is the average number of wafer starts per week for the process flow, $NS$ is the number of mask layers in the process flow performed by the given set of machines, and $LY'$ is an inflated line yield given by

$$LY' = \frac{(1.0 + LY)}{2},$$

where $LY$ is the reported line yield for the process flow. (This inflated line yield allows for half of the total line yield loss to load photolithography equipment, or equivalently, the average wafer that is scrapped makes it through exactly half of the layers before being discarded.) Considering all process flows, the total stepper operations per day is summed up, then divided by the number of steppers in service to obtain the value of the metric.

The CSM Program reports the throughput scores of its participants for 5X I Line steppers, 4X DUV steppers and all stepper/scanners. However, the reader may note that the approach taken in equations (5) and (6) can be applied to any type of equipment in order to estimate the throughput of that equipment type.

**Die throughput**

The CSM Program combines the wafer throughput metric with the integrated yield metric to obtain an *integrated throughput metric* that may be computed for an equipment type or set of types of interest. For example, when applied to steppers, this metric reduces the wafer throughput to account for line yield and die yield losses, in effect measuring the equivalent number of perfect wafer layers processed per machine per day, assuming a
product with die area of 0.5 cm² was in production. For each process flow, the number of good wafer operations $GWO$ is estimated as

$$GWO = \frac{WS}{7} \left( NS \right) \left( LY20 \right) \left[ 1 - e^{-\frac{(0.5)D}{(0.5)D}} \right]^2,$$  \hspace{1cm} (7)

where $WS$ is the average number of wafer starts per week for the process flow, $NS$ is the number of steps in the flow performed by the equipment set of interest, $LY20$ is the line yield metric calculated for the process flow using (1), and $D$ is the defect density calculated for the process flow calculated using (3).

This integrated throughput metric helps to assess the effectiveness of the participants in addressing the intricate trade-offs between line yield, die yield and equipment throughput in an attempt to maximize good die output.

**2.3. Labor Productivity Metrics**

It is common at many semiconductor manufacturing companies to measure worked productivity in terms of the number of wafer processing steps completed per person per day. (Terminology for one wafer completing one process step includes “activity” or “move”. Thus some fabs speak of the number of activities completed per operator per day, or the number of wafer moves completed per operator per day.) A difficulty with this form of metric is that the granularity of what constitutes a “step” varies from company to company.

As a practical alternative, the productivity of clean-room staff can be measured similar to equipment productivity. That is, we can measure the number of wafer layers completed per person per day. The number of wafer layers $WL$ completed per day for a given process flow is estimated as
where $WS$ is the average number of wafer starts per week for the process flow, $WD$ is the number of working days per week, $NS$ is the number of mask layers in the process flow, and $LY'$ is the inflated line yield as defined by equation (6). The total number of wafer layers completed per working day in the fab can be computed by summing $WL$ for each process flow.

We define the direct labor productivity metric as

$$DLP = \frac{\sum WL_i}{NO}$$

(9)

where $WL_i$ is the number of wafer layers completed per day in process flow $i$ and $NO$ is the number of full-time-equivalent (FTE) operators employed by the fab. We define the total labor productivity metric as

$$TLP = \frac{\sum WL_i}{NS}$$

(10)

where $NS$ is the total full-time-equivalent staff employed by the fab, including operators, technicians, engineers, supervisors, managers and administrative staff.

2.4. Space Productivity Metric

Clean-room space is expensive. It represents not only significant capital expense, but also significant operating expense (especially utilities) to maintain the required environment and airflow. We can measure space productivity similar to the previous productivity metrics, simply by dividing the wafer layers completed per day by the amount of clean room space. The space productivity metric $SP$ is defined as
where $WL_i$ is the number of wafer layers completed per day in process flow $i$ and $CS$ is the total square footage of clean-room floor space.

### 2.5. Speed-Related Metrics

#### Cycle Time Metric

Fabrication cycle time is the elapsed time from when blank silicon substrates enter the first step of the fabrication process flow until a completed wafer exits the last step. It includes all elapsed time, not just time the wafer is actually undergoing processing. At all CSM participants, wafers are transported between steps in lots or cassettes, most commonly accommodating 25 wafers. Except for line yield losses, lot integrity is typically preserved all the way through the process flow. Thus the cycle times that are tracked and statistically averaged by CSM participants are *lot cycle times*.

To account for differences in the number of manufacturing steps needed to make semiconductors of varying complexity, the CSM program tracks cycle time per wafer layer. *Cycle time per layer*, defined for each process flow, measures the average duration (expressed in fractional working days) that is consumed by production lots of wafers from time of release into the fab until time of exit from the last step of fab (before EDS), divided by the number of mask layers in the process flow. To obtain a metric score at the fab level, the CSM Program computes

$$
FCTPL = \frac{\sum_i (WS_i) CTPL_i}{\sum_i WS_i},
$$

(12)
where $FCTPL$ is the weighted-average fab cycle time per layer. $WS_i$ is the number of wafer starts per week in process flow $i$ and $CTPL_i$ is the cycle time per layer in flow $i$.

**Other speed-related metrics**

Two other speed-related metrics tracked by the CSM program have straightforward definitions. Process development time $VT$ is the time required to qualify a new process flow, measured from time of first wafer start until date of qualification. Yield ramp time $RT$ is the elapsed time from process qualification until mature die yield is achieved. These durations are reported directly by the CSM participants.
3. Actual Manufacturing Performance

Figures 3-1 through 3-21 display metric scores for ten CSM participants during the period 1996 - 2000. The ten participating fabrication plants include fabs operated by AMD, Conexant, Cypress and Micrus (a joint venture of IBM and Cirrus Logic) in the USA, ST Microelectronics in Europe, and Nec, Oki, TSMC, UMC and Samsung in Asia. All of these fabs were built during the period 1993-1996, and all operate CMOS process flows to fabricate digital devices on eight-inch (200mm) wafers. Generally, each fab was selected by its owning company to participate in CSM because it was the best performer among fabs it operated of the vintage indicated above. Their individual identities are disguised using labels M1 – M10 consistently across the figures, e.g., M1 represents the same fab in all figures. Scores are tabulated by month or by quarter, enabling the reader to track performance through time.

3.1. Yield

Figure 3-1 displays line yield scores. As may be seen, fab-wide average line yield per twenty layers reaches scores of about 98 percent. There seems to be some closure in performance over time among the participants.

Figures 3-2 through 3-7 display Murphy defect density scores. Scores are broken out by technology generation (0.5um, 0.35um and 0.25um) as well as by memory vs. logic. Not all participants reported die yield data in every category: for example, fabs M1, M5 and M8 produced only memory devices; fabs M7, M9 and M10 produced only logic devices, while the others reported production of some of each. As may be seen, defect densities of logic devices in each generation are driven down over time to about 0.2 fatal defects per cm² or less. Defect densities of memory devices in each generation are driven down to about 0.1 fatal defects per cm² or less (measured after laser repair). For production of a die with an area equal to 0.5 cm², these defect densities correspond to die yields of about 91% and 95%, respectively.
Note that most of the participants ultimately tend to reach the same mature defect density, but there is considerable disparity concerning when each fab was able to commence production in a given technology and how long it took them to reach mature die yield. Those companies able to qualify new process technology at an early date and ramp to mature die yield quickly probably enjoyed much higher sales revenues.

Figures 3-8 through 3-13 display integrated yield scores for the participants, combining line yield and die yield performance. These graphs further emphasize the disparity in starting times, starting points and yield ramp times for each technology. Convergence to a common mature yield is evident in most categories. As can be seen, mature performance for memory devices approaches 93% and for logic devices it approaches 89%.

### 3.2. Equipment productivity

Figure 3-14 displays wafer throughput scores for I-Line 5X steppers operated by the participants. These machines are the most numerous exposure machines in use at the participants. As may be seen, there is a wide variation in scores: two fabs (M2 and M3) achieved more than 1,000 wafer exposures per day per machine, while the others could only reach 600 or less wafer exposures per day per machine. Figure 3-15 displays similar data for 4X DUV steppers, a more advanced machine than the I-Line 5X stepper. Again, the range of peak scores is over 1,000 to less than 600.

Figure 3-16 displays wafer throughout scores considering all photo exposure machines in the fab. Most participants “mix and match” photo exposure machines, employing cheaper and faster (and less precise) machines to perform the easiest photo exposure steps while utilizing the most expensive and sophisticated machines to perform the most difficult steps. Three fabs achieved throughputs of about 900 wafer exposures per machine per day, three others achieved about 600 wafer exposures per machine per day, and the others achieved less.
Finally, Figure 3-17 displays integrated wafer throughput scores for photo exposure machines at the participants. The scores integrate yield losses with the wafer throughput scores considering all photo exposure machines. As may be seen, the gap in scores has narrowed from 600-1000 to 400-700, indicating that some fabs with lower stepper throughputs achieve higher yields.

### 3.3. Space productivity

Figure 3-18 displays space productivity scores. Fab M3 achieved almost 0.4 mask aligns per square foot of space per day; most participants achieved half that much or less.

### 3.4. Labor productivity

Figures 3-19 and 3-20 display direct labor productivity and total labor productivity scores, respectively. As may be seen, there is a very wide range of scores. Direct labor productivity ranges from 85 wafer layers per operator per day down to less than 20. Total labor productivity ranges from 45 wafer layers per person per day to less than 10. It is apparent that indirect staff number about as much as direct staff at the participants.

### 3.5. Speed-related metrics

Figure 3-21 displays cycle time per layer scores for the participants. Two fabs (M1 and M2) achieved cycle times below 1.5 days per layer; most participants achieved cycle times in the range 2 – 2.5 days per layer.

Graphs are not provided for $VT$ (process qualification time) and $RT$ (yield ramp time) scores, since these are metrics computed once for each new process flow rather than computed continuously over time. Performances observed for $VT$ need to be handicapped depending on how pioneering is the process technology. For a very new technology, i.e., the first at its geometry, the best observed figure for $VT$ was 7 months, compared to an
average of 12 months. For new process technologies similar to those already in production, the best observed figure was 4 months, compared to an average of 7 months.

Performances observed in yield ramp times need to be handicapped depending upon whether or not other fabs have already ramped up similar process technologies. This is because effective solutions to yield problems become embodied in the hardware, software and consulting offered by suppliers of fabrication equipment. For a process technology introduced about one year after the leader introduced a similar technology, the best observed yield ramp time was 7 months, compared to an average of 12 months.

3.6. Summary of Actual Performance

In the first half of the 1980s, there was considerable alarm in the management of US semiconductor manufacturers, as they faced Japanese competitors achieving superior yields and rapidly expanding capacity with government support. But a decade later, the world had changed.

Compared to CSM scores calculated for the period 1989 – 1995,⁴ there is much more closure in mature yield performance 1996 - 2000. Leadership die yield and leadership line yield are not distinguished by region. However, there is considerable disparity in the starting time and starting yield for new process technologies, as well as considerable disparity in the time required to ramp to mature die yield. In general, all of the speed-related metrics (VT, RT and CT) are major discriminators of performance in the industry. (This implies major differences in sales revenues among the participants, even when their products are similar.)

---

We also have seen that equipment and labor productivity also are major discriminators of performance. This suggests that there were major differences in finished wafer costs among the participants.
Figure 3-1. Line Yield

Line yield per 20 layers (percent)

Time

M1
M2
M3
M4
M5
M6
M7
M8
M9
M10
Figure 3-2. CMOS Logic Device Defect Density
0.45 - 0.6 micron CMOS process flows

Defect Density (fatal defects per square cm)

Time

95 96 97 98 99 00
Figure 3-3. CMOS Logic Device Defect Density
0.35 - 0.4 micron CMOS process flows

Defect density (fatal defects per square cm)

Time

M2
M3
M4
M7
M10

2001
Figure 3-4. CMOS Logic Device Defect Density
0.25 micron CMOS process flows

- M3
- M9
- M10

Defect Density (fatal defects per square cm)

Time

98 99 00 01
Figure 3-5. Memory Device Defect Density (after repair)
0.45 - 0.5 micron CMOS process flows

Defect Density (fatal defects per square cm)

Time

M2
M3
M5
M8
Figure 3-6. Memory Device Defect Density (after repair)
0.33 - 0.4 micron CMOS process flows
Figure 3-7. Memory Device Defect Density (after repair)
0.25 - 0.29 micron CMOS process flows

Defect density (fatal defects per square cm)

M1
M3
M5
M8

Time

97 98 99 00 01
Figure 3-8. CMOS Logic Device integrated Yield
0.45 - 0.6 micron CMOS process flows
Figure 3-9. CMOS Logic Device Integrated Yield
0.35 - 0.4 micron CMOS process flows
Figure 3-10. CMOS Logic Device Integrated Yield
0.25 micron CMOS process flows
Figure 3-11. Memory Device Integrated Yield
0.45 - 0.5 micron CMOS process flows
Figure 3-12. Memory Device Integrated Yield
0.33 - 0.4 micron CMOS process flows

Integrated Yield (after repair)

Time

96 97 98 99 00 01
Figure 3-13. Memory Device Integrated Yield
0.25 - 0.29 micron CMOS process flows
Figure 3-14. I-Line 5X Stepper Productivity

- M1
- M2
- M3
- M4
- M5
- M6
- M7
- M8
- M9
- M10
Figure 3-15. DUV Stepper Productivity

- M3
- M8
- M9
- M10

Wafer operations per stepper per day vs. Time (1997-2001)
Figure 3-16. Stepper Productivity (all types of steppers)
Figure 3-17. Integrated Stepper Throughput
Figure 3-18. Space Productivity

Mask layers per sq ft per day

<table>
<thead>
<tr>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
<th>M7</th>
<th>M8</th>
<th>M9</th>
<th>M10</th>
</tr>
</thead>
</table>

Time

95 96 97 98 99 00 01
Figure 3-19. Direct Labor Productivity

Mask layers per direct labor per day

- M1
- M2
- M3
- M4
- M5
- M6
- M7
- M8
- M9
- M10

Time

95 96 97 98 99 00 01
Figure 3-20. Total Labor Productivity

Mask layers per total labor per day

Time

95 96 97 98 99 00 01
Figure 3-21. Cycle Time Per Layer

[Diagram showing cycle time per layer for different months (M1 to M10) over time from 1995 to 2001.]
4. Key Practices Underlying Performance

The CSM survey teams spent two or three days at each participating fab. During this period, they toured manufacturing facilities, interviewed a cross-section of factory staff, and held sessions to review engineering, managerial and organizational practices in various areas. These areas include the introduction of new process technology, process control, yield improvement, equipment efficiency improvement, cycle time reduction, on-time delivery improvement, computer-integrated manufacturing and automation, teams and work groups, and the development of human resources.

Comparing findings from these sessions to the manufacturing performance scores, the CSM team identified six key practices correlated with leadership performance. These practices, discussed in turn below, are as follows:

- Automate information handling, and make manufacturing mistake-proof.
- Collect detailed process, equipment and test data, integrate the data and analyze it statistically.
- Wisely manage the development and introduction of new process technology.
- Reduce lost time and reduce process time on steppers and other bottleneck equipment.
- Implement intelligent scheduling and WIP management.
- Reduce division of labor, up-skill the workforce and develop a problem-solving organization.

4.1. Automate information handling

*Information handling* concerns the transfer of instructions to operators and machines concerning what processing activity to perform and how to perform it, and the transfer to engineering and managerial databases of information concerning the results of processing and the monitoring of equipment and process. Information handling can be performed manually or it can be automated (i.e., electronic). Automation of information handling is
strongly correlated with performance in the CSM survey. The typical progression of automation of information handling is as follows:

- **Auto-recipe down-load.** A *recipe* is the specific machine settings and instructions to carry out a processing cycle. To initiate a processing cycle, the recipe identifier may be manually entered into the processing machine by an operator, or it may be automatically down-loaded by computers. When automated, the chance for human error in recipe selection is sharply reduced. Most CSM participants operate with 100% of their process equipment utilizing auto-recipe down-load. At several fabs, recipe parameters stored in process machines or in cell controllers could be automatically and instantaneously updated as well. At others, recipes had to be updated machine by machine, a time-consuming task that exposes manufacturing to the risk of incorrect or inconsistent recipes across machines.

- **Automated WIP tracking.** The audit trail of work-in-process (WIP), i.e., the recording of when each step was performed on each manufacturing lot, by what machine and what operator, is known as *WIP tracking*. WIP tracking systems are used by operators to identify the candidate lots awaiting processing at each equipment type, to record the selection of a lot for initiation for processing, and to record the completion of the process step on the lot. In most fabs, WIP tracking is manual in the sense that keyboard entry is used to retrieve and record these data. In leading fabs, these functions have been partially or totally automated, eliminating keystroke operations. This saves cycle time and improves data quality.

- **Automated metrology upload.** After completion of major process steps, there typically are one or more measurement steps to insure the process was performed correctly. Alternatively, it may be desired to record one or more measurements of actual conditions during processing. These data may be the subject of statistical process control as well as off-line engineering analysis. In most fabs, these data are keystroked into process control systems and/or into engineering databases. In leading
fabs, these data are automatically uploaded, increasing data completeness and accuracy as well as saving cycle time.

- **Fully automated and interlocked SPC.** At leading fabs, statistical process control (SPC) is applied to all critical process and equipment parameters in an automated fashion. That is, metrology and process data are automatically uploaded into SPC systems and into engineering databases. There is no keystroking of these data. SPC calculations are automatically made. If an out-of-control (OOC) situation is detected, the SPC system automatically notifies the appropriate personnel to initiate the out-of-control action procedure (OCAP). Moreover interlocks are implemented so that further processing of the affected lots and/or further operation of the process or equipment involved are electronically inhibited. These systems help to contain line yield losses, and they increase organizational focus on process problems.

- **Automated process control.** The final stage of the automation of information handling is the implementation of feed-back and feed-forward process control systems. These systems sharply reduce the involvement of process engineers in the day-to-day operation of the fabrication process, serving to reduce cycle time and increase consistency of the manufacturing process. Typically, such systems give process engineers the confidence to reduce requirements for processing test wafers or send-ahead wafers as well as allow process equipment to be more flexibly used, thereby improving equipment productivity and further reducing cycle time.

### 4.2. Integrate and analyze data

Leading fabs make thorough and swift analyses of manufacturing data. Root causes of losses of yield and throughput are quickly identified and clearly understood. Effective improvements to process, equipment and operation are deduced and implemented.

To carry out effective analysis of yield losses, a complete audit trail of product, process and equipment is consolidated in one relational database fitted with convenient and
powerful statistical analysis tools. Data subject to this analysis include all WIP tracking data, equipment tracking data, in-line metrology and process measurements, and end-of-line product test data, including “maps” of the distribution of failed dice across the wafer surface and by position of the wafer in the lot. As an example analysis, end-of-line die yields by lot may be re-sorted in the sequence the lots passed through a particular process tool, in order to identify correlation between yield and elapsed time since cleaning or other preventive maintenance procedures performed on the machine.

Leading fabs also perform extensive in-line defect monitoring (i.e., scanning the wafer surface for defects), both by means of optical scanning and by means of electrical testing of simple structures printed in the scribe lines between the die on the wafer. Intelligent sampling schemes and classification schemes are developed according to the size and nature of the defects liable to be present at various stages of the manufacturing process. Sometimes, multiple levels of sampling are performed. For example, a relatively speedy electrical test might be performed first; if a significant number of failed structures are detected, then a more time-consuming optical scan would be undertaken to quantify and characterize the defects that are present.

Leading fabs differentiate random defect losses from systematic losses. This is done by analyzing yield loss vs. area and by plotting yield loss vs. die position on the wafer. Systematic losses are mitigated by making modifications or adjustments to product design, process specifications and/or equipment. Once defect losses are characterized and traced to sources, the minimum defect capabilities of process steps and equipment are determined based on the defects found in the best-performing lots. A defect “budget” is established, targeting the defect reduction to be achieved in each process step and equipment in order to achieve an overall satisfactory die yield.

4.3. Manage the development and introduction of process technology

If a new process technology fails to provide satisfactory yields, no amount of manufacturing efficiency can make up for the loss of revenue. There were a number of
instances among the CSM participants when the transfer from R&D of a new process technology went poorly in the sense that the process did not yield for an extended period of time. During this time, substantial revenues were no doubt missed. A critical area of managerial and engineering practice concerns the development and transfer of new process technology so as to facilitate a smooth and prompt ramp-up into mass production. Several effective practices in this regard emerged from the CSM survey.

Copy exactly. Almost all of the observed major problems in new process introduction had to do with differences between manufacturing facility and development facility. Under the copy exactly policy, equipment sets, recipes, chemicals and materials used in mass production are required to be identical to that used in process development. Even information systems and are made identical and databases are electronically copied. Strict enforcement of this policy mitigates the risk that manufacturing is unable duplicate process results achieved by R&D. In the ideal case, a new fabrication facility is constructed and populated with enough equipment to support development of the process technology. Once development is complete, the manufacturing organization is brought in and the equipment is replicated up to a scale to permit mass production.

In its purest form, the copy exactly policy is suitable only for the case that the market for the process technology can fill an entire fabrication facility, and it is undesirable to tinker with the technology specifications over its life. Developed and perfected by Intel, the policy is quite effective for their microprocessor business, but is an awkward fit for many others.

Concurrent development. Under this policy, both development of new process technology and mass production using older technologies are carried out in the same facility. The information systems are more flexible and sophisticated than in typical fabrication facilities, able to accommodate both manufacturing and development activity. Specifications for processing development lots, including experimental specifications, are input to the information system so that their handling is usually little different from production lots.
The manufacturing staff also is more sophisticated than typical. Processing of development lots is performed by operators, and the installation and early use of new equipment needed to process development lots are handled by sustaining equipment engineers rather than development engineers.

Once a new process technology is qualified for mass production, there is no transfer from development to manufacturing. Manufacturing has been involved from the start and is already proficient at operating the technology.

This practice was perfected by foundries operating a variety of process technologies at moderate volumes and is well suited to their business.

A general theme of good practice in process development and transfer is what we term *complexity management*. In leading fabs, there is a deliberate attempt to minimize the number of engineering variables that must be simultaneously confronted. The timings of changes to wafer size, process technology and device are always staggered. For example, a new process technology is transferred and ramped up using a pure shrink of a mature device printed on the same size wafer. This way, no device or wafer size variables are introduced that might obscure or slow down recognition and resolution of process problems. Similarly, a change wafer size will be made while freezing the process technology and the set of devices in production. Whenever more than one of these three variables (process, device, wafer size) was changed at the same time, the transfer and ramp-up was difficult and very time-consuming.

Another good application of complexity management arises in process development itself. An entire process technology is a sequence of process *modules*, each module the portion of the process involving a major equipment step plus associated preparation steps (cleaning, heating, coating, etc.) and post-processing steps (stripping, cleaning, metrology, etc.). In each generation of process technology, development engineers face a
choice of re-using certain modules from the previous generation or replacing it with a new module.

Suppose the average process module is suitable for use in three generations of process technology. Then, on average, each generation of process technology should involve one-third new modules to be developed and should re-use two-thirds modules from the previous generation. In general, the more new modules in a technology, the more difficult is the task of process integration, i.e., tuning and adjustment of the modules so they work together most effectively. From the point of view of complexity management, the number of new modules should be kept close to one-third in each generation. A new technology with, say, 80 percent new modules will be very difficult to de-bug and it will be difficult to ramp up the yield. A new technology with, say, 10 percent new modules might be easy to qualify and ramp up, but this means more than the average number of modules will become obsolete and need to be replaced in the succeeding generation. The next generation is likely to be very difficult to integrate.

4.4. Eliminate lost equipment time

Historically, the major focus for improving equipment efficiency in the semiconductor industry concerned the reduction of non-available time (e.g., breakdowns and subsequent repairs and re-qualifications, preventive maintenance). Industry-wide, there has been some success in this regard; equipment availability has improved. In the current survey, we find that efforts at leading fabs have expanded to address the reduction of idle time and time lost during utilization of bottleneck equipment. Such lost time includes gaps between consecutive production lots, time lost waiting for test or sampling results, time consumed by recipe changeover, and time lost because the machine is processing at a slower rate than ideal. Effective practices to re-capture this lost time are described as follows.

Linked photolithography cells involve a series of process steps (pre-cleaning, photoresist coating, pre-baking, exposure, develop, post-cleaning) linked together into a sort of
transfer line. Different device-layers may involve different settings at several or all of these steps. In some fabs, the entire linked system must be flushed before inputting wafers to be processed with a different device-layer, a loss of 30 to 40 minutes. In other fabs, software has been introduced to switch over the recipes of various stations one by one as the new wafer type moves down the transfer line. Using this software, if the exposure tool is accommodates multiple reticles and the changeover of reticles is automated, then there is little or no lost time when changing over from processing one device-layer to another. This permits a “cascade” of lots requiring different photo recipes to be processed continuously through the linked photo cell. The productivity of the bottleneck photo tools can be dramatically improved, especially when there is a wide device mix in production.

Another loss of potential throughput at photo tools is the requirement to perform sample or “send-ahead” wafers. Under this requirement, the first wafer of a photo lot is exposed, developed and inspected before exposing any more wafers from the lot. Based on the inspection, the parameters of the photo recipe may be adjusted before processing the remainder of the lot. The exposure machine may be held idle for 15 – 20 minutes while the first wafer is sent ahead, or it may be changed over to process a previous recipe and then changed back again after the results of the inspection are available. The enforced idleness obviously reduces throughput; the repeated (and otherwise unnecessary) reticle changes also reduce throughput, especially if recipes can not be cascaded without flushing the linked photo cluster.

Leading fabs have introduced feedback control mechanisms at photolithography that automatically adjust photo recipe settings based on metrology results. These process control systems provide photo engineers with the confidence to sharply reduce or totally eliminate the requirement for send-ahead wafers, thereby improving throughput of the exposure machines.

As mentioned above, another important source of lost throughput arises when process tools operate at rates less than ideal. Weak power supplies, inadequate gas flow, dirty
optical paths, sensor malfunctions and software bugs all can contribute to excessive time to complete process cycles. Without automated monitoring of process cycles performed by the machine, such losses remain largely out of sight and out of mind.

To detect such losses, leading fabs measure equipment performance in great detail. Using SECS-II ports, machine event logs are extracted and uploaded to an analysis database. These logs record start and stop times of the many individual elements of the process cycle, e.g., pumping load lock down to vacuum, robotic arm transfer to processing chamber, etching to endpoint, etc. Logs from different cycles and from different machines are compared to detect abnormally long duration for a particular element, evidence of a hardware or software problem in the machine.

From the review of large amounts of actual performance data, from vendor specifications for equipment performance, and/or from basic scientific principles, leading fabs establish formulas that express the theoretical or ideal time to complete a process cycle as a function of the recipe parameters. These theoretical times become the basis for monitoring the processing rate in real time of constrained tools. If actual elapsed time is significantly longer than theoretical, an alarm is triggered, indicating there has been some deterioration inside the machine. The internal problem has not caused the machine to fail yet, but the processing rate has deteriorated, suggesting investigation and correction would be beneficial.

For the traditional area of increasing equipment availability, leading fabs have effectively employed the total productive manufacturing (TPM) paradigm. Monitoring and inspection of equipment have been increased, cleanliness of equipment has been improved, and, as a result, failures have been reduced. Cleaning, inspection and monitoring procedures that were never performed or were performed by equipment maintenance technicians became tasks routinely performed by machine operators. Maintenance technicians became more involved in equipment trouble-shooting, repair,

---

5 For example, exposure time is a function of desired exposure energy and lamp intensity, and implant time is a function of desired dose and beam current
major preventive procedures and projects to modify and improve machines. These tasks were previously the domain of equipment engineers. Released from these tasks, equipment engineers devote time to establishing maintenance and inspection standards, planning and engineering improvements to the equipment, and training and mentoring of their staff.

4.5. Plan and schedule intelligently

Fabs with low cycle time and superior on-time delivery performance have excellent WIP management and scheduling practices. Important characteristics of these practices are described as follows.

At most fabs, floor scheduling follows the lot dispatching paradigm. Due dates are assigned to lots at time of release into the fab. A target cycle time is established for the entire process flow, then broken down by step. The lateness of each lot may be judged based on comparing its due date to the remaining target cycle time to the end of the flow. Lots that are furthest behind (or least ahead) of schedule are prioritized, either using a least slack rule or a critical ratio rule.

A weakness of the lot dispatching paradigm is that it is difficult to keep lot due dates up to date. If the demand changes, if a downstream lot is scrapped, or if a lot passes another lot containing the same product, then the due dates become incorrect. The lot dispatching paradigm may drive the factory to allocate capacity for the purposes of putting lots of the same product back in their original order, even though this does not improve on-time delivery and it does not reduce customer-perceived cycle time. This capacity could have been allocated to reduce cycle time and improve on-time delivery of other products.

Leading fabs follow a different paradigm that we shall term as the WIP management paradigm. Rather than lot due dates, scheduling works to a target fab out schedule. The target cycle time and the fab out schedule imply a target profile of WIP for each product. At each process step, the priority of a lot waiting at that step is established based on the
surplus or deficit of downstream WIP of the product in the lot. Changes in demand, lot
scraps, and lots passing one another are properly accounted for.

Another weakness at most fabs concerns the allocation of target cycle time to process
steps. This is typically done proportional to process times, actual cycle times or simulated
cycle times. Leading fabs recognize that specification of the target cycle time profile is
equivalent to specifying the target WIP profile. Care is exercised to allocate buffer cycle
time to steps performed on bottleneck equipment rather than to steps performed on low-
utilization equipment. Thus the target WIP profile at leading fabs concentrates a higher
percentage of total fab WIP at the bottleneck equipment type. This permits a lower total
level of fab WIP necessary to sustain a given utilization level of the bottleneck and hence
a given fab throughput rate.

In some process technologies, only a subset of the machine of a particular type are
qualified to perform a given process step. For a different step, a different and perhaps
overlapping set of machines is qualified. This happens when considerable engineering
work is required to “tune” and qualify each machine to perform a given step. In this
situation, simple prioritized dispatching may result in WIP left over and machines left
over that are incompatible. At leading fabs, the likelihood of such a problem is mitigated
by employing scheduling systems that simultaneously consider all machines and all WIP
waiting for a given equipment type. This larger view is optimized to find the best
allocation of WIP to machines. These scheduling systems enable a reduction in the WIP
level (and hence in the cycle time) required to maintain a given utilization level.

The cycle time performance and capabilities of a fabrication line are to a great extent
determined by the installed equipment set, the machine qualifications that are made, and
the wafer input to the line. There are significant differences in the detail and accuracy of
capacity planning and production planning⁶ among the participants. At most fabs,
capacity planning and production planning either ignores or approximates detailed

⁶ “Production planning” means the process of establishing a target fab out schedule and a schedule for the
release of new wafers into the line.
machine qualifications and tooling issues such as reticles and probe cards. Process times are rough averages, or are replaced by rule-of-thumb volume limits. At leading fabs, analysis of such issues is rigorous. Machine qualifications and tooling additions are explicitly planned and scheduled in order to sustain a desired cycle time performance and a desired volume ramp. Actual process times are extracted from machine logs, and formulas are developed to predict process times for new specifications. Wafer releases are approved only if proven to be feasible with respect to equipment and process capabilities and cycle time targets.

4.6. Develop a problem-solving organization

Leading fabs recognize yield and throughput losses more quickly, and they more quickly devise and implement improvements to recoup those losses. Considerable effort is expended to improve the problem-solving ability of the work force and to increase its initiative for uncovering opportunities to improve. There are two general themes we have recognized in this endeavor.

*Up-skilling the workforce.* Leading fabs obtain more participation in problem solving. Everyone in their organizations feels a responsibility to assist the overall engineering effort to improve process and equipment. At the lowest level, the TQM and/or TPM paradigms are used to organize operators and technicians into teams for problem detection and solution. These teams are mentored and advised by engineers. The teams carry out process control and investigate to find root cause of process problems. They carry out inspection and light maintenance of machines, learning more about proper machine condition and operation, and they investigate to find root cause of equipment problems. Often, operators have increased education levels, with two years of technician schooling subsequent to their high-school graduation.

Recognizing their more sophisticated role, their job titles are revised. “Self-sustaining technician,” “Lady setter,” and “Self-help lady” are examples. The ideal is that process or equipment trouble occurring on the line is recognized and rectified by the operating staff.
whenever possible. Technicians and sustaining engineers still are involved in problem solving, but now the operating staff are in effect their assistants, to which they can delegate much of the effort, especially monitoring, data collection and other in-the-fab work.

Given the new role of operators, equipment technicians are able to take on work formerly handled by sustaining engineers or maintenance managers. Sometimes they are given new titles, such as “Equipment owner” or “Equipment key man.” The idea is that this kind of technician now writes the maintenance and inspection specifications that operators and junior technicians follow. This kind of technician, on duty every shift in every area, becomes the on-the-floor equipment expert to which operators and other technicians turn for help. Problems that might otherwise await the return of a staff engineer on the next weekday morning get addressed when they happen.

Reduced division of labor. The merging of operator and equipment or process technician jobs, and the merging of equipment technician and sustaining engineer jobs, as described above, are examples of reductions in the division of labor. At some participants, manufacturing and equipment maintenance departments were formally merged; at others, this reduction in the division of labor was less evident from the organizational chart. Regardless of their organizational structure, leading fabs have reduced the number of hand-offs required to get problems recognized and solved.

An important kind of reduction in the division of labor concerns coordination or consolidation of different engineering skills. The general intent is to accelerate problem solving by instilling broader knowledge and responsibility in staff engineers. This was undertaken in different ways at different participants. Some fabs had merged equipment engineering and process engineering groups. The process knowledge of former equipment engineers was increased; the equipment knowledge of former process engineers was increased; and their ease and facility of working together was increased. Some fabs had implemented a “module management” organizational structure, defined around the various processes of fabrication (photolithography, etching, diffusion, thin
film deposition, etc.). Each module included all the operators, technicians, supervisors, and sustaining engineers that work in the area. Each module team can flexibly assign its staff to address problems that may arise and to develop the skills and teamwork of its organization.

Another important means of reducing the division of labor concerns yield analysis and yield improvement. In most fabs, a yield engineering group is charged with recognizing and analyzing yield problems. Once the cause of the problem is isolated to a particular process step or equipment type, the problem typically is handed off to a process or equipment engineer to determine root cause and a permanent fix. In leading fabs, those with detailed knowledge about the process and the equipment join this problem-solving effort early. Moreover, these individuals have strong facility with the data analysis tools, so they can accelerate characterization of the problem and its resolution. In some cases, process engineers carry out the yield analysis; in others, the yield engineers have strong process and equipment knowledge themselves and can carry through to root cause analysis and permanent fix. Regardless of the organizational structure, the key idea is the broad skills of the individual engineer: facility and skill in statistical analysis, and knowledge to propose appropriate process or equipment modifications, concentrated in the same individual.

5. Conclusions

Dr. Robert Helms, CEO of International SEMATECH, recently remarked, “It used to be that the big companies eat the small. Today, the fast run over the slow.” The findings of the eight-inch CSM survey confirm this observation. The most important performance differences concern speed. Leading companies introduce new process technology earlier. They qualify the technology faster. They ramp up the yield and the volume more quickly. Their cycle time is shorter. And their process times are shorter, so their equipment throughput is higher.

These results are achieved by means of six key practices. They automate information handling. They wisely manage the development and transfer of new process technology, minimizing the number of simultaneous engineering variables and mitigating the difficulties of technology transfer. They integrate and analyze process, equipment and test data to more swiftly uncover and resolve losses of yield and throughput. They detect and eliminate lost equipment time. They intelligently schedule and manage WIP, and they carefully plan their equipment installations, qualifications and volume ramps. Finally, the leading fabs develop strong problem solving organizations, up-skilling their organizations and reducing the division of engineering tasks and the division of technical knowledge.