An Integrated Framework for Yield Management and Defect/Fault Reduction

Charles Weber, Bizhan Moslehi, and Manjari Dutta

Abstract—**An integrated framework for yield management and defect/fault reduction is presented. A 3-D space consisting of a quality axis, a process integration axis, and a scaling axis encompasses all process and manufacturing parameters. Cross-functional teams of process, equipment, operations, and materials personnel proactively explore this space, and provide process engineers with a stable and capable environment for process development and manufacturing activities.**

I. Introduction

YIELD improvement is a critical component of the drive towards manufacturability. Traditional, somewhat "reactive" approaches to yield improvement rely heavily on responding to yield crashes by "crisis management" and "fire fighting" efforts, which are often hasty, too late, and expensive. In more successful approaches yield improvement efforts are proactive, preventive, consistent, and continuous: yield is managed by the implementation of well planned strategies capable of achieving targeted defect density, fault density and yield milestones.

A defect is defined as anything that *may* cause a product to fail, whereas a fault is any form of defect that *induces* product failure. Thus, inadequate process control, reliability problems, particulates, and other forms of contamination in the wafer environment are all sources of defects, and defects are the cause of all faults. However, only a fraction of all defects become faults. A yield management philosophy that promotes the detection, prevention, reduction, control, and elimination of sources of defects, therefore, contributes to fault reduction and yield improvement.

Defect and fault density requirements vary substantially with the maturity of a process and the minimum feature sizes of the associated products. This paper describes a comprehensive framework for yield management that customizes yield improvement strategies to the nature and state of a process. Implementation of the resulting defect/fault control methodologies enables the concurrent pursuit of leading edge process development, customer prototype manufacturing, and cost-competitive volume production.

II. Integrated Framework for Yield Management

A fundamental framework for yield management consists of the 3-D space shown in Fig. 1, where the axes represent quality, process integration, and scaling. Defect density and the capability indices (C_p and C_pk) are the primary parameters that characterize the quality axis. Intrinsic data cycle time and product complexity are associated with the process integration axis. Minimum feature sizes of VLSI/ULSI circuits quantify the scaling axis.

Manuscript received December 9, 1993; revised October 14, 1994.

The authors are with the Technology Development Center, Integrated Circuits Business Division, Hewlett-Packard Company, Palo Alto, CA USA.

IEEE Log Number 9409809.

A. The Quality Axis

The quality improvement effort consists of four discrete phases: research (R), control (C), transfer (T), and manufacturing (M) [1]. Throughout all phases the defect density of all failure modes decreases, and the degree of control over all parameters increases. Fig. 1 gives typical values for circuit defect density (DD or D) and C_pk. The defect densities correspond to simple functional yield and the C_pk numbers are based on specification limits that reside infinitesimally within the brink of values that induce an electrical fault.

In the research phase, specifications are not determined yet, and distributions of parameters are typically neither stable nor normal. Therefore, at this stage, systematic process problems are the dominant contributors to defect density. The capability indices (C_p and C_pk) are meaningless and cannot be utilized to quantify the degree of process control. The research phase ends with the establishment of a full VLSI/ULSI process together with a set of "at-risk" design rules, and specifications for each process parameter [1].

During the control phase, distributions of all relevant process parameters are narrowed and centered. The process usually yields working samples when the average of the C_pk values of all process parameters is around 0.5. In addition, circuits with an area of 0.4 cm^2 will yield with statistical significance and the defect density of these VLSI circuits range from 5 to 10 defects/cm^2. By the end of the control phase C_pk for most parameters runs around 1 and the systematic process problems have been minimized. Random failures become the dominant contributors to defect density at ~1 defect/cm^2. At that stage of process development, reliability issues must be settled, and design rules must be finalized. Circuits with areas of 3 cm^2 yield with statistical significance, which implies fully functional customer samples can be fabricated in sufficient quantities [1], [2]. Key quality milestones during the control phase include producing working samples (WS) in the early stages, engineering samples (ES) in the middle of this phase, and customer samples (CS) at the end.

In the process transfer phase (from research & development to manufacturing), fabrication of prototypes and development of process variants can all occur concurrently. Quality improvement efforts must continue during the transfer phase,

0894–6507/95$04.00 © 1995 IEEE

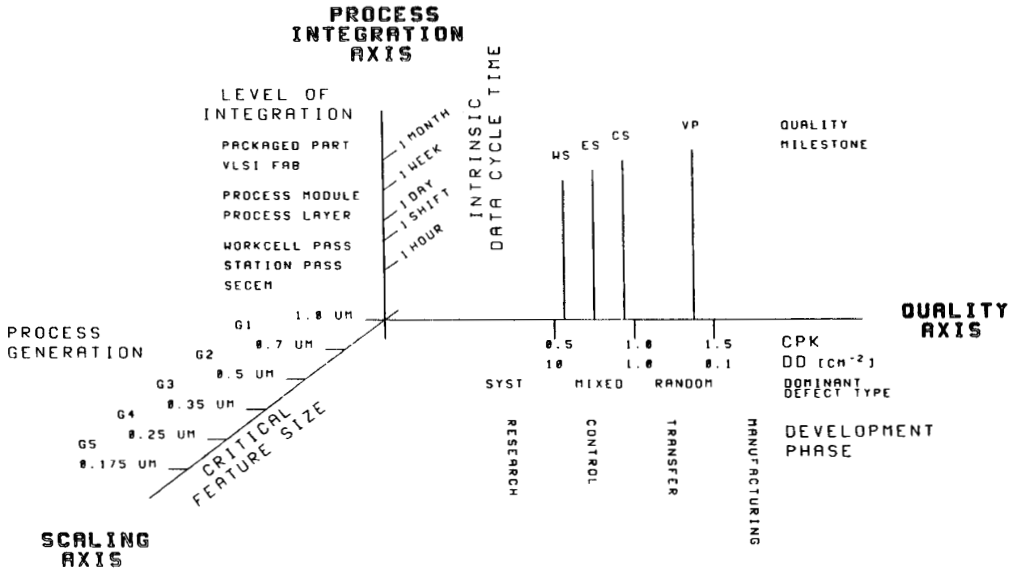


Fig. 1. 3-D space comprising an integrated framework for yield management. The acronyms "WS", "ES", "CS", and "VP" respectively stand for "working sample", "engineering sample", "customer sample", and "volume production".

because profitable high volume production (VP) of VLSI/ULSI circuits requires a fault density of about 0.1 to 0.3 defects/cm^2. At that point, a circuit with an area of 3 cm^2 will yield between 40–80%. Such fault densities can only be achieved if process problems of a systematic nature have been all but eliminated, which requires C_pk values greater than 1.5 on relevant parameters. This is known as the manufacturing phase.

TABLE I
MOST COMPLEX PRODUCTS OF SPECIFIC PROCESS SEGMENTS

Process Segment Type (Level of Integration)	Most Complex Product
Packaged Part	Fully functional and reliable VLSI circuit in a package
VLSI Fab	Fully functional VLSI circuits on a wafer
Process Module	Parametric test structures with multiple conducting layers
Process Layer	Parametric test structures with a single conducting layer
Workcell	Multiple films or single-level, non-electrical structures
Station	Modified pre-existing film or structure
SECEM	None

B. *Process Integration Axis*

Process integration is defined as combining a series of short process segments into longer process segments that yield more complex products. Products are defined as tangible entities, which are produced by a process segment, and can be evaluated quantitatively. Processed wafers and packaged VLSI circuits are thus examples of products of different process segments.

The nature of the most complex product fabricated by a process segment associates the segment with one of seven levels of integration exhibited in Table I. Table II indicates which methods of defect detection are appropriate for each level of integration, and Fig. 1 illustrates that data cycle time increases geometrically with each level of integration. A designer of an experiment must therefore tradeoff the ability to obtain data from complex products with the cycle time of the experiment.

The quality of the lowest level of integration, site, environment, consumables, equipment, and materials (SECEM), is ideally determined on a real-time and *in situ* basis. For example, particle detectors that continually monitor clean room air will automatically sound an alarm on the monitoring computer system of the facility if the particulate count exceeds a specified limit. Vacuum particle sensors or in-line impurity analyzers provide real-time, *in-situ* monitoring in process equipment.

All higher levels of integration are associated with specific process segments that consist of one or more fabrication steps and one or more inspection/metrology steps. In a station segment, for example, wafers pass through a single piece of process equipment that modifies the state of the wafers, and separate metrology equipment performs measurements on the modified wafers. A particle-per-wafer-pass (PWP) experiment or process-induced-particle (PIP) measurement on a single piece of process equipment is an example of a station segment. A surface scanner searches for particles on a set of unpatterned wafers, before and after they are run through a single piece of process equipment with a fabrication cycle or a dummy cycle. The difference in particle counts gives an indication of the effect of the piece of process equipment on the wafers. However, conventional surface scanners cannot evaluate patterned wafers or films with very rough surfaces. Station segments that etch patterned wafers or generate films with rough surfaces therefore require special patterned wafer inspection systems as defect detection tools. Station segments that generate films with rough surfaces can also utilize special grazing angle wafer inspection systems.

TABLE II
DEFECT DETECTION METHODS FOR VARIOUS LEVELS OF INTEGRATION

	<---------- Method of Defect Detection ---------->					
Level of Integration	Real-Time, In-Situ	PWP	Auto Inspect	Parametric Test	Functional Test	Life Test
Packaged Part						*
VLSI Fab				*	*	
Process Module				*		
Process Layer			*	*		
Workcell		*	*			
Station		*	*			
SECEM	*					

In a workcell segment a wafer passes through more than one piece of process equipment before it undergoes a defect detection step. Under ideal workcell conditions one person can perform all aspects of the processing, inspection, and metrology steps without traveling more than six meters. A lithography loop consisting of prime, spin, exposure, develop, automatic inspection, and critical dimension metrology steps is an example of a workcell segment. An etch area workcell segment can consist of hardbake, plasma etch, resist strip, critical dimension metrology, and automatic inspection steps. The sequence wet clean, gate oxidation, oxide thickness measurement, polysilicon deposition, polysilicon thickness measurement, and surface particle count comprises a workcell segment in the films area. The surface particle count can be performed on a test wafer that travels with the lot throughout the workcell segment.

A process layer is a process segment that contains at least one thin film deposition, at most one patterning step, and at least one etch step. The P-active layer, N-doped polysilicon layers, and metal interconnect layers are examples of conducting process layers that are electrically testable. An inter-layer dielectric patterned with contact holes and etched is an example of a nonconducting process layer. It can only be electrically tested if it is sandwiched between two patterned upper and lower conducting layers [3]. Table II indicates that automatic inspection can evaluate the quality of both conducting and nonconducting process layers. Automatic inspection at the completion of every constituent workcell segment of a process layer helps localize sources of defects within the process layer. Automatic inspection of critical layers prior to parametric testing can establish direct correlation between defects and electrical faults, but only if the test pattern contains electronic microstructures that yield with statistical significance [2], [3].

A process module is a process segment that yields circuitry that can be parametrically, but not functionally, tested. Electrically testable process layers constitute the simplest process modules. Process segments that yield NMOS, PMOS, or bipolar devices as product are examples of more complex modules.

A process segment that yields products that can be functionally tested without dicing wafers is defined as a VLSI/ULSI fab segment. These segments are typically divided into four process module segments: isolation, gate, contact, and interconnect. The isolation module consists of all process steps from the start of the CMOS process through field oxidation. The gate module covers all process steps from field oxidation through the completion of the field effect transistor. In the majority of state-of-the-art, ASIC-oriented semiconductors this occurs after silicidation of the polysilicon gates and the source/drain area. The contact module consists of all process steps from the completion of transistors through the etching of the first metal layer. A local interconnect layer is, therefore, part of the contact module. All process steps between metal-1 etch and the time the wafers exit the fabrication facility constitute the interconnect module.

Multi-purpose mask sets that evaluate all four modules have been developed [1], [3]. They can characterize more than 95% of all electrical fault mechanisms of a CMOS manufacturing process. In addition, they may contain structures that examine and analyze issues such as plasma damage, thermally dependent dielectric breakdown [4] and electromigration [5]. Many reliability problems can thus be addressed by fabricating process module segments with relatively short data cycles. However, the multipurpose mask sets cannot identify complex electrical faults such as latch-up, whose detection requires fabrication of a complete CMOS device.

Routine processing of a VLSI/ULSI circuit such as a RAM through the complete fab segment covers most of the complex electrical fault mechanisms of a manufacturing process, and provides valuable information on random fault density. However, many reliability problems can only be characterized by dicing the wafers, packaging the dice, and performing life tests. A process segment that goes through all these process steps is called a packaged part segment. Fig. 1 depicts customer samples at a higher level of integration than working samples and engineering samples, because they require the routine completion of packaged part segments.

Table I lists the most complex products that can be obtained from each level of integration. The SECEM level, due to its real-time nature, requires no wafer processing, and thus yields no products. A station segment can either add a film to a wafer or make a simple modification to an existing film/structure, because it contains no more than one fabrication step. The short series of fabrication segments that comprise a workcell segment can add a few films to a wafer or make slightly more sophisticated modifications to a film/structure, but they cannot yield any electrically testable product.

Fabricating the simplest electrically testable devices, single-conducting-layer parametric test structures, requires the realization of a conducting process layer segment on a previously insulated wafer. Process module segments contain more than one conducting layer and can thus generate more complex parametric test structures like transistors and contact strings

as products. VLSI fab segments yield fully functional VLSI circuits on a wafer. Packaged part segments, which consist of a VLSI fab segment and additional dicing and assembly steps, yield fully functional and reliable VLSI circuits in a package.

C. The Scaling Axis

Persistent efforts to increase the density and performance of integrated circuits have resulted in continual scaling of the minimal integrated circuit feature sizes. Historically, integrated circuit scaling has been driven by the release of successive generations of DRAM's, which has required the critical IC features sizes to shrink by a factor of about 0.7 every 2–3 years. This trend is expected to continue for the foreseeable future, although lately the evolution of microprocessors has been driving the scaling of key interconnect layers [6]–[8].

Fig. 1 illustrates the picture most major IC manufacturers face in the 1990's. The labels G1, G2, G3, G4, and G5 on the scaling axis respectively denote the 4, 16, 64, 256 Mbit, and 1 Gbit DRAM equivalent process generations. The numerical values adjacent to the generational labels represent the expected critical feature sizes corresponding to these process generations.

As critical feature sizes shrink, increasingly smaller defects may develop into faults. Unfortunately, defect frequency grows geometrically with defect size [9]-[14], which makes the potential impact of defects on VLSI/ULSI yield even more severe. The detection, prevention, reduction, control, and elimination of smaller and smaller defects thus becomes a key component of any yield management strategy, even if increasingly expensive tools and methods must be deployed. For example, semiconductor manufacturers are currently purchasing automated defect detection/inspection equipment that can detect defects as small as 0.1–0.2 μm. These tools are capable of exploring the defect environment of G3, G4, and G5 on low levels of process integration. However, they fall short of completely exploring G5.

Exploring the defect environment at higher levels of integration also requires a coherent test structure program. Fortunately, the basic test structures that debug VLSI/ULSI processes are well known [15]-[18] and scaleable. Electrical test structures that detect random defects that cause the most common faults in integrated circuits have been converted into multipurpose structures, and assembled as multipurpose test chips on a CAD station [3], [19]. Many of these designs have been scaled to cover three consecutive process generations [1]. Scaling the chips required only minor design modifications that incorporated new physical phenomena. The effect of defects on IC product yield has also been studied extensively [20]–[23]. Specifically, VLSI/ULSI memory chips follow yield models that are well understood. They have thus become commonly used vehicles for yield improvement that can be scaled from process generation to process generation.

D. Unique Defect Environments

Every axis in Fig. 1 can be divided into discrete domains and ranges. The quality effort undergoes four distinct process development phases (research, control, transfer, and manufacturing). The process integration axis encompasses seven distinguishable levels of integration listed previously. The scaling axis is divided into discrete process generations that cover a specific range of critical feature sizes. The space in Fig. 1 can thus be divided into discrete sectors identified by development phase, level of integration, and process generation. Each sector represents a unique defect environment that requires specific sets of tools and methods.

Test masks, for example, can be customized for each sector [1]. In order to obtain statistically significant defect density data, the area of VLSI circuits on test masks must increase as defect density decreases [2]. The increasing size of VLSI circuits begins to crowd out parametrics, requiring a redesign of all test masks as the VLSI fab segment matures from one process development phase to the next. Every new test mask has less area available for parametrics than its predecessor. Simpler test structures are thus fabricated on short cycle using specialized test masks for process layer and process module development [3]. The test structures on these short cycle test masks also grow with decreasing defect density so they also have to be redesigned at each development phase transition. (Efficient methods of parametric test structure design have been developed specifically for this purpose [19], [24].) Fortunately, most of the structures are scaleable, which permits process engineers to reuse many test chips designs in subsequent process generations. Level of integration (above process layer), process development phase, and process generation, thus specify the nature of a test mask associated with a sector of the space in Fig. 1.

III. Yield Management Strategy

Yield management strategy essentially entails exploring the space in Fig. 1 in an orderly manner. Fig. 2 illustrates how a hypothetical semiconductor supplier could provide an optimal defect environment for both process development and manufacturing under ideal circumstances. A series of planes emerge from the origin of the space at regular intervals of about 2 years. Equipment purchases, designs of experiments, lot scheduling, and human resource allocation are all coordinated to achieve a gradual but timely increase in understanding of the defect environment.

Figs. 3–6 show actual data that support the approach in Fig. 2. Fig. 3 exhibits fault density (FD) trend data for three process generations ($G_N, G_N + 1$, and $G_N + 2$) during various phases of VLSI process development. The solid curves represent five-lot averages of VLSI memory fault density data compiled from more than 6000 wafers over a period of four years. The dotted curves depict test structure data from over 3000 defect monitor wafers [1], [3], [18]. They reflect the sum of the five-lot averages of the two dominant failure modes of the VLSI process: metal bridging and polysilicon bridging. The criteria outlined in the Quality Axis section of this paper determine the development phase boundaries of each process generation.

Fig. 4 illustrates how wafer-to-wafer yield distributions tighten as a process matures. The data reflect realizations of a single VLSI circuit. The mean and standard deviation of wafer-

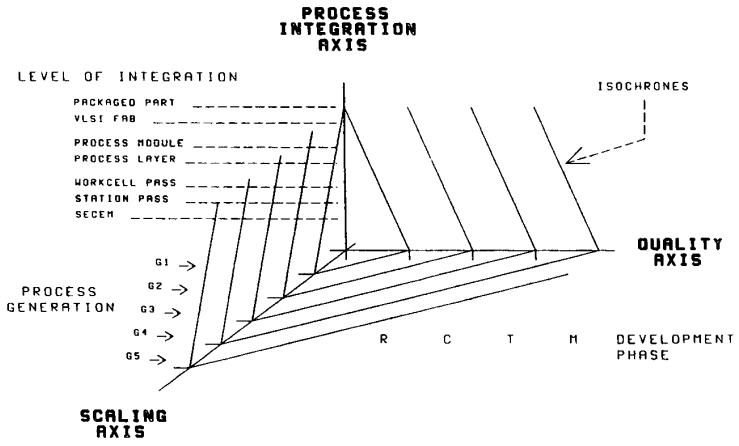


Fig. 2. Defect environments for process development and manufacturing. G1 through G5 refer to successive process generations. The letters "R", "C", "T", and "M" respectively denote the research, control, transfer, and manufacturing phases of process development.

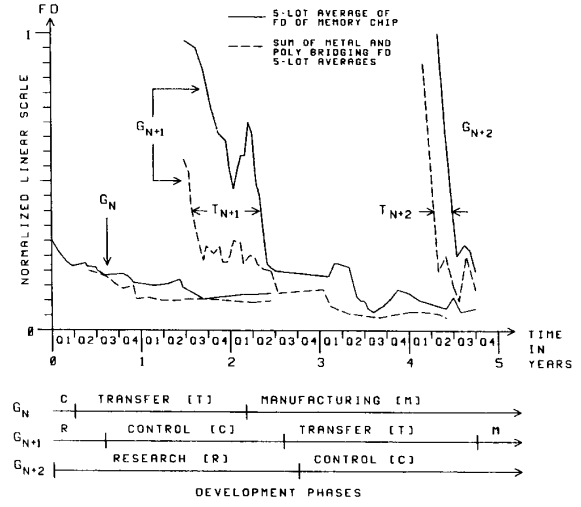


Fig. 3. FD trends during various phases of process development. The lines below the timeline distinguish the development phases of three concurrent process generations G_N, G_{N+1}, and G_{N+2}. Fault density is graphed on a linear, but normalized scale. T_{N+1} and T_{N+2} respectively represent the module integration lag for G_{N+1} and G_{N+2}.

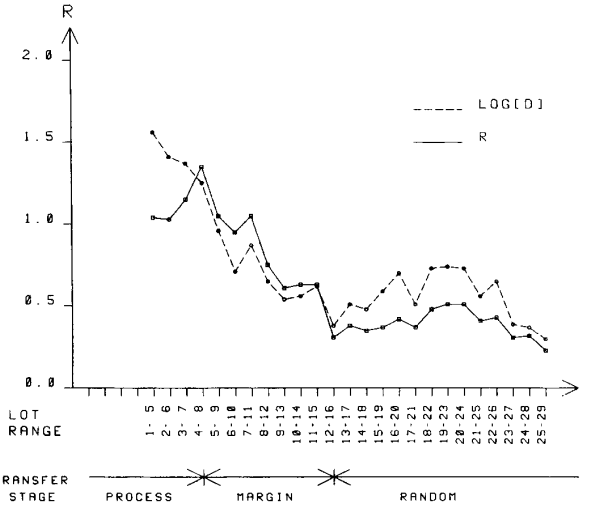


Fig. 4. Defect density and wafer-to-wafer yield distribution of a series of lots during a process transfer. Both curves depict running five-lot averages. The solid line represents the ratio of the standard deviation to the mean of the wafer-to-wafer yield distribution (R). The dashed line depicts the logarithm of the expectation of defect density for the same lots. The base of the logarithm is normalized in order to illustrate the correlation between R and the logarithm of defect density.

to-wafer yield have been calculated for each lot, and the ratio of these numbers (R) is tracked as a running five-lot average. The solid curve in Fig. 4 depicts this parameter on a linear scale, whereas the dashed curve reflects the five-lot average of defect density as calculated by the Poisson yield model on a logarithmic scale. The Poisson yield model states that

$$Y = \exp(-AD) \tag{1}$$

where Y, A, and D respectively denote the yield expectation, the area of the VLSI circuit, and the defect density. The base of the logarithm has been normalized in order to illustrate the correlation between R and $\log(D)$.

The data in Fig. 4 represent the first 29 lots of a technology transfer into a new fabrication facility. The control phase of the process had been completed previously in a research and development facility. During the early stages of the transfer, the process exhibits many of the fundamental process problems identified during the control phase in the R&D facility. These affect the yield of every wafer in a lot, which tightens wafer-to-wafer yield distribution of every lot. Splitting lots at sensitive process steps identifies and helps eliminate known process problems common to all wafers, which simultaneously reduces defect density and broadens the distribution of wafer-to-wafer yield. Once the fundamental process problems have been eliminated, the transfer enters the process margin stage, where all process parameters gradually come under control. Problems only tend to occur on some wafers of each lot; the number of

problematic wafers decreases from lot to lot, and the number of experimental splits diminishes accordingly. Therefore, the defect density decreases dramatically and the wafer-to-wafer yield distribution tightens. In the final stage of the transfer systematic process problems have been all but eliminated, and lots splits occur infrequently. Random defects become the dominant yield limiters, although comparing wafer-to-wafer yield to the binomial distribution suggests the presence of residual systematic components of defect density. Defect density decreases at a much slower rate, and the yield distribution tightens only marginally. The correlation between R and the logarithm of defect density is very strong during the margin and random stages of the process transfer [25].

The data in Figs. 5 and 6 respectively show defect count as a function of defect size as measured by a laser surface particle counter and an automatic defect detection/inspection system. Fig. 5 shows cumulative particle size distribution of the measured PWP values of filtered and static (unfiltered) hot sulfuric/peroxide baths measured down to 0.15 μm on silicon wafers. Results for two systems (System #1 and System #2) are shown. Each system has two baths (Bath A and Bath B). In each system, Bath A is equipped with filtration and recirculation capability and Bath B is a still bath (no filtration recirculation). The results indicate a factor of 5 to 6 improvement in PWP values as a result of filtration. Fig. 6 represents the histogram of optical defect densities as a function of defect size. The data shows measured defects on 93 six-inch wafers at polysilicon post-develop inspection step. About 30 cm^2 of each wafer were inspected. The results of Figs. 5 and 6 indicate that the PWP or defect density values increase with decreasing the defect size.

The approach in Fig. 2 differs from traditional methods of defect reduction in the following ways:

1) The sectors in Figs. 1 and 2 have well-defined boundary conditions. Crossing a boundary at a certain date can thus be designated as a milestone. Defect/fault density targets can therefore be set and met at a previously specified date. This means defect/fault density and its associated yield are being managed.
2) Defect/fault reduction efforts occur continuously and concurrently at all stages of process development, at all measurable dimensions and at all levels of integration.
3) Traditional manufacturing parameters, such as yield, defect density, and cycle time, are treated as process parameters and are thus subjected to similar statistical quality control procedures. For example, wafer-to-wafer yield variability is reported for every lot. This parameter gives a good indication of the maturity of a process. As systematic process problems are removed, the wafer-to-wafer yield distribution approaches the binomial distribution [25]. (See Fig. 4).
4) Quality improvement is pursued most aggressively at lower levels of integration. If the individual components of a process segment are clean and under control, then integration problems are more readily identified. Defect/fault density reduction at higher levels of integration always lags behind that of lower levels of integration. In Fig. 3, G_N represents a process in the transfer phase. The VLSI memory fault density (solid line) slightly exceeds the sum of fault densities of its two dominant process module level failure modes (dotted lines), obtained from test structures. Years 2 and 3 represent the control phase for the next generation process, $G_N + 1$. The module failure modes first drop dramatically as systematic module problems are removed, and settle at a background level of fault density. About a year later, process integration problems are resolved, and VLSI memory fault density drops to a level slightly above the background level. In year 3 the background level starts to drop as $G_N + 1$ is transferred to a brand new fabrication facility. VLSI memory yield follows the same pattern with a lag of $T_N + 1$. Preliminary data indicate that history will repeat itself in the next generation process, $G_N + 2$.
5) A plethora of analogous experiments can be performed at lower levels of integration. For example, the data in Figs. 5 and 6, respectively, indicate that the PWP or defect density values increase with decreasing the defect size. To date, experiments of this type confirm the well known formula

$$D/D_0 = (S/S_0)^{-n} \tag{2}$$

where D/D_0 represents the ratio of defect densities, S/S_0 denotes the ratio between critical feature sizes, and n is a number between 2 and 3 [10], [11]. This relationship frequently holds for the background levels of electrically testable process module failure modes once all systematic process problems have been eliminated.
6) Smaller geometries and defect sizes are first pursued at lower levels of integration. Small defects near the beginning of the process can cause larger defects as additional process layers are added, which may ultimately lead to faults. Conversely, Fig. 5 shows how purchasing a filtration system for a sink, which is intended to remove larger defects, also contributes to the reduction of the number smaller defects.
7) Correlation between levels of integration is established wherever possible, resulting in an early warning system for yield and process problems. Short cycle data at lower levels of integration can prevent yield crashes in real time. Fig. 3 shows a clear correlation between VLSI memory circuit defect density and the defect density of the dominant process module failure modes during the transfer phase of process development.

IV. Defect/Fault Reduction Methodologies

Fig. 2 clearly indicates that defect/fault detection and reduction efforts occur simultaneously across process generations and at all levels of integration. A "total systems" approach to yield management therefore requires placing the necessary defect detection tools into the fabrication facility and/or the test area. Effective utilization of such equipment depends upon sound methodologies for fault identification and reduction, which may differ slightly for each type of fault.

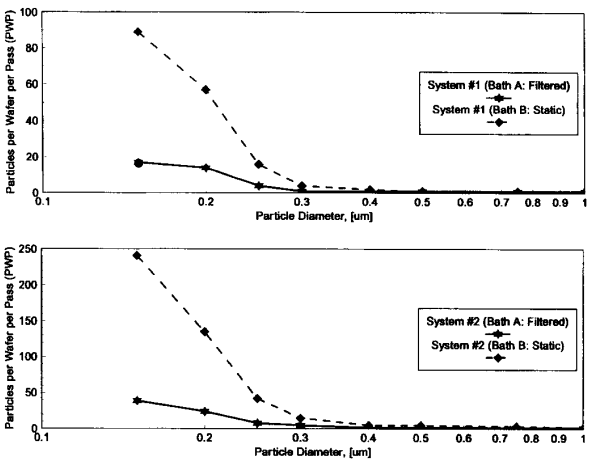


Fig. 5. Cumulative particle size distribution and measured PWP values of filtered and static (unfiltered) hot sulfuric/peroxide baths measured down to 0.15 micrometers. Results for two systems (System #1 and System #2) are shown. Each system has two baths (Bath A and Bath B). In each system, Bath A is equipped with filtration and recirculation capability and Bath B is a still bath (no filtration recirculation).

A. *Faults from Optically Detectable Defects*

Optical/visual defects, especially particulates, are still considered to be the primary source of electrical faults. Therefore, on-line efforts to detect and reduce optical defects constitute a crucial and integral part of any yield management program. An on-line methodology based on the use of automated inspection tools typically consists of the following steps: identify, quantify, prioritize, understand, isolate, eliminate, verify, and document. Table II shows that detection of optical defects occurs at the station, workcell, and process layer levels of integration. Perpetually fabricating and evaluating many of these low-level segments in parallel will identify and quantify defects from most process steps in a VLSI fab segment on short notice, and localize the source of the defects to a few process steps. However, these "short-loop" monitors do not cover all process steps of a VLSI fab segment.

A more comprehensive analysis of a semiconductor process requires inserting automatic inspection steps into a VLSI fab segment at key locations. This sequential analysis illustrates how defects accumulate as a lot moves through the line and how they cause problems for subsequent process layers. For example, small defects in an inter-layer dielectric may be decorated and grow into larger defects in metal layers deposited at a later stage of the process, interfering with metal deposition, metal patterning or metal etch steps. Automatic inspection tools would classify this defect as a metal defect, which may explain why the interconnect module typically exhibits a higher density of large defects than in the isolation, gate or contact modules. (Defects greater than five times the minimum feature size of the semiconductor process are defined as "large.")

A substantial portion of these optical defects may cause electrical faults. It is therefore crucial to separate the "killer" defects from the "cosmetic" defects and to focus on removing the "killers." Large defects receive high priority, as do well understood defects that are known to cause specific electrical faults. For example, LPCVD polysilicon particles are known to cause single bit failures in static memory chips. LPCVD systems are therefore monitored routinely at low levels of integration. An increase in the count of polysilicon particles on an automatic inspection tool will therefore induce an immediate response. Conversely, functional testing of a large sample of VLSI memory chips may identify a high level of "dead column" failures, which may be historically associated with problems in a specific process layer, e.g. metal-2. A rise in dead column failures will thus trigger a high priority search for optical metal-2 defects at lower levels of integration.

A higher level of understanding of the nature of a defect generally helps isolate its source. Therefore, when a large defect is detected by an automatic inspection tool, its coordinates are transferred to an automatic optical review station or a scanning electron microscope equipped with an energy dispersive X-ray analysis system (SEM/EDX) for further analysis. The optical review station is used for classifying defects and developing and implementing Pareto driven defect reduction plans based on defect types and sizes. The SEM/EDX zooms in on the defect and analyzes its composition, which enables engineers to reduce the number of its potential sources. For example, a defect composed of elements present in the dielectric-1 and metal-1 materials may have originated in the dielectric and grown a metal shell during metal deposition, or it may just be a metal defect on a dielectric background. Dissecting the defect by focused ion beam (FIB) would most likely decide which scenario is the correct one.

Experiments that test remedies for defect problems generally start at low levels of integration, whereas verification occurs at higher levels of integration. Initial comparisons between a set of wafers subjected to the remedy and a set of control wafers may occur at the station or workcell level. Running a split lot through a process module segment will quantify the remedy's effect on fault density, and inserting automatic inspection steps into the process module segment will confirm results from lower levels of integration. The effect on the yield of VLSI circuits can only be assessed by completing a VLSI fab segment, and if the remedy could induce a complex reliability problem the VLSI circuits may need to be diced and packaged.

Any defect type that warrants corrective action must be documented for future reference. The documentation should contain pictures of a defect from automatic inspection or review systems and an SEM/EDX system, as well as defect/fault density data from all levels of integration. A precise description of the remedy must also be included, especially if it involves any process changes. Process changes in the transfer or manufacturing phase mandate an engineering change notification (ECN), which requires the signatures of responsible engineers and managers from all pertinent departments.

B. *Faults from Non-Visible Defects*

Functional analysis of VLSI circuits frequently identifies electrical faults, which subsequent failure analysis cannot associate with a defect. For example, a particle about the size

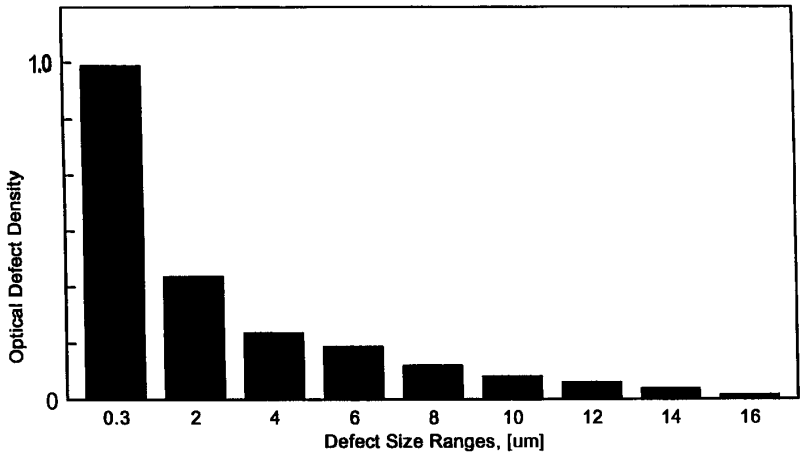


Fig. 6. Histogram of optical defect densities as a function of defect size. The data represent measured defects on 93 six-inch wafers at polysilicon develop inspect. About 30 cm^2 of each wafer were inspected. Defect density data are normalized.

of the gate length of a MOSFET may land on the active area of a future transistor prior to the sensitive channel implant, which could alter the characteristics of the device to be fabricated and cause a single-bit fault. A wet chemical clean or strip process following the channel implant could remove the culprit particle after the damage has been done. From the point of failure analysis a "non-visible" defect (NVD) has created this fault.

For circuits with submicron features faults caused by NVD's can also come from some unexpected sources. For example, bacteria attached to wafers get incinerated without an optical trace when they are processed through a furnace, yet they leave their metal atoms behind. Each bacterium can potentially introduce enough localized contamination to significantly affect the doping levels of a device. A single bit fault may result [26].

To date, the best remedy for faults from NVD's has been prevention. The majority of NVD's are visible at some point in the process. Therefore, continuous monitoring of particulates, bacteria and other defects at the station and workcell level of integration, where they are still visible, will increase the level of understanding of the defects' nature. If they are killers, action needs to be taken immediately. If no obvious correlation between defects and faults exists, proactive elimination of the defects may still be beneficial to product yield, because the defects may cause faults and subsequently disappear. Unless the remedy for removing a certain defect is prohibitively expensive, the benefits of prevention typically outweigh the costs.

Some defects such as metallics may be inherently non-visible. For instance, process materials like wafers, chemicals or gases, contain trace amounts of metal contaminants that can systematically increase diode leakage by acting as recombination centers. Local concentration of certain metals can even cause random single-bit faults. Metallic contaminants also come from equipment sources, such as high temperature furnaces or sputtering in ion implanters. Unfortunately, the remedies for metallic contaminants tend to be quite expensive. They typically involve purchasing ultra-pure materials, *in situ* monitoring and tool-level improvements.

C. *Faults Related to Process Margins*

Historically, faults related to process margins have been eliminated by tightening and centering the distributions of process parameters until random failures become the dominant contributors to defect density. However, as film thicknesses and feature sizes shrink, systematic defects may remain the yield limiters. For example, G4 and G5 in Fig. 1 contain films that are less than 70 Å thick and features that are less than a quarter micrometer wide. Such films and features can be measured much less accurately than those of previous generations. Wafer level nonuniformities on many process steps and intra-chip parameter variations caused by lithography and mask making are also becoming a significant fraction of the error budget for process control. In G4 and G5 process margins may therefore tighten to the point where parametric yield loss may become unavoidable. Engineers may thus be forced to obtain the statistical distributions of individual process steps by realizing electrical test structures with process layer, and process module segments. Engineers would utilize these distributions to simulate and optimize the margin of a VLSI fab process segment on TCAD tools prior to fabrication [27], [28].

V. Implementation (Tactical Planning)

An implementation plan that encompasses many process generations at different stages of development and many levels of integration presents enormous communication and coordination challenges. Therefore, all constituents of an organization must cooperate in order for any yield management program to succeed. The nature of these challenges varies from level of integration to level of integration. A large number of analogous experiments need to be performed at lower levels of integration with data cycles ranging from a few minutes to about a week. For example, process engineers can perform analogous PWP experiments on every piece of thin film equipment in the fabrication facility once per shift (or more often). At high levels of integration the number of

experiment types are much smaller and their cycle times are longer. For example, from the control phase onwards process integration primarily consists of the repetitive fabrication and evaluation of one basic VLSI fab segment. Typically, less than 5% of all process steps deviate from the recipes of the core process.

Hewlett-Packard's Integrated Circuit Technology Development center has created specific entities to address the problems of different levels of integration. Formal departments typically cover the higher levels of integration, whereas cross-functional teams deal with lower levels of integration. All entities try to deploy analogous methods of defect/fault detection and yield management.

A. Cross-Functional Teams

The SECEM team deals with the lowest level of integration. Engineers and technicians from the process, equipment, and materials department jointly specify the limits for all contaminants levels for the clean room air, gases, ultrapure water, acids, solvents, test wafers, and wafer starting material. The team also assures the quality of these items on a continuous basis.

A dozen workcell teams cover the station and workcell levels, each assigned to a specific equipment cluster. They consist of process engineers, equipment personnel, and operations personnel, whose duties also include daily monitoring of individual pieces of process equipment and clusters of process equipment. These teams are in charge of most PWP/PIP monitoring and generating automated inspection data.

Four module focus teams (MFT) cover process layers and the four basic modules of the VLSI fab segment: isolation, gate, contact, and interconnect. They consist of process engineers from relevant workcell teams, process integration engineers that represent the various process generations and members from the yield and reliability groups. A module focus team is thus a multi-phase, multi-generational forum for all process integration issues pertaining to a specific module. MFT's are typically assigned research deliverables for process generation $N + 2$, control deliverables for generation $N + 1$, and transfer deliverables for generation N (N represents the process generation currently in the transfer phase.)

Module focus teams also serve as the first line of defense against electrical faults. They run lots that monitor the faults of each module on a weekly basis with specialized test chips [1], [3]. They also help specify the defect density requirements for constituent workcell teams. Workcell teams subsequently contribute to fault reduction by reducing defect densities in the most strategic areas.

Fig. 3 indicates that VLSI memory fault density tends to lag the fault density of the dominant module failure modes by T_{N+1} in process generation G_{N+1}, and by T_{N+2} in process generation G_{N+2}. These quantities of time, known as the module integration lags of each process generation, depend substantially on communication. For example, a few transistor architecture issues including the choice of an isolation scheme had not been settled at the beginning of the control phase of G_{N+1}, and the nature of some frequently occurring faults was not well understood. Defect task forces, which did not include the architects of the full VLSI process, were assembled on an *ad hoc* basis at the beginning of the second year, resulting in a dramatic drop in fault density of the constituent modules of G_{N+1} (dashed curve) within a few months. However, systematic problems in the VLSI process mandated recipe changes in many process steps, and no formal lateral communication channels between the process architects and the engineers in charge of the recipes existed at the time. As a result, T_{N+1} consistently exceeded 9 months until the fault density of the VLSI memory chip fell below 1 defect per square centimeter.

The previously described MFT's were formed in the third quarter of the fourth year, in order to improve communication between the architects of the VLSI process generations under concurrent development and the engineers responsible for controlling the recipes of constituent process steps. The fault density of the majority of module failure modes of G_{N+2} dropped on short notice; only metal bridging over topography remained a challenge. Management assigned the metal bridging problem to a subset of the interconnect MFT, which came up with a plausible solution within a few weeks. The fault density of metal bridging over topography fell by an order of magnitude, which is illustrated by the dramatic drop in the fault density of module failure modes of G_{N+2} in the second quarter of the fifth year (dashed curve). Due to active participation of the architects of the G_{N+2} VLSI process in module focus teams, the module integration lag T_{N+2} never exceeded 10 weeks, which is slightly longer than the fabrication cycle time of a VLSI fab segment.

B. Analogous Methods

Pursuit of analogous methods reduces the coordination and communication challenges. If all personnel use analogous experimental strategies, common test chips, common metrology tools, analogous measurement methods, and the same data analysis tools, more people will have a better understanding of the resulting data. A delegate from each module focus team therefore meets with his/her counterparts on a regular basis in order to discuss analogous methods issues.

Documentation has been the key to promoting analogous methods. Documentability, testability, and manufacturability have been designed into parametric test chips [19]. An information model, developed to keep documentation as simple as possible, mandated the use of analogous layout components and test software. Generating these components and software has since been automated [24], and an attempt to extend this approach to process documentation is currently under discussion.

C. Management Commitment to Concurrent Process Development

Fig. 2 illustrates that defects and faults must be fought concurrently, continuously, and consistently in all phases of process development, at all levels of integration, and in every process generation. Clearly, prevailing in this tenuous struggle cannot occur without a strong and persistent commitment

by management. Unfortunately, the price of semiconductor equipment and materials continues to escalate from process generation to process generation, which may tempt cost conscious managers to trim an item or two from the budget in times of fiscal discipline. Defect reduction capabilities that tend to affect lower levels of integration typically get cut from the budget first.

The yield management framework in Figs. 1 and 2 illustrates the consequences of any such resource reduction. The space in Fig. 1 is partitioned into discrete sectors, each representing a unique defect environment. The space must be explored sector by sector from the origin outward without leapfrogging over any sector. When a process segment crosses a sector boundary (milestone), equipment, materials, methodologies, and personnel resources for exploring the adjacent sector must be in place, in order for progress to continue. A lack of investment in low levels of integration therefore causes stagnation on all fronts of the yield improvement struggle.

Defect/fault reduction in any sector cannot proceed unless all sectors beneath it have been explored, because the efforts at higher levels of integration are orders of magnitude less efficient. For example, the intrinsic fabrication cycle of a VLSI fab segment is at least four times as long as that of its constituent process module segments and about a hundred times as long as a typical station segment. It thus contains many more potential sources of defects. In addition, a VLSI circuit is a much more sophisticated product than a test structure, which makes quantification and classification of defects in a VLSI circuit much more complicated. It is therefore no surprise that the defect density of the VLSI fab segment for G_{N+1} in Fig. 3 is significantly higher than the defect density of any of its constituent module segments, (although it tends to be much lower than the sum of all constituent module segments).

Removing the systematic faults of the constituent modules at lower levels of integration highlights process integration issues. Fault reduction in the VLSI fab segment without activity at the module level is inherently more complex, because separating module specific problems from integration problems would be much more difficult. Management support of defect reduction at low levels of integration therefore shortens the learning curve for process development.

Management commitment to keeping a previous generation running in a process development facility at low volume after completion of the process transfer can also shorten the yield learning curve, because integrated circuit process development strategies generally favor changing the equipment set as little as possible from process generation to process generation. Achieving a high yield on a VLSI circuit of the previous generation demonstrates that the fabrication facility and the majority of the equipment base are in working order. In the mature process random defects dominate. Differences in defect density between a process in the late transfer phase and a process in the control phase, beyond what is suggested in (2), can thus be attributed to systematic problems in the latter process. The more mature process serves as a baseline for the more advanced process. For example, G_N in Fig. 3 serves as the baseline process for G_{N+1} throughout the control phase of G_{N+1}, and G_{N+1} serves as the baseline process for G_{N+2} throughout the control phase of G_{N+2}.

Concurrent advanced research can also improve the yield of previous process generations. For example, a process in the early control phase may exhibit problems that require a process change to eliminate a specific source of defects. Retrofitting the previous process generation with the process change in a timely manner may improve its yield during the manufacturing phase. A retrofit during the transfer phase should however be avoided, because it can dislocate the performance of the receiving entity.

Cross-functional teams remind management of its commitment to concurrent process development. The SECEM team becomes the primary advocate for defect reduction at low levels of integration, and the module focus teams alert management of any imbalance in resource allocation between process generations. Indiscriminate budget cuts become much more difficult.

VI. Summary

The concurrent pursuit of leading edge process development, customer prototype manufacturing, and cost-competitive volume production mandates the implementation of well planned yield management strategies capable of achieving targeted defect/fault density and yield milestones. A framework for yield management, shown in Fig. 1, consists of a 3-D space whose axes represent quality, process integration, and scaling. The space is divided into distinct sectors, each representing a unique defect environment that requires specific sets of tools and methodologies. Yield management strategy entails exploring the space in the manner illustrated in Fig. 2. Empirical data depicted in Figs. 3–6 illustrate how yield management strategy deviates from traditional approaches to yield improvement. Fault reduction methodologies follow a standardized protocol for optical defects. The proactive approach is stressed for random defects and faults from nonvisible defects, but new methodologies may need to be developed to cope with margin problems in future ULSI processes [27], [28]. Implementation of these defect/fault reduction methodologies presents enormous communication and coordination challenges, which are overcome by the establishment of cross-functional teams, and the pursuit of analogous methods. Finally, management commitment to concurrent process development is considered essential to the success of any competitive yield improvement program.

Acknowledgment

The authors would like to thank all the employees at the IC manufacturing facility in Ft. Collins, Colorado and the technology development center in Palo Alto, California who were involved in processing the wafers that yielded the data exhibited in Figs. 3–6. The authors' gratitude also extends to Richard Clark, Dave Crockett, Dennis Hayes, Nancy Kaw, and Vijay Varma for their help with testing wafers and compiling data. Special recognition goes to Patrick Chang and Jeff Rosner for their leadership role in establishing module focus

teams, and to John Kelly, Alan Marty, Yoshio Nishi and Martin Scott for supporting that effort.

REFERENCES

[1] C. Weber, "Generic test chip formats for ASIC-oriented semiconductor process development," in *Proc. IEEE/ICMTS*, 1993, pp. 247–252.

[2] U. Kaempf, "Statistical significance of defect density estimates," in *IEEE/ICMTS*, 1988, pp. 107–113.

[3] C. Weber, "A standardized method for CMOS unit process development," *IEEE Trans. Semicond. Manufact.*, vol. 5, no. 2, pp. 94–100, May 1992.

[4] K. C. Boyko and D. L. Gerlach, "Time dependent dielectric breakdown of 210 angstrom oxides," in *IEEE/IRPS*, 1989, pp. 1–5.

[5] J. Towner, "The importance of the short circuit failure mode in aluminum electromigration," *J. Vac. Sci. Technol. B*, vol. 5, no. 6, p. 1696, Nov./Dec. 1987.

[6] G. E. Moore, "Progress in digital integrated circuits," *IEDM Tech. Dig.*, p. 11, 1975.

[7] Y. Nishi, "VLSI research and development in U.S. and Japan," in *Proc. Mat. Res. Symp. VLSI V*, 1990, pp. 3–11.

[8] Y. Nishi, "ULSI technology towards the next century: Driven by DRAMs or MPUs," in *Proc. IEDM*, 1992, pp. 13–15.

[9] H. G. Parks and E. A. Burke, "The nature of defect size distributions in semiconductor processes," in *Proc. IEEE/ISMSS*, 1989, p. 131.

[10] R. Glang, "Defect size distribution in very large scale integration chips," in *Proc. IEEE/ICMTS*, 1990, pp. 57–60.

[11] A. V. Ferris-Prabhu, "The role of defect size distribution in yield modeling," *IEEE Trans. Electron Devices*, vol. ED-32, no. 9, pp. 1727–1734, 1985.

[12] C. Stapper, "Modeling of defects in integrated circuit photolithographic patterns," *IBM J. Res. Develop.*, vol. 28, pp. 461–474, July 1984.

[13] C. Stapper, "Modeling of integrated circuit defect sensitivities," *IBM J. Res Develop.*, vol. 27, pp. 549–557, Nov. 1983.

[14] J. P. de Gyvez and D. Chennian, "IC defect sensitivity for footprint-type spot defects," *IEEE Trans. Computer-Aided Design*, vol. 11, no. 5, May 1992.

[15] C. Alcorn, D. Dworak, N. Haddad, W. Henley, and P. Nixon, "Kerf test structure designs for process and device characterizations," *Solid-State Technol.*, pp. 229–235, May 1985.

[16] W. Lukaszek, W. Yarbrough, T. Walker, and J. Meindl, "CMOS test chip design for process problem debugging and yield prediction experiments," *Solid-State Technol.*, vol. 29, no. 3, pp. 87–93, Mar. 1986.

[17] M. Buehler, *et. al.*, "CMOS process monitor," in *Proc. IEEE/ICMTS*, 1988, pp. 164–168.

[18] C. Weber, "Standard defect monitor," in *Proc. IEEE/ICMTS*, 1988, pp. 114–119.

[19] ——, "Standardization of test structure design," in *Proc. IEEE/ICMTS*, 1991, pp. 151–156.

[20] C. Stapper, "The defect-sensitivity effect of memory chips." *IEEE J. Solid-State Circuits*, vol. SC-21, no. 1 pp. 230–234, 1974.

[21] J. Cunningham, "The use and evaluation of yield models in integrated circuit manufacturing," *IEEE Trans. Semicond. Manufact.*, vol. 3, no. 2, pp. 60–71, May 1990.

[22] H. G. Parks, "Yield modeling from SRAM failure analysis," in *Proc. IEEE/ICMTS*, 1990, pp. 169–174.

[23] N. Higaki, S. Ando, and M. Taguchi, "A measurement technique for analyzing bit-line mode soft errors in half-micron design DRAMs," in *Proc. IEEE/ICMTS*, 1990, pp. 179–184.

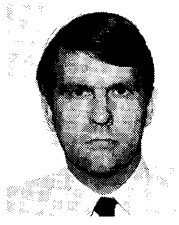
[24] T. Ternesien d'Ouville, J. P. Jeanne, J. L. Leclerq, D. Caloud, and L. Zangara, "Automatic test chip and test program generation: An approach to parametric test computer aided design," in *Proc. IEEE/ICMTS*, 1992, pp. 145–149.

[25] U. Kaempf, "The binomial test: A simple tool to identify process problems," *IEEE Trans. Semicond. Manufact.*, vol. 8, pp. 160–166, this issue.

[26] K. Yabe, *et al.*, "Responding to the future quality demands of ultrapure water," *Microcontamination*, p. 37, Feb. 1989.

[27] C. Yu, T. Maung, D. Bartelink, K. J. Chang, D. Boning, J. Chung, and C. J. Spanos, "Use of short-loop electrical measurements for yield improvement," *IEEE Trans. Semicond. Manufact.*, vol. 8, pp. 150–159, this issue.

[28] D. Bartelink, "Statistical metrology—At the root of manufacturing control," presented at the 40th American Vacuum Society's National Symposium, 1994.



Charles Weber received the A.A. degree in physical science from the American College of Switzerland, the B.S. degree in engineering physics from the University of Colorado, Boulder, and the M.S. degree in electrical engineering from the University of California, Davis.

In 1982, he joined the Hewlett-Packard Company as a process engineer in an integrated circuit manufacturing facility, where he developed his research interest in the productivity of manufacturing systems. In 1985, he became a member of the technical staff of HP's Integrated Circuit Technology Development Center, working primarily in the field of microelectronic test structures. In 1989, he designed a series of productivity-oriented test chips for HP and SEMATECH, including short-loop monitors for process control, defect reduction, and improved manufacturability.

Bizhan Moslehi received the B.S. degree from Arya-Mehr University of Technology, Tehran, Iran, and the M.S. and Ph.D. degrees from Stanford University, all in mechanical engineering. He also holds the M.S. degree in engineering science from Stanford University.

He is a technical contributor at the Technology Development Center of Hewlett-Packard's Integrated Circuit Business Division in Palo Alto, California, where he is working on a program aimed at yield management, defect prevention, and reduction and microcontamination control. Among his major responsibilities are fab defect improvement and microcontamination issues, including wafer cleaning, defect monitoring tools, defect analysis, cleanroom environments, process equipment defect reduction, and defect source identification. Previously, he worked at National Semiconductor, Fairchild Semiconductor, and Applied Materials. At Fairchild, he was a major contributor to the Open Area SMIF Isolation Site (OASIS) project with the aim of elimination of cleanrooms. He is on the editorial advisory board of *Microcontamination*. He has written extensively on a range of microcontamination issues and one of his articles on basic physics of particulates received an IEEE award in 1985.

Dr. Moslehi is an elected member of Sigma Xi. He has been a member of American Association for Aerosol Research (AAAR), Fine Particle Society (FPS), Institute of Environmental Sciences (IES), and the Materials Research Society (MRS).

Manjari Dutta received the B.S. degree in chemistry from the University of Calcutta, India, in 1974.

She is a Process Engineer at Hewlett Packard's Integrated Circuit Technology Development Center at the Deer Creek site in Palo Alto, CA. She has eight years experience in the semiconductor industry in the areas of metallization, ion-implantation and photolithography. She has worked at Toshiba and AMI Gould as a sustaining engineer in metallization and implantation and later joined Avantek's Si bipolar R&D division, where she was involved in the development of its 0.7 μm ISOSAT process. Her major responsibilities at Avantek were the development of a trench isolation and its subsequent planarization process. She was also instrumental in the successful transfer of the complete ISOSAT process from R&D to manufacturing. At present she is involved in defect reduction activities at both the tool and module levels.