Analysis and Reduction of Variability in Scanning Electron Microscopy Measurements of Critical Dimensions

by

Elisabetta Cortesi

S.B., Materials Science and Engineering
Massachusetts Institute of Technology, 1987

Submitted to the Department of Materials Science and Engineering
and the Sloan School of Management
in Partial Fulfillment of the Requirements for the Degrees of

Master of Science in Materials Science and Engineering
and
Master of Science in Management

at the
Massachusetts Institute of Technology
June 1998

Signature of Author ______________________________
Department of Materials Science and Engineering
Sloan School of Management
May 8, 1998

Certified by ______________________________
Professor Lionel C. Kimerling, Thesis Advisor
Department of Materials Science and Engineering

Certified by ______________________________
Professor Lawrence M. Wein, Thesis Advisor
Sloan School of Management

Accepted by ______________________________
Lawrence S. Abeln, Director Master's Program
Sloan School of Management

Accepted by ______________________________
Linn W. Hobbs, John F. Elliott Professor of Materials
Chairman, Departmental Committee on Graduate Students

Analysis and Reduction of Variability in Scanning Electron Microscopy Measurements of Critical Dimensions

by

Elisabetta Cortesi

Submitted to
the Department of Materials Science and Engineering and
the Sloan School of Management
on May 8, 1998
in Partial Fulfillment of the Requirements for the Degrees of

Master of Science in Materials Science and Engineering
and
Master of Science in Management

ABSTRACT

This thesis describes work done during a Leaders for Manufacturing internship at Intel. At the broadest level, this work relates to the importance of controlling and monitoring measurement processes just as one controls the "fundamental" processes being measured. Without such control there can be no confidence in the integrity of the data describing the fundamental process.

More specifically, the project assessed the variability that characterized the critical dimension measurements of one specific layer. It was shown that there was significant operator variability, related primarily to several common types of mismeasurement, that could not be monitored using standard production data. Other potential sources of variability were also investigated but were found to be less important.

Various steps were undertaken to reduce the observed operator variability. As part of this effort an anonymous, automated feedback system was developed and piloted to give operators feedback on their measurements using a standard structure. Although the data from the pilot was inconclusive, the need to monitor measurement variability seems clear.

Finally, the thesis recommends changing the production system so that information on measurement processes can be ascertained from standard production data. It also makes specific recommendations that while not addressing the control of the measurement process, could make the system less susceptible to variation.

Thesis Advisors

Lionel C. Kimerling, Professor, Department of Materials Science and Engineering
Lawrence M. Wein, Professor, Sloan School of Management

Acknowledgments

The author wishes to acknowledge the Leaders for Manufacturing Program for its support of this work.

I also gratefully acknowledge the support of the Intel Foundation's Robert Noyce Memorial Fellowship during my time at LFM.

The work described in this thesis was done at Intel's Fab 9 in Rio Rancho, New Mexico. My thanks to the many people I met, talked to, pestered, and worked with for their unfailing openness and generosity that made the internship a truly wonderful experience. My thanks especially to Eric Stubee, my company supervisor, for guidance and perspective on Intel and the project throughout the internship; to Art Zadra, who became an unofficial supervisor in litho, for his enthusiasm in pursuing measurement issues and for his invaluable assistance with respect to really getting things onto the floor; to Gary Newell, for his immense help with data, automation, SEMs, nuts and bolts of all sorts, and for many laughs; to Wendy Michel, for limitless help getting me up to speed on everything to do with SEMs and for pointing me in the right direction many times; to Dave Williams, for graciously working to finish the automated system pilot after I left, even in the face of a daunting workload; to Nicole Shaver for making me laugh as hard as I can remember and for reminding me that sometimes it's just a karma thing; to Cheryl Morgan for her patience and dedication without which the automated part of the automated system would never have happened; to everyone in the litho process engineering group for helping keep me on track and integrated with the rest of litho; and to all the SEM operators and trainers in Fab 9 whose help and active participation and good humor made this work both possible and rewarding.

I would also like to thank my advisors, Professor Lionel Kimerling and Professor Larry Wein, for their direction and support during the internship.

LFM has been a remarkable experience for me and for this I am grateful to Roger Little for giving me the professional opportunities that helped develop my interest in the business and manufacturing issues that eventually brought me to LFM; to the LFM faculty and staff for their dedication and willingness to help throughout; and especially to all the other '98s for their support, friendship, humor, and readiness to both think deeply and have a great time.

Finally, I am indebted to my family and to the many friends who have been unstinting in their faith and in offering nourishment, perspective, and diversion at all the right times and in more ways than I can name.

Table of Contents

List of Tables

List of Figures

1.0 INTRODUCTION

1.1 Control of Measurement Processes

Every measurement of a ("fundamental") process is itself a process and is therefore characterized by what Shewhart[1] first characterized as "common cause" and "assignable cause" (Deming[2] preferred the term "special cause") variability, just as is the fundamental process. Despite both the increasingly widespread use of statistical process control (SPC) in manufacturing environments and the exhortations of Deming[2] and others[3] that "statistical control of the process of measurement is vital; otherwise there is no meaningful measurement,"[2] it is not unusual for measurements to be simply assumed to be reliable. The temptation to trust measurements implicitly rather than address the possible complexities of the measurement process may also be exacerbated by measurements often being considered non-value-added parts of the total process. Nevertheless, as long as measurements are used to control and evaluate the fundamental (and presumably value-adding) process, measurements must be considered an integral part of the fundamental process and controlled accordingly, with the "implicit requirement that the variation in the measurement system be small when compared to the process variation of the characteristic to be measured."[3]

A good understanding of the characteristics of the measurement process is crucial for the health of the fundamental process. While an uncharacterized measurement process is not necessarily out of control, there can be no confidence in those measurements. A measurement process that is out of control can have a devastating impact on the process it is being used to control since the assumption of reliable data is now violated. A measurement process that is in control but with a variability that is relatively large compared to the variation of the fundamental process can also be problematic since measurements with relatively large variability have the potential to overwhelm the process variability budget for the fundamental process and to cause Type I and Type II errors in dispositioning material, rejecting good material or accepting failing material. Measurement process "specifications" are thus generally related to the specifications of the fundamental process the measurements are being used to control. In turn, specification limits for the fundamental process, especially if the process window drops off very sharply, will need to take into account the measurement variability if the quality of the product is to be protected.

The advantages of SPC for fundamental process control apply equally to the control of measurement processes. Specifically, such a methodology supports better process stability by invoking "rules" that advise against process correction in response to the inherent variability of the process (the result of common causes) but support process correction given special causes of variation that represent an actual change in the process. As Deming describes with the example of marbles falling through a funnel[2], process adjustment because of variation due to common causes results in increased variability compared to letting the process continue to run (intuitively, the process will "even itself out" over time with respect to common causes that are characteristic of the process). If real shifts, changes resulting from special causes, are detected, however, it is advantageous to observe the shift as quickly as possible to minimize elapsed time to correct the process. Statistical process control is also useful in addressing measurement processes because its separation of common causes from special causes allows one to address an in-control but large-variability process by changing the measurement process – e.g., through changes in equipment, maintenance, or operator training – and to address special causes by looking at irregularities of particular machines or operators.

As with control of fundamental processes, more than one chart or set of data may be required to provide assurance that the measurement process is in control. Different sources of variability may require quite different approaches, some more consistent with (or a subset of) the data collected in the normal course of production and some not. However, if a major source of variability is not accounted for in the data collected, the measurement process can neither be properly known to be in control nor the potential impact on the fundamental process accurately assessed.

1.2 The Internship

This thesis addresses many of the measurement control and variability issues described briefly above for the particular case of scanning electron microscopy (SEM) measurements of critical dimensions in a semiconductor fabrication facility (fab). The internship work was performed at Intel Corporation in their Fab 9 (Rio Rancho, NM) facility. The work was performed from within the lithography process engineering group and was focused on the measurement variability issues surrounding a particular layer in a particular process (the contact layer for a relatively new process). The motivation for the investigation of this particular layer was a combination of the leverage that could be achieved by improving the capability of the contact layer as well as a general suspicion at the start of the internship that there might be SEM measurement variability confounding process difficulties.

1.3 Thesis Structure

Chapter 2, Internship Background, will set the stage for the project with a brief description of Intel Fab 9; the fabrication process in very general terms; the rationale for choosing the focus of the project (measurements of critical dimensions for the contact layer); the relevant measurement "organization"; and basic SEM operation. Chapter 3 explores the potential costs of high measurement variability, or measurement error, in the fab context, including Type I and Type II errors. Chapter 4 addresses potential sources of SEM measurement variability and estimates their magnitudes in order to assess the relative importance of various types of measurement error. Chapter 5 looks at the specific ways in which the most common types of mismeasurement occur while Chapter 6 discusses ways that were used, as well as other ways that could be used, to reduce the occurrence of these measurement errors. Chapter 7 provides detail on the anonymous, automated feedback system designed and piloted during the course of the internship. Finally, Chapter 8 provides a summary of conclusions and recommendations.

2.0 INTERNSHIP BACKGROUND

2.1 Intel Fab 9

The internship was carried out at Intel Corporation's Fab 9 facility in Rio Rancho, New Mexico. Intel Corporation is a leading semiconductor device manufacturer of microprocessors and also of flash memory (total 1997 revenues of $25.1 B). Approximately seven years old, Fab 9 at the time of the internship processed 6-inch silicon wafers, primarily into flash memory. Intel's Rio Rancho campus also includes Fab 7, also processing 6-inch wafers primarily into flash memory, and Fab 11, processing 8-inch wafers into microprocessors. Fab 7 and Fab 9 run many of the same processes to produce many of the same products and thus were part of a "virtual factory," wherein the fabs strive to match each other in terms of equipment, process, and yield. While Fab 11 is at times a useful resource for Fab 9, the interaction is rather limited since the fabs run different processes, with significant equipment differences, to make very different products. Fab 7 and Fab 9 may also experience a higher level of cost pressure and less market security than does Fab 11 because of the commodity nature of the flash memory business.

In general, Fab 9, similar to other Intel fabs, is organized with a number of functions that serve the whole fab, some that serve a whole shift (four shifts total: "front end" of the week days and nights and "back end" days and nights), some that serve an entire cluster (e.g., lithography), and some that serve a particular cluster during a particular shift. Upper management and functions such as safety, training (at the more administrative levels), automation, etc., provide services to the entire fab. Shift supervisors, one per shift, have responsibility for the manufacturing floor for that shift. Day to day floor activity is controlled most directly by the supervisors for each shift that have responsibility for a particular cluster. Manufacturing technicians (MTs) report to and are hired by the (cluster) supervisor. Process engineering is also divided by cluster but with engineers supporting all shifts (working weekday days).

As mentioned previously, the work for this thesis was conducted from within the lithography (litho) process engineering organization and involved substantial work with the litho manufacturing organization (all shifts) as well as with the automation and training organizations.

2.2 Process Overview

The processes used can be described very briefly and generally as follows. A product and a process to make the product are designed, developed, and transferred from the technology development organization to the fab(s) for production ramp-up. The process starts with a bare silicon wafer which is modified by numerous steps to build the desired structure. Often, these processing steps are grouped into "layers" where a layer typically involves the deposition of a thin layer of organic "resist" on the wafer; patterning of that resist (using masks through which the resist is selectively exposed); removal of excess resist (to leave the desired resist pattern on the wafer surface); execution of the process step desired (selectively, since the resist protects the rest of the wafer); removal of the resist; and wafer cleaning. The process step used may be, for example, an etch step, to remove material; an ion implantation or diffusion step, to change the electronic properties of the exposed material; or a deposition step, to add material to the exposed area. The lithography cluster is responsible for creating the appropriate resist pattern on the wafer. Wafers then go to another cluster to receive the corresponding process step and subsequent clean before returning to litho for confirmation of the structure just created before beginning on the next layer.

Litho Metrology

In addition to the actual process steps, a number of metrology (or inspection/measurement) steps may also be included. In the lithography area, metrology includes measurements after patterning the resist to ensure that the pattern is generally "clean" (no macroscopic problems), that resist pattern dimensions are correct, and that the alignment with previous layers is correct. It also includes measurements after processing and cleaning to ensure that the actual dimensions are correct (e.g., that holes that have been etched in existing material for contacts are the correct size).

The macroscopic check is manual while the check for alignment to previous layers is done using an automated registration tool and the check for dimensions is typically performed using a SEM. The measurements after the resist is patterned ("develop check") are to ensure that bad material is not created because of a bad resist pattern since at that point the wafer can be "reworked" (the resist stripped, reapplied, and repatterned) without significantly affecting the wafer. The check of dimensions after the actual process step has been performed ("final check") is to ensure that the material (or more specifically, certain critical dimensions on the wafer) is within specification; it is, however, too late to rework if it is not correct. Measurements may not be performed on all lots at all layers if the process capability is high enough to justify either not measuring at all or measuring only a fraction of the lots.

Production Data

Although metrology steps are often not considered to add value to the wafer, the metrology data is integral to fab operations. Process control charts are generated using the measurement data for a particular step. Critical dimension SEM (CDSEM) data for a particular layer in the process, for example, is recorded and plotted, with previous data and with specification limits, at the completion of data collection. If the data indicate that the lot is within specifications, the lot continues to the next step; if not, the lot is stopped and proceeds, if at all, with a modified sequence of steps, generally as determined by engineering (the lot might, for example, be sent to rework, or it might continue processing after engineering consideration of risk and appropriate documentation).

Process engineers also use metrology data to identify process trends or shifts as well as to evaluate opportunities for process changes (that might make the process more robust, for example). In addition, the metrology data are used at a higher level in surveying fab performance to assess the capability of a layer or process. Process capabilities are formally reviewed periodically to try to bring the entire process up to a minimum desired capability, with layers that are below the desired capability targeted for improvement efforts. These data may also be used to determine the necessity of actually making certain measurements – e.g., if the process has a history of very high capability, it may be a candidate for reduced measurement frequency.

2.3 Choice of CDSEM Measurement of Contacts for Internship Study

Although Fab 9 was running several processes at the start of the internship, it was clear that the project should focus on one particular process (call it, arbitrarily, P1). P1 was a submicron process for producing flash memory that was just coming up at Fab 9 but was ramping quickly and expected to comprise the bulk of processing in a short time. Two other processes were also running at the time but were ramping down. Working with process P1 therefore appeared to offer the greatest leverage for the project.

Further, it was evident from looking at historical data for a similar Fab 9 process, one of the ones that was ramping down, as well as preliminary data for P1, that the contact layer was a source of particular process difficulty. The contact layer etches a hole that when filled with metal will provide contact to the source and drain regions of the device. The contact layer, for example, exhibited the lowest process capability (Cpk) and was the largest contributor to scrap for the period studied.

While develop check CDSEM measurements of contacts are part of the process and are used to check that the resist pattern has been put down properly, they do not provide a highly reliable indicator of final contact dimension since the etch process significantly impacts the actual contact dimension and since SEM image quality at develop check is limited by the insulating resist layer. Since the final check measurements reflect the actual contact (after etching, rather than just the resist pattern for it), and because measurements are much easier to make when the wafers are not coated with resist, the final check measurements provide a much more meaningful basis for controlling the process as well as for dispositioning material.

Although contact final check CD measurements are more reliable (have much better image quality) than develop check measurements, and although many of the process issues leading to low Cpk and/or high scrap rates were related to the etch process itself, there was also some suspicion that the contact final check CD measurements themselves might be adding a significant amount of variability and thus hampering efforts to diagnose problems and improve the process. It was, however, impossible to accept or reject such a hypothesis based on the production data available since production wafers vary themselves, particularly when adjusting the process to find better operating parameters. This, combined with the fact that particular measurements are not linked to particular operators, made it impossible to determine whether observed variability was due to the process or to the measurements or to both.

Since there was already an established group addressing the process issues related to the contact layer, the internship project focused first on investigating the measurement variability (if any) exhibited on contact layer final check CD measurements. The project continued, after there was indeed shown to be measurement variability, by looking at sources of variability and their relative magnitudes, and finally by implementing changes and suggesting ways to reduce that variability.

It should be noted, also, that measurement "quality" in a general sense can be addressed both in terms of accuracy (how well the measurement represents "reality," as determined by other techniques) and in terms of precision (essentially repeatability, the degree to which the same measurement value is recorded for the same structure). This thesis is focused exclusively on precision issues since problems with accuracy do not impact the efficacy of process control as long as measurements are consistent, and do not impact the quality of disposition decisions as long as specification limits are set using values that are consistent with the measurements themselves.

2.4 Fab 9 CDSEM Measurements

This section briefly presents relevant details on the organizational structure, general training approach, basic SEM operations, and SEM equipment qualification in Fab 9.

Organization

As previously noted, CDSEM measurements fall within the litho area. A subset of the litho MTs on each shift are trained and certified to do CDSEM measurements. In general, because of ergonomic considerations, CDSEM measurements are done on a rotation basis, with any particular technician working only a certain number of hours per day at SEM or rotating on a weekly basis in and out of SEM. To accommodate this rotation, each shift has on the order of 15 operators per shift certified for SEM operation, giving a total fab population of SEM operators of on the order of 60 MTs.

Some of the SEM operators are also certified as SEM trainers, meaning that they can train and certify other MTs. There are on the order of six SEM trainers per shift with trainers primarily running production, with an occasional training "stint." There are no formal meetings for SEM trainers within or across shifts.

Generally on each shift one of the trainers is designated the "peer trainer" although this does not appear to be a formal position. On some shifts, the peer trainer was essentially the shift "master trainer" and was the generally acknowledged expert. On other shifts, however, there did not appear to be a clearly acknowledged "peer trainer."

In addition there are cluster training coordinators (CTCs), one per cluster per shift, who are also MTs. While the CTCs spend a significant fraction of their time running production, they also have formal training responsibilities. Litho CTCs from all shifts meet together periodically to discuss training issues, including changes to the baseline training package (brief description below) which must be approved in writing by each of the CTCs. They also carry out specification change notifications, making sure that each MT understands and acknowledges any changes. However, CTCs are not necessarily certified trainers for all parts of the cluster. Thus, while three of the four litho CTCs were actually certified SEM trainers, one was not even certified as a SEM operator. In that instance SEM CTC-type issues were handed off by the CTC to an informal alternate CTC who was a SEM operator and trainer.

Finally, there are also shift training coordinators (STCs), one per shift, who are not MTs and who operate at a much more general level. In practice they appear to have little day to day interaction with the floor, at least in the case of SEM training.

For most shifts there is also an informal SEM "coordinator" designated by the litho shift supervisor. The coordinators are MTs and are often, but not necessarily, trainers; they usually have a lot of experience with SEM and thus are relative experts; and they spend a lot of time at SEM on a regular basis. While the coordinators don't necessarily have training responsibilities, they do tend to have a significant amount of informal influence on the way SEM measurements are performed on a daily basis.

Training Materials

The baseline training package for SEM (same one for all shifts) is generally "owned" by an MT. It consists of both written training material and a checklist of items (generally to be demonstrated at the SEM) for each of which the trainee must sign-off themselves and be signed-off by a trainer. Over time, there appears to have been engineering input to the document and the checklist but this is quite informal as the official sign-off requires only the signatures of the baseline package owner and the CTCs for each shift.

The picture book (one for the fab, located at the SEM) is the definitive reference for details of SEM measurements for particular processes and layers. It is currently owned by an "engineering technician," someone originally from the manufacturing floor but now reporting to engineering. Any changes to the picture book are made by, or through, the owner. Changes must be coordinated with document control but no other sign-offs are required. Thus, while informal consultation with engineering layer owners and with the CTCs as to major changes was welcomed, no formal sign-off by either group is actually required.

SEM Operation

When the operators are at SEM they typically run measurements on many different layers (on the order of a dozen), at both develop check and final check, for two or sometimes more processes and for numerous products. Generally three or four operators at a time are working on SEMs placed adjacent to each other. For each lot, the automation system loads the appropriate SEM "recipes" (specifying, e.g., current, voltage, field control method (FCM), measurement magnification, etc.) and takes the operator to the designated measurement sites on the wafer. Proper SEM set-up, however, which can significantly impact image quality and therefore critical dimension values, is largely manual. After SEM conditions have been set, but before the program actually goes to a wafer measurement site, the recipe takes the operator to a metal grid which is used to adjust the wobble, astigmatism, and focus for those particular SEM conditions, essentially manually aligning the beam.

At each measurement site, the operator adjusts (if necessary) the stigmators and focus to get the best image and then places the measurement gates in an appropriate location. Using algorithms that are set by the recipe, the SEM moves the gates in a predetermined direction looking for what the algorithms define as "the" (correct measurement) edge. The SEM moves the gates to that location and also displays the measurement value on the CRT, at which point the operator must accept or reject the gate placement (and therefore the CD value). If the operator accepts the placement, the measurement value is collected by the automation system and stored; if not, the operator re-places the gates and repeats the process until an acceptable measurement is made before moving to the next site. After the requisite sites are measured the wafer is returned to the cassette and control charts are displayed showing the mean and standard deviation of the just-completed set of measurements, plotted with previous measurements of the same layer against control limits.

SEM Measurement Process and Equipment Capability

There is a well established protocol for demonstrating that a newly installed SEM (or other piece of equipment) has a measurement performance that is consistent with the other SEMs in order to be qualified to run production material. Assessments of measurement processes often use the Precision/Tolerance (P/T) metric, where P is defined as six times the sigma of the measurements (repeated measurements of a single site) and T is defined as the minimum of six times the sigma for the total process or simply the upper specification limit less the lower specification limit. While this metric does provide an effective evaluation of the quality of a measurement process relative to the fundamental process it is measuring, such an assessment is not performed on a regular basis and would also likely not (unless specifically designed to do so) include the effects of operator variability.

The SEMs are somewhat unique in the fab in the sense that operator performance can still have a significant effect on actual data. In much of the rest of the fab, in contrast, standard operations require that MTs handle material correctly but generally data is acquired by the equipment, without MT intervention. Operator variability is therefore an insignificant issue in much of the fab operation – one notable exception to this being the repair or maintenance of equipment, where a large manual or MT-influenced component (and thus opportunity for variability) is still present. Control requirements are thus somewhat different for SEMs. As an example, automated tools to measure registration (wafers placed in the machine, operator starts sequence, machine takes all the data) are checked with monitor wafers for tool-induced shift on a regular basis; these measurements determine whether the tool is production-ready or whether it needs adjustment. On the SEMs, pitch monitors are used on a regular basis to confirm that SEMs are well adjusted and production-ready with respect to magnification calibration. However, there is no monitor that includes operator variability, which, because of the basic equipment operation, can still be significant.

2.5 SEM Image Generation

Scanning electron microscopes (SEMs) are now standard semiconductor fab metrology tools for measuring critical dimensions. (Postek and Joy[4] provide a good general reference and description of the use of SEMs for measuring submicron critical dimensions as of the late 1980s. Reimer[5] provides another useful general reference for low-voltage SEM.) SEMs started being used in fabs when shrinking device dimensions made the resolution of optical microscopes (limited by diffraction effects to roughly the wavelength of the light, or about 0.5 microns) insufficient. Because SEMs use a beam of electrons (with a wavelength of on the order of 10^{-5} microns) to image the sample, feature resolution is no longer limited by diffraction effects. Although other effects do become significant, SEM resolution appears to be sufficient to meet semiconductor industry needs in the near term.

An image is generated by scanning the sample with a focused electron beam and using the electrons emitted from the sample at each "location" to construct an image. Higher magnifications are achieved by scanning smaller areas but still producing images with the same number of pixels. Most SEMs use detected secondary electrons (generated when an electron from the beam knocks an outer orbital electron out of the sample material which then escapes, if it is generated close enough to the surface, and is subsequently detected) to form the image. Using "secondaries" to create the image thus results in an image dominated by surface features since secondaries generated too deep will not escape the sample. (Backscattered electrons, electrons from the beam that go into the sample, bounce around, and bounce back out, are also present but are easily distinguished from secondaries because of their much higher energy and are generally not used for image generation.)

Because the image created by a SEM is a collection of secondary electron emission intensities, rather than a direct image of the sample as an optical microscope or a transmission electron microscope would reveal, the SEM image of a hole, e.g., for a contact, may not look exactly like the hole. More efficient generation and detection of secondaries results in brighter areas of the image with, e.g., more sharply angled walls generating secondaries more efficiently than horizontal surfaces, but also with secondaries generated near the bottom of a deep hole being less likely to escape and reach the detector. Thus, the particular geometries of the contact hole result in an image like that shown in the schematic in Figure 1. By placing the measurement gates at the inside of the grey ring (correct placement indicated by dotted lines) the dimension of the bottom of the contact will be measured.

The beam is generated by extracting electrons from a tip and applying an accelerating voltage. This accelerating voltage is an important parameter in SEM recipe formulation. Although higher accelerating voltage generally improves resolution, the low voltages used for semiconductor metrology (0.5 to 5 keV) are advantageous to minimize both charging and damage to the wafer.

Sample charging occurs when either more electrons are absorbed than emitted (negative charging) or more electrons are emitted than absorbed (positive charging). High accelerating voltages may thus result in negative charging if the electrons can not escape. Certain accelerating voltages can also, however, result in positive charging (very efficient removal of electrons). Specific accelerating voltages at which equilibrium is maintained depend on sample material and surface topography. Charging, of either type, is a problem because it reduces the quality of the image and can deflect the beam, resulting in a distorted image or, in extreme cases, causing the image to wander on the screen.

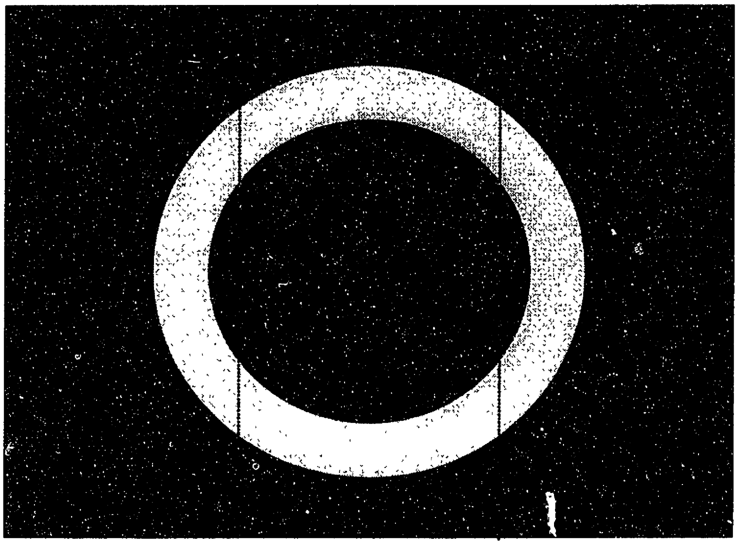


Figure 1 Schematic of SEM image of final check contact. The vertical dotted lines placed at the inside edge of the inner "grey" ring indicate correct gate placement for measuring the dimension of the bottom of the contact (the critical dimension). Dimensions not to scale.

Another important parameter in a SEM recipe is the emission current. In general, higher beam current will produce a better image because it will generate a stronger signal. However, higher currents will accelerate charging effects, which could be a problem under some conditions and with some samples. Typical beam currents for CD measurements are between 1 and 10 μA.

Many SEMs also now have a recipe option particularly for imaging of holes in insulating materials, such as contacts. By using an electrode to keep a slight positive charge on the surface of the wafer, secondary electrons emitted near the bottom of the hole are more likely to be collected and contribute to the image. This is the so-called field control method (FCM).

3.0 COSTS of MEASUREMENT VARIABILITY and MISMEASUREMENT

Large measurement variability has the potential to significantly increase the costs of a production system, primarily through material dispositioning errors and poor process information. A brief discussion of the contact layer and the effects of incorrect CDs will provide a context for the discussion of costs of measurement errors.

Contacts

The contact layer essentially etches a hole through a stack of insulators to reach silicon at the bottom of the hole. When the hole is filled with metal, a contact is created between the source or drain region below and other layers to be created on top. In the case of the contact layer it is important to know the dimensions of the hole, and more particularly, the dimensions of the bottom of the hole, the critical dimension. (The bottom of the hole and the top of the hole are somewhat different in size because the hole walls are not absolutely vertical.) Figure 2 shows a schematic of the contact layer (after etching, before metal deposition); dimensions and wall angle are not to scale. If the contact (hole) dimensions are outside the process window (outside defined specifications) then device performance is compromised. If the contact is too large, the distance between contact and gate is reduced and can lead to leakage or, in extreme cases, to shorting. If the contact is too small, on the other hand, then contact resistance is increased.

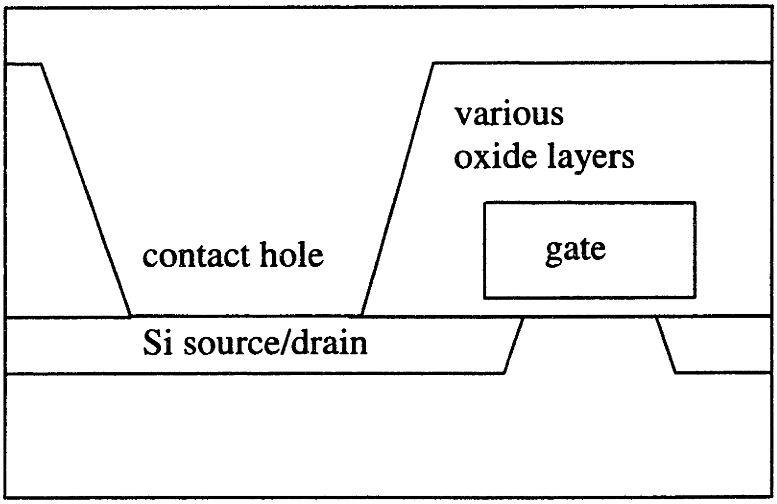


Figure 2 Schematic of a contact after etching (before metal deposition) in cross-section. Dimensions not to scale.

Measurement Variability and Mismeasurement

Just as fundamental processes exhibit a natural variability, so do measurement processes. In each case it is important to refrain from continually adjusting the process based on normal variation but at the same time to be sensitive to true process shifts. A measurement variability that is large relative to what is being measured can result in both Type I and Type II errors relating to dispositioning material off the fundamental process. It can also result in incorrect modifications to the fundamental process in so far as the measurement data is used to control the process.

The current situation is characterized by two rather different and distinct sources of the variability observed on measurements of final check contact layers. The first source of variability is that described by common cause variability, namely the variability associated with correct but not identical measurements done repeatedly and by different operators on different SEMs at different times. The second, and much larger, source of variability results from incorrect measurements (primarily measurements of the wrong dimension). The causes of these mismeasurements will be discussed in more detail in Chapter 5. For now, it is relevant to note that because of the particular geometry of the contact layer and the most common measurement issues, mismeasurements tend to be characterized by values that are too high rather than too low.

3.1 Type I and Type II Errors

Type I Errors

Type I errors, or false alarms, are false failures of good lots. The usual procedure is that when a contact lot fails (or appears to fail), the lot is held and remeasured. In general, if the second measurement appears to be in control, then the lot passes and is sent on; if not, the lot fails and is held for dispositioning. While this procedure is quite effective for reducing the number of "permanent" Type I errors, that is, Type I errors that are acted upon, there are remeasurement as well as integrity costs.

The remeasurement costs stem from the labor associated with actually measuring the lot twice and from delaying the progress of the lot through the line. The integrity costs come with the bias toward acceptance introduced by testing until a part passes rather than making disposition judgments based on all of the data. For example, if a bad lot has a 90% chance of being measured correctly as failing, but each failure is followed by one remeasurement, the probability of the part passing (in error) is actually 0.1 + 0.9(0.1) or 19% as compared to 10%. More remeasurements clearly result in even greater bias.

There is a further, informal bias that is exaggerated by a system that essentially assumes that a failing data point is a bad measurement. The bias is a result of operators knowing the expected value of the critical dimension and thus perhaps remeasuring or otherwise adjusting the measurement, consciously or unconsciously, if a critical dimension appears to be extreme or failing. The ability operators have to "dial-a-CD," because of the manual nature of the measurements, means that the data are particularly "open to suggestion," especially if operators have more confidence in the expected value than in their own measurements. It might therefore be useful to remove the display of dimensions on the CRT so that operators are less likely to be intentionally or unintentionally biased by results as they are accumulated. Thus, while operators would see the final result (the mean and standard deviation of the sites just measured) plotted on the relevant control chart, they would not see individual values as they were taking the measurements. At the same time, operators need to have confidence in the quality of their measurements.

The costs described above are essentially the costs of a "remedied" Type I error. However, Type I errors are not necessarily always remedied, in which case other sorts of costs also enter into the picture. A good lot, for example, could be measured as failing because of an incorrect measurement by an operator and that same operator could remeasure the lot and make the same error again, causing the lot to fail. Costs of this unremedied Type I error then include the costs of chasing down what is (erroneously) believed to be a process problem (the search may be limited to litho but may also extend to etch, may involve bringing tools down, etc.); the costs of processing a lot to that point in the line and then scrapping it (same cost as if it had really failed); and the opportunity costs of not processing the lot to end-of-line and generating revenues on the sale of good product.

Type II Errors

Type II errors are false acceptances. These errors are more insidious because, by definition, they will not be caught in-line and are thus only discovered, after a possibly significant delay, when the material reaches electrical test at end-of-line. Again, because of the geometry issues, the most likely event is that a too-small contact will be measured as passing. A variety of costs are associated with this type of error. Direct costs will be incurred in processing a lot to end-of-line that will then fail because of the previously-undetected problem. There are also opportunity costs since the bad lot is taking the processing slot of what could be a good, revenue-generating lot. Perhaps most importantly, because of the mismeasurement, the actual process problem (that, for example, results in contacts that are too small) may not be caught until that lot reaches the end-of-line, or until another lot actually fails at SEM, leading to still greater losses as more lots are affected before the problem is corrected.

Magnification of Errors

The costs of Type I and Type II errors can be easily magnified if they occur on particularly important sets of measurements. This might occur, for example, in the situation where changes to the process are being investigated, with process decisions being made based on critical dimension measurements of a relatively small number of lots, possibly all measured by the same person because the lots are being expedited. If data are thus compromised by consistently incorrect measurements, the process will be incorrectly set, leading not only to material that at the very least is sub-optimal and which may be out of specification, but also to more time being required in the future to re-engineer the process another time once the error of the supposedly corrected process is observed and confirmed.

3.2 Estimating the Costs of Type I and Type II Errors

Type I Errors

As described in the previous section, the costs of Type I measurement errors are: the direct costs of remeasuring the lot; potential yield costs due to increased handling; the costs (if any) of the delay experienced by that lot as well as by others in line behind it; and the costs of "promoting" to some degree a "dial-a-CD" mentality. Although it is quite difficult to precisely estimate these costs, it may be helpful to bracket the minimum and maximum likely costs of each type.

The direct costs of remeasuring a lot consist of the cost of operator time and the cost of additional equipment use (general wear and tear, expendables if applicable, etc.). In terms of operator time, the costs of remeasurement are zero as long as the remeasurement can be accommodated by the standard set of operators (that is, additional operators, or operator hours in the case of overtime, are not hired to do the remeasurement). The maximum reasonable cost for remeasurement would be operator time at an overtime rate for a feasible minimum time, perhaps an hour. In the case of SEM use, incremental equipment cost for remeasurements, as long as the volume of remeasurements does not result in the need for another SEM in order to meet production requirements, appears to be negligible.

The yield costs of increased handling, for a single remeasurement (or modest number) appear to be insignificant compared to the total handling. These costs could be ascertained through yield studies of wafers whose handling through the line was carefully monitored.

When a lot must be remeasured, it is certainly delayed at that inspection step. The delay at that inspection step may or may not result in a delay in that lot, or other lots behind it, reaching the customer. If no delay is introduced, because the SEM delay merely substitutes for other delays further downstream or because the product is being built to stock, then costs of the delay would be zero. If, however, SEM were a production constraint and all available product were being shipped to customers, then one way to assess the cost of remeasuring a lot would be through a Theory of Constraints approach[6,7]. In this model, if SEM were the constraint, the rate of wafers being "processed" through SEM would determine the rate of wafer completion, and thus revenue generation, for the fab over a given period of time. In this case, each measurement that must be repeated reduces throughput and the cost of each repeated measurement could be measured by the reduction of revenues in the period of interest by one lot divided by the total number of times a lot is processed at SEM in the standard flow.

Although it does not appear to be appropriate to the current situation to consider SEM a constraint either in terms of equipment or people, suggesting that the cost of the delays is negligible, it is also true that the desire on a larger scale to "size" the fab optimally in terms of MT staffing levels could lead to MT time essentially being a constraint. That is, without sufficient "slop" in operator time to accommodate a remeasurement, that remeasurement can in fact delay outs and reduce revenue, assuming all product made is shipped.

The dial-a-CD mentality can have significant, if difficult to quantify, costs. In this case, Type II errors are essentially being created if operators don't believe the "bad" measurements they're getting and adjust the measurements to something they believe is more "correct," with associated costs similar to other Type II errors as described below.

Type II Errors

As described in section 3.1, the costs of Type II measurement errors are: the direct and opportunity costs of processing to end-of-line a lot that is not actually good; and the costs of duplicating the problem on additional lots because the real problem is not detected.

The direct costs of processing a lot to end-of-line that will then fail there are comprised of the cost of operator time; the cost of additional equipment use; and materials costs. In terms of equipment and materials, incremental costs could be ascertained or could be estimated very roughly as a percentage of the total process. Again, there is the caveat associated with the equipment that these estimates assume that the number of "error lots" is low enough that additional equipment is not required in order to meet commitments (e.g., that equipment is not a constraint). Labor costs also follow the arguments above: in so far as operators have time to absorb error lots there are no real labor costs. However, as soon as operator time becomes the constraint, operators processing an "error lot" rather than a good lot are directly reducing revenue while incurring the same costs. Thus, labor costs are zero at a minimum but at a maximum translate to the lost revenues corresponding to the number of error lots multiplied by the fraction of the process remaining (in the current case, a rough estimate might place the contact layer at approximately 2/3 of the way through the process). This estimate of costs is not of the total costs of a bad lot, but only those costs associated with scrapping the lot at end-of-line rather than at the appropriate (first available) spot.

Although in relatively small numbers the costs above may not be prohibitive, Type II errors have the potential to impact more than just the lot on which the error occurs by virtue of the fact that in the event there is a real process problem represented by what should have been an out of control point, the problem goes undetected for some period of time. At the very least, one more lot is processed improperly before the problem is properly detected by the next lot measured at SEM (assuming that the next lot reaches SEM much sooner than the initial lot reaches end-of-line). It is also quite plausible, however, that a number of lots might be processed improperly in the time between when the first lot should have been caught (but was not) and when the next lot is measured at SEM and caught. If the second and/or subsequent lot(s) is also mismeasured the problem clearly grows very quickly.

3.3 Cost of Poor Measurements Not Resulting in Type I or Type II Errors

It is also possible to have, for certain ranges of actual critical dimensions, highly variable or incorrect measurements that do not actually result in a Type I or Type II error. Although there is no dispositioning error in this case, there are still associated costs to the system, which could be significant. These costs relate to reduced sensitivity to shifts in the fundamental process, artificially tight process constraints, and reduced quality resulting from a skewed distribution.

Mismeasurements or highly variable measurements can reduce the efficacy of statistical process control of the fundamental process by pushing the control limits out, thereby reducing the sensitivity of the charts, and therefore the system, to actual process shifts. The impact will depend, of course, on the relative magnitude of the variation of the process relative to the variation of the measurements. At best, there are no process shifts and therefore no costs associated with a scenario such as this one. At (reasonable) worst, a significant process shift is not promptly detected and is very large, with many lots affected, when finally detected.

With respect to specification limits, measurement variability cannibalizes the process variability budget so if measurement variability is high relative to the process limits, the process may be artificially constrained to a narrower window, leading to higher costs, greater expenditure of engineering resources, etc.

In the current case, because large errors in measurement tend to be high values, there may a consistent offset to the measurements. These data would suggest to the layer owner a need to adjust the process, e.g., to recenter it, when in fact it may be perfectly centered. An incorrect centering may lead to more failing lots since the process, whatever its capability, Cp, is now skewed in the process window (e.g., center is low compared to desired center). In addition, there may also be a yield cost because the process, although within the window, no longer has its distribution centered in the window[8].

4.0 SOURCES and MAGNITUDES of CDSEM MEASUREMENT VARIABILITY

Given the potentially significant costs of large measurement variability and mismeasurement, a key part of the internship was to determine whether there was significant variability, and if so what the sources and magnitudes of variability were. To help frame the search for sources of variability, a simple fishbone, or Ishikawa[9], diagram was created as a way to categorize possible sources of measurement variability under the traditional headings of Materials, Environment, Equipment, and People. Variability information available from standard production data was used when available to estimate magnitudes. In some cases, variability data had to be generated from special experiments. This information was then used to compare sources of variability by magnitude of possible deviation from baseline values or by standard deviation, as measured in arbitrary units (arbitrary units held constant throughout this thesis).

4.1 Sources of Measurement Variability

Figure 3 is a simple fishbone diagram showing possible sources of CDSEM measurement variability.

Starting with Materials, there is certainly variation of the actual CDs to be measured: within-field, within-wafer, within-lot, and lot-to-lot variability. The measurement variability being addressed, however, does not expect all measurements to be the same – rather it expects repeated measurements of the same material to be the same. The actual process variation is thus not of particular interest for this work. (The only way in which the process variability may interact with measurement variability is that smaller contacts appear to be more difficult to measure than larger ones, where more difficult means that the image quality is not as good and therefore repeatable measurements are more difficult to achieve.)

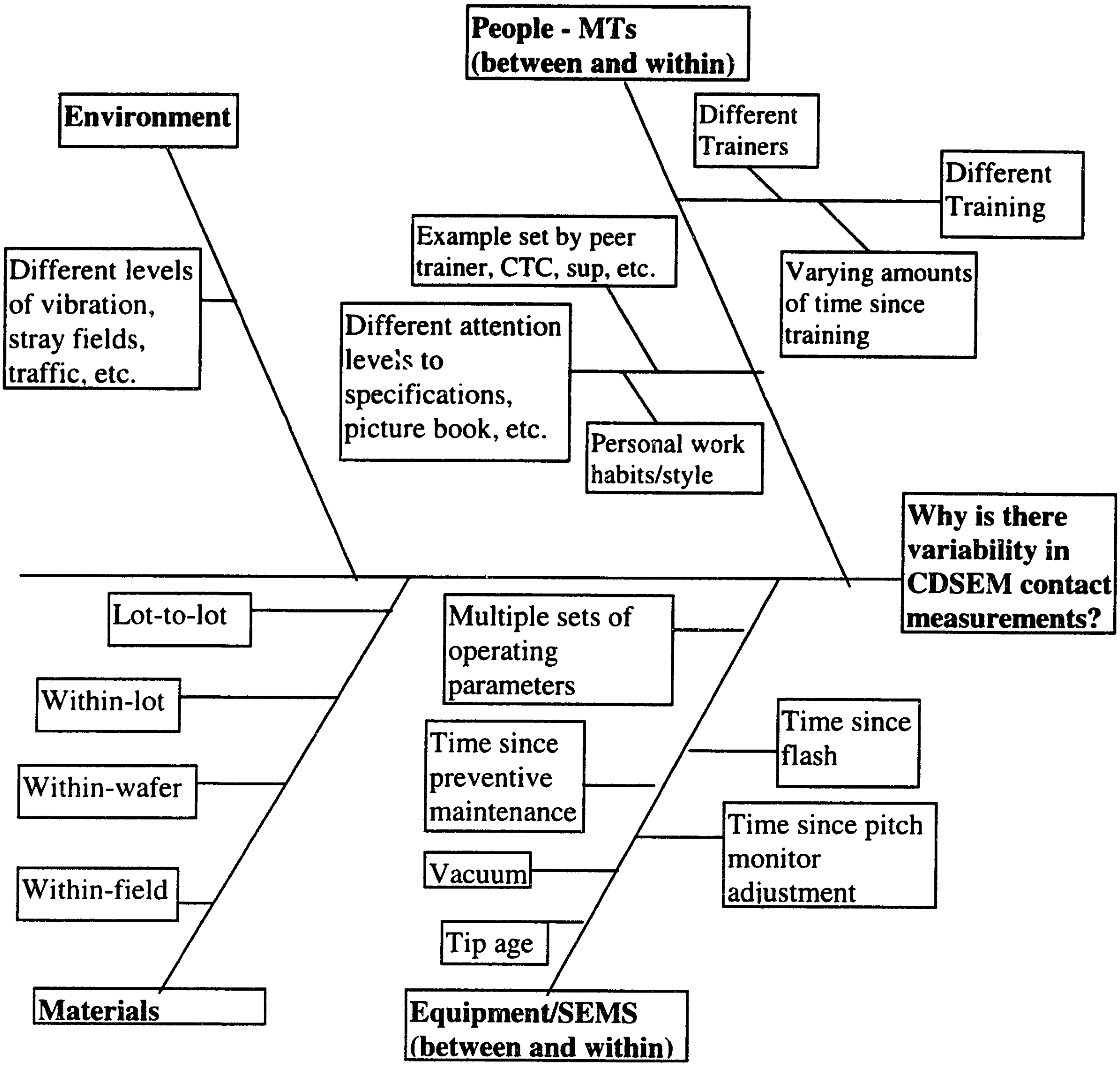


Figure 3 **Fishbone diagram showing possible causes of CDSEM measurement variability.**

Environment also does not seem likely to be a major contributor to measurement variability since fab conditions are well controlled, SEMs are stationary, and electromagnetic shielding on this equipment is believed to be quite effective. One possible source of variability could be between day weekday shifts, when there is generally more engineering traffic, and night and weekend shifts. These effects could be confounded, however, with variability inherent to particular shifts, also associated with day or night, through the many possible people issues that will be discussed. Cumulative probability plots of production SEM measurements on the final check contact and other layers, however, do not show a systematic difference between day and night shifts (where weekend days also counted as days).

With respect to Equipment, there are many possible differences, either between SEMs or on a single SEM over a period of time. These include differences such as: elapsed time since last preventive maintenance, elapsed time since "flashing" (a procedure used periodically to clean the tip; after flashing the tip typically exhibits less stability for a short time, requiring more frequent adjustment for optimal image quality), tip age, vacuum quality, etc. This work did not study these factors but only addresses them as a whole by looking at cumulative probability plots (Figure 4, for example) over different time periods and observing the magnitude of the differences between SEMs or between different time periods for "average" lots, those at the center of the distribution (0.5 cumulative percentage). For the contact layer at final check, the largest difference between SEMs or over different time periods for a single SEM was approximately one arbitrary unit (and on the order of one quarter the magnitude of typical operator mismeasurements). Thus, the internship work continued to focus on the people factors contributing to observed variability.

There are a large number of ways in which People differences result in significant measurement variability. The most common of these are displayed on the fishbone chart and discussed further below. At the broadest level, operators are likely to measure differently because they received different training or because they devote differing amounts of attention to making sure their measurements are correct.

Operators may receive different training because, while they do use the same package of written materials, they are trained and certified by one of a relatively large number of trainers. Differences between trainers are then passed on to operators, who each add their own variability. In addition, different operators will have been trained differing lengths of time ago. While some operators may get better and more consistent with time, personal variations may become magnified over time as the exact content of the training becomes more distant in time and memory.

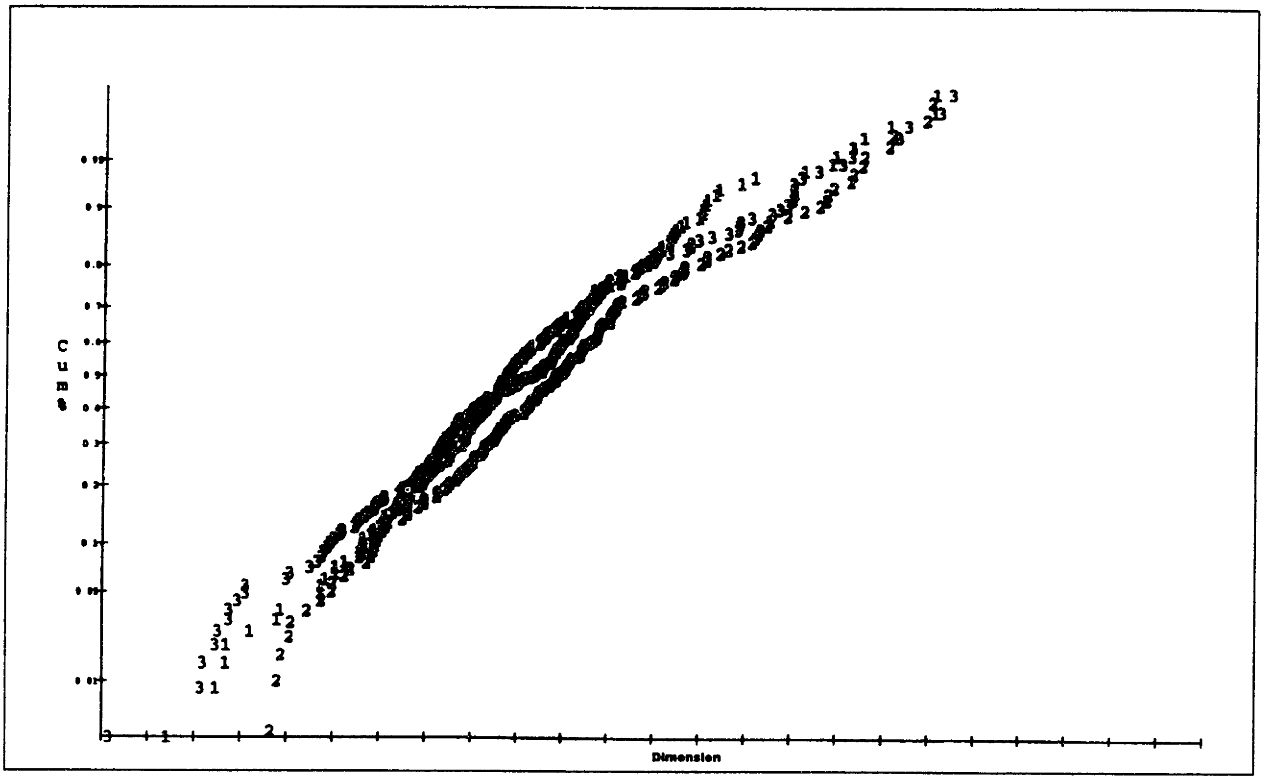


Figure 4 **Example of a cumulative probability plot of final check contact measurements.**

Operators may also measure differently as a result of their own personal work habits. Those work habits that are critical for consistently correct, low variability measurements include: using the picture book when they are unsure how to measure a certain feature; being diligent in absorbing specification changes or other updates; being diligent in always achieving good focus and astigmatism adjustment; and being reluctant to “dial-a-CD.” These habits as they apply to SEM measurements are strongly influenced by the personal style and motivation of the operator, but they can also be encouraged by trainers, cluster training coordinators, supervisors, or anyone else who has significant interaction with the operators. Good SEM habits, as others, can also be influenced by thoughtful incentive and responsibility structures that can help frame an operator’s desire to strive for high quality measurements.

4.2 Measurement Variability Observed from Standard Production Data

Some measurement variability data can be gleaned from standard production data. Data files available generally contain: lot ID, layer, product type, raw measurement points, the mean of those points, the standard deviation of those points, the time and date the measurement was completed, and the SEM on which the measurement was done. From these data, cumulative probability plots can be made relatively easily. With a large number of data points (typically on the order of hundreds) these plots are a convenient way to graphically illustrate differences in measurement distributions given an assumed random (equal) distribution of actual critical dimensions.

Cumulative probability plots by layer, for example, clearly show SEM to SEM variability, systematic differences in measurement values, that may result from SEMs that are slightly misadjusted from each other. The cumulative probability plots describe the success achieved in keeping the SEMs matched to each other, generally through the use of a pitch monitor, a standard structure of known dimension on a single wafer. At regular intervals the pitch monitor is measured on each of the SEMs in both x and y directions at each set of SEM conditions used in production. If the SEM reading deviates from the known value the SEM is adjusted accordingly. The correspondence is not perfect, however, because measurements of the standard structure are somewhat ambiguous in the sense that many different but valid measurement values can be obtained depending on rather small differences in gate placement on the structure. Other details of the pitch monitor measurements also suggest that a more robust system could be developed. However, because the effects, as determined by cumulative probability plots, were significantly smaller than operator variability, and also were being tracked with standard production data, I did not opt to further address pitch monitor procedures.

Cumulative probability plots (by layer by SEM) that show shift differences can also be constructed. Any shift differences almost certainly do not reflect product differences, but could indicate environmental differences (days vs. nights) or differences in measurement "techniques" from shift to shift if trends are consistent. Ideally, shifts should look the same: the value of a critical dimension should not depend on whether it was measured Monday during the day or

Thursday during the night. Although there are shift to shift differences for the contact layer at final check (a typical large difference is on the order of one arbitrary unit) there are no consistent differences that would suggest either a strong environmental factor or strong biases associated with particular shifts.

4.3 Measurement Variability Undetected by Standard Production Data

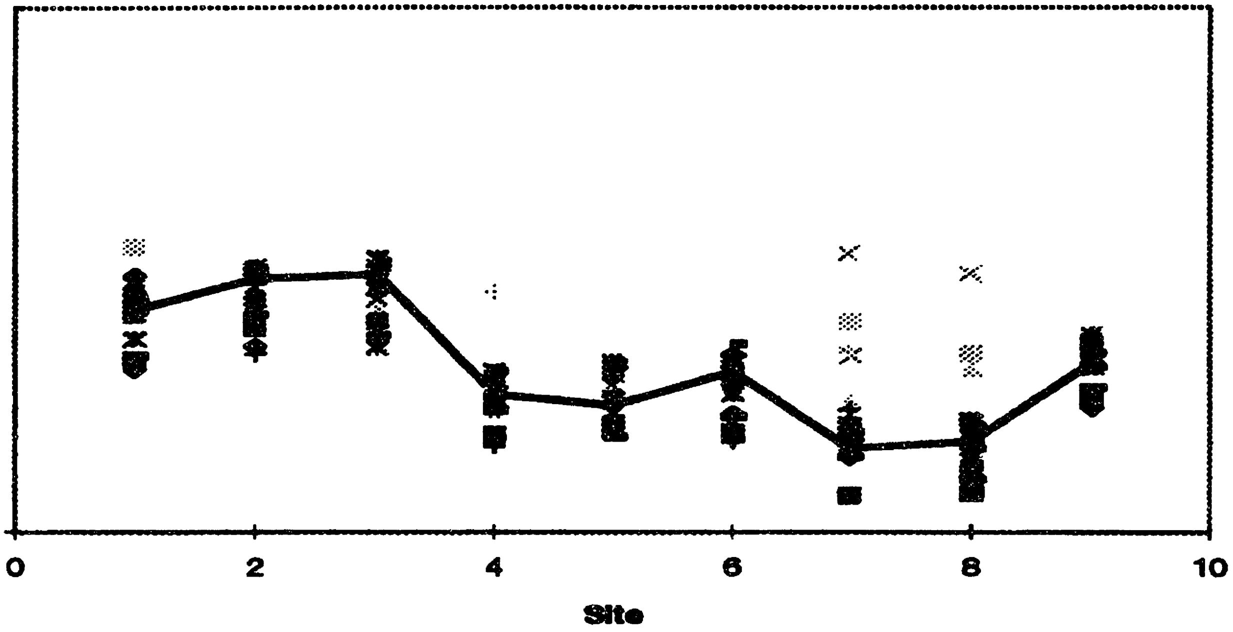
Section 4.2 described the types of measurement variability that can be observed from standard production data. There are, however, important sources of SEM measurement variability that can not be observed from standard production data, including operator variability and mismeasurement. While cumulative probability charts can be constructed from production data to give a reasonable estimate of SEM to SEM variability, the same can not be done for operators because while the data record does include the SEM used, it does not provide reliable information as to which operator made the measurement. In the immediate case, the lack of such information is an automation issue and a work habit issue. At root, however, it is a cultural issue that reflects a historical reluctance to make MTs responsible for their individual measurements. Thus, it was necessary to perform a special experiment in order to assess operator variability on the contact final check (or any other) layer, as described in Section 4.3.1.

Production data also can not give an indication of how robust the SEM recipes themselves are with respect to repeatability. SEM recipes, like many others, are generally passed to a fab intact either from another fab in the virtual factory or from the development organization. Changes may be made after that but require a formal approval process. Often, the development fab may start with recipes from another process and if they look adequate will leave them as is, confirming their utility but not necessarily optimizing them. In addition, recipes are often defined for a particular process, e.g., P1, but the interactions with other processes that might be running beside it in the same fab are not taken into account. A designed experiment addressed this recipe issue (Section 4.3.2) but found the associated variability differences to be insignificant.

4.3.1 Preliminary Study of Operator Variability

An important task of the internship was to determine whether there was in fact significant operator variability in measurements of final check contacts. A preliminary study was performed in Fab 7 and Fab 9 which addressed both develop check and final check contact wafers although the internship project continued with a focus on final check measurements for the reasons already described. Develop check and final check wafers for the preliminary study were patterned using a focus-expose matrix in order to provide a range of conditions on each wafer that would approximate a realistic production critical dimension spread. Using this matrix also had the added benefit, because the critical dimension varied significantly over the wafers, of preventing operators from thinking they knew what "the answer" was, or that all the sites would look the same.

The wafers were first measured using a much more automated SEM in Fab 7 (used mostly for engineering work). The values from the automated SEM, confirmed by engineering, formed the baseline values against which operator measurements were later compared. Two operator volunteers from each of the four shifts from each fab were asked to measure the final check wafer, using recipes similar to production recipes, once per shift for three shifts (the same was also done for the develop check wafer). Measurement values were then plotted against the baseline values for Fab 7 and Fab 9 for each of nine sites measured. Plots of these results for the final check measurements are shown in Figure 5 for Fab 7 and Fab 9 respectively. (The solid line represents the baseline values.)



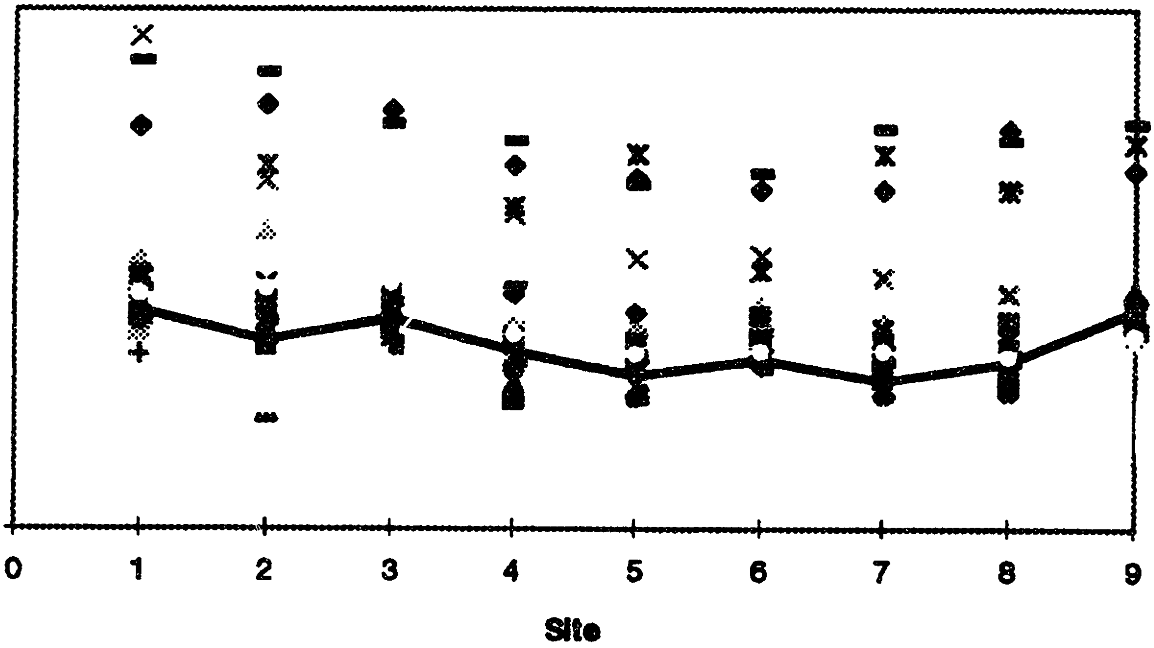


Figure 5 Results of preliminary study on final check contact measurements for Fab 7 (top) and Fab 9 (bottom). The solid line represents the baseline values for each wafer used. Vertical scales are identical.

The data showed, first, that a significant number of Fab 9 measurements were high – the highest were on the order of four arbitrary units larger than the baseline values. It was also evident that these large deviations occurred much less frequently in Fab 7. It was encouraging to note, however, that (in Fab 9) while some operators consistently measured high, others consistently measured very close to the baseline values, suggesting that the large operator-to-operator measurement variability was not necessarily characteristic of the system but rather reflected a training or other operations issue. This was supported by the somewhat bimodal nature of the data, with many points around baseline but a significant fraction rather high. As will be discussed in the next chapter, despite their relatively common occurrence the large deviations from baseline resulted from special cause mismeasurement, e.g., measurement of the wrong dimension, rather than from common cause variability around a basically correct measurement. In addition, the data show that in general an entire set of measurements (all nine sites) is high, as opposed to being variable within a set, confirming other observations that a primary source of variability and mismeasurement is SEM set-up. Thus, if set up properly for the first measurement, the rest generally follow; and conversely, if the first site is incorrect the rest will also generally follow.

This study was important because it definitively showed that a not insignificant fraction of measurements on the contact final check layer incorporated relatively large deviations from the baseline values. It thus provided compelling justification for investigation into causes of and remedies for such deviations or mismeasurements. Further, common cause operator variability can be estimated by looking at the data clustered around the baseline, with deviations not large enough to represent the usual types of mismeasurement. The standard deviation of this "good" data is approximately 0.4 arbitrary units. Inclusion of all measurements, however, results in a standard deviation of on the order of 1.3 arbitrary units.

4.3.2 Designed Experiment Investigation of SEM Operating Conditions

A Taguchi experiment designed using RS/1[10] and fit with a quadratic model was used to investigate whether the SEM operating conditions being used were optimal or whether they might be contributing significantly to measurement variability. Operating condition variables used in the matrix were accelerating voltage, beam current, and FCM, with the settings for each reflecting the range of conditions that were already being used on the SEMs for other layers and processes (and which reflected the historical ranges of conditions for Intel SEMs generally).

The design called for 11 "inner" runs (combinations of the factors just mentioned) and four outer runs for each inner run where the outer runs consisted of: a measurement of approximately process of record (POR) -sized contacts; a measurement of smaller-than-POR contacts; and a replicate of each. Each "measurement" was the standard deviation of five different contacts measured at a single site. (As will be described briefly later, the contacts at a particular site were quite uniform in size; different contacts were measured because repeated measurements of the same area tend to change the measurement.) Thus, a small standard deviation reflects conditions that give good measurement repeatability (low measurement variability), primarily because they result in an image that is unambiguous. Conditions that were sub-optimal typically resulted in an image for which it was very difficult to distinguish the proper edges for measurement, particularly the inside edge of the grey ring, and thus difficult to accurately repeat measurements. A summary of the runs performed as well as a ranking of results is shown in Table 1. The best ranking (1) indicates the lowest mean of the four (outer run) standard deviation values for that inner run.

The final check POR conditions are those in run 11 – while not the worst, they are also not the best. The mean standard deviation for the final check POR run was more than 50% greater than that for the best case, run 3. The develop check POR conditions for process P1 are those in run 6, which have a mean standard deviation of more than double the best case in run 3, confirming that simply using the develop check conditions for both develop check and final check wafers is not appropriate.

Table 1. Runs for SEM operating conditions designed experiment.

Run	Accelerating Voltage	Current	FCM	rank (best = 1)
1	-	0	+	4
2	-	-	+	8
3	-	+	+	1
4	-	+	-	2
5	0	0	-	6
6	-	-	-	10 (P1 develop check POR)
7	+	-	-	11
8	0	-	+	9
9	+	-	+	3
10	+	+	-	7
11	+	+	+	5 (P1 final check POR)

The model fit to the data by RS/1 suggested an optimal condition as represented by run 4, followed by that represented in run 3, which appeared to be the best set of conditions. To confirm what were believed to be optimum conditions, a different operator made 16 measurements on a single (POR-sized) site using the conditions of run 3 (suggested by the data to be the best), run 4 (suggested by the model to be the best), and run 11 (the process of record). In this case, the conditions of run 3 again appeared to be better than those of POR based on average standard deviation (POR 34% greater than run 3) while the conditions of run 4 actually appeared to be slightly worse than POR (POR 15% smaller than run 4). An F-test on the data from run 3 compared to run 11 (16 measurements each), however, gave a p-value of 0.28, indicating that the difference in variance is not significant at an $\alpha = 0.05$ level. A t-test on the same data gave a p-value of 0.36, indicating that the difference in means was also not significant at an $\alpha = 0.05$ level.

While the improvement in variability was statistically insignificant, repeatability at least as good as POR was determined and a set of operating conditions was suggested that is of interest because it uses the same voltage as that used for P1 develop check layers. Because changing the accelerating voltage essentially involves readjusting the SEM alignment, the opportunity to reduce the number of voltage changes from layer to layer during measurements is of interest. Although it is difficult to estimate the quantitative impact of changing SEM conditions, and especially voltage, from layer to layer, it is clear that each such change requires adjustment before the next measurement and that there appears to be a "settling time" before the SEM again reaches optimum stability. For this reason, it is standard practice to run sets of lots together on a particular SEM that utilize the same operating conditions. Currently, the Fab 9 SEMs run on the order of six different sets of operating conditions. Any opportunity to reduce the number of condition changes could be advantageous in terms of system efficiency.

It was thus of possible interest to Intel to consider making such a change to POR for this layer. In order to support a proposal for such a change, the proposed condition (that from run 3) was compared to POR using production wafers. Using n = 36 data points, an F-test indicated that the variance was not statistically different (larger variance would suggest the proposed change not be made and would at least require much more extensive justification), and a paired t-test indicated that the means were not statistically different (so if the change were implemented there would not need to be an investigation into changing the target). However, although there may be system gains to rationalizing SEM conditions, relative to operator variability issues already discussed, the variability associated with SEM conditions is insignificant and was not felt to represent the best opportunity to reduce measurement variability.

4.3.3 Within-Site Variability

As mentioned in the previous section, in the current experiments as well as in other work at Intel, a common assumption is that the contacts at a particular site are uniform in size. This assumption seems reasonable based on measurements by a single operator on a single SEM of several sets of ten different contacts at a single site using POR conditions. The average standard deviation over the sets of ten contacts was approximately 0.2 arbitrary units. While it is impossible to separate out within-machine variability or within-operator variability, it does suggest that even in the "worst" case – where all observed variability would be attributable to actual dimension differences – the variability is small. The variability between sites, on the other hand, can be significant but reflects on the process rather than on the measurements and thus will not be considered further.

4.4 Summary Comparison – Magnitudes of Sources of SEM Variability

The sources and magnitudes of measurement variability and mismeasurement described in this chapter are summarized in Figures 6 and 7. Figure 6 shows that the effect of operator mismeasurement is significantly larger than either the effect of using different, imperfectly matched SEMs or the effect of other (presumed) common causes that result in observed (but not consistent) differences between shifts. The work described in this thesis thus focused primarily on variability caused by mismeasurements because of their large magnitude; their relatively high frequency as indicated by the preliminary study; and because these measurement errors had not been previously addressed in any significant way, were not being monitored, and in fact could not be monitored with existing production data. Variability resulting from SEM to SEM differences is not insignificant but was both being monitored and addressed using production data. Figure 7 compares the standard deviations representative of all the final check preliminary study measurements (assumed to represent both common cause and special cause variation); of only those measurements that were relatively close to the baseline (assumed to approximate common cause variability rather than measurement errors); and of multiple contacts on a single site. The plot makes clear the relatively large magnitude of the observed special cause variability (mismeasurements).

Chapters 5 and 6 will describe in greater detail the causes of the observed operator variability and measurement errors. They will also detail the efforts made to reduce measurement variability, both through training improvements and updates as well as through development of an automated, anonymous feedback system for operators.

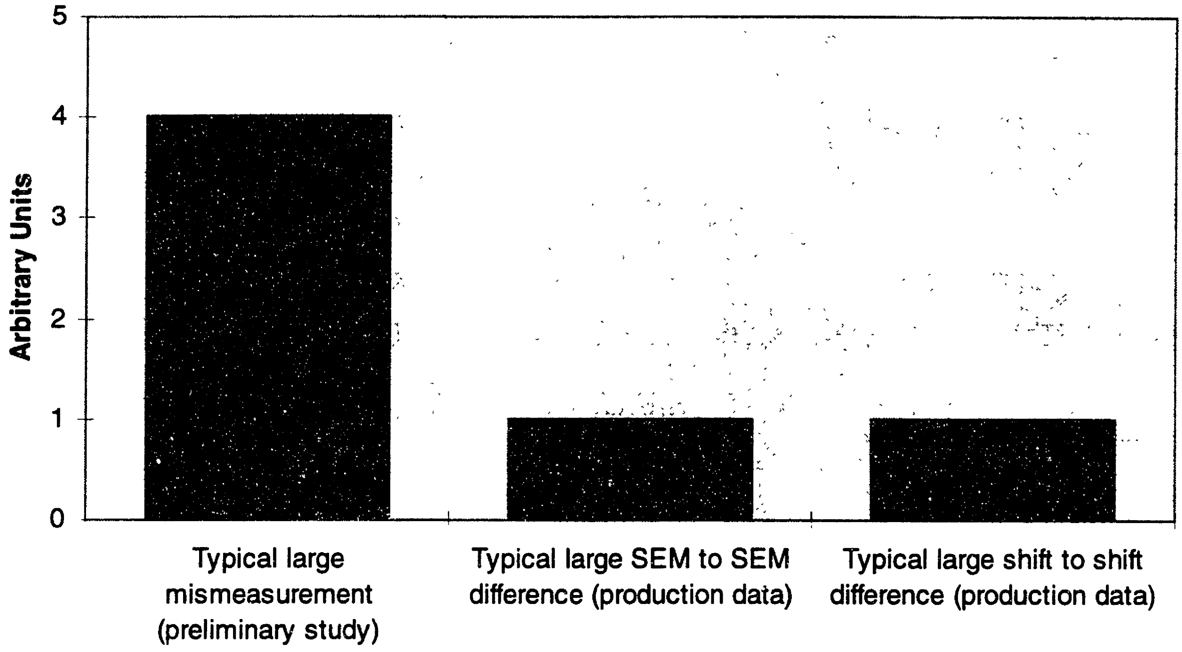


Figure 6. Magnitudes of common measurement differences.

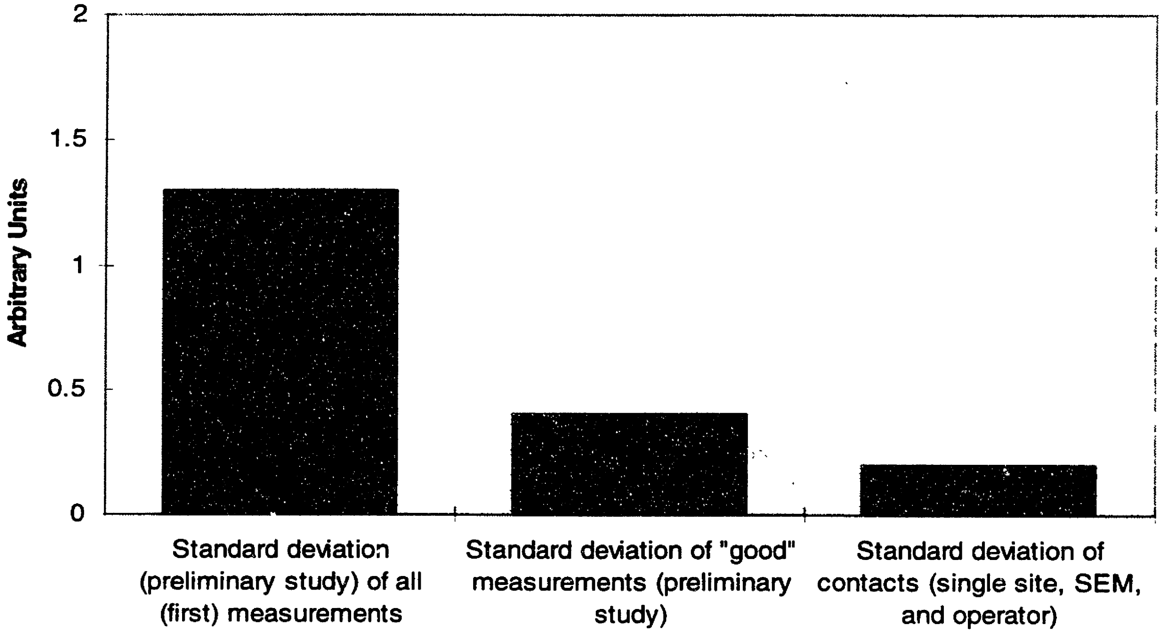


Figure 7 Magnitudes of different types of variation.

5.0 CAUSES of OPERATOR VARIABILITY and MISMEASUREMENT

As noted in Chapter 4, mismeasurement of CDs appeared to be the most significant contributor to measurement variability for contact final check layers. In the process of starting to look more carefully at operator measurements, conducting the preliminary study, and doing informal retraining after the preliminary study, the principal forms of incorrect measurements became clear. The most common problems were: simply measuring the wrong dimension (typically measuring the inside of the bright ring as sketched in Figure 1, rather than correctly measuring the inside of the greyish ring), because of confusion with another layer, lack of familiarity with the layer, or failure to correct misplaced gates; measuring the wrong dimension, as above, because of an incorrect SEM set-up such that the correct dimension is not clearly visible (typically, screen brightness turned down too low, or poor set-up, in wobble, astigmatism, focus, or a combination, that smears out the inner ring); and measuring the wrong contact. All represent training issues of various sorts. Common cause operator measurement variability (assumes correct dimension being measured) relates mostly to sub-optimal but not severely compromised set-ups. Because of the relative magnitude of the effects of actual mismeasurement, the project focused on reducing mismeasurements although many of the training updates undertaken would also be expected to help reduce common cause variability in so far as they are successful in promoting better, more uniform set-ups and measurements.

As mentioned above, one common mistake was MTs having good SEM set-ups but mismeasuring by measuring the inside of the bright ring rather than the inside of the grey ring. A common cause of this mistake appeared to be confusion with other layers, namely with the contact layer from a previous process, and with the via layer (also a hole) from process P1, where there was a similar geometry in the sense of a bright outer ring and a darker inner ring but correct measurements were made at the inside of the bright ring because of differences in the hole profile. Some operators seemed to assume that all holes were measured the same way, regardless of process or layer. Thus it became apparent that any successful retraining efforts should also include the via layer of the current process to avoid simply pushing the measurement errors from one layer to the other.

Despite the confusion described above, it appears that many operators, if presented with a good set-up, measure the correct dimension. It was not uncommon, however, for operators to have a poor set-up (e.g., they "inherited" the SEM with the brightness too low to see the inside ring, or failed to adjust the SEM for a good image) that resulted in the inner grey ring not being visible. In this case operators did not always seek to correct the set-up but rather simply measured the structure they saw, resulting in incorrect measurement of the inside of the bright outer ring. This type of error thus results in a larger CD value than it should be – on the order of four arbitrary units too high, as observed in the preliminary study.

While not resulting in as large a deviation, one occurrence observed in production that would contribute to greater variability was measurement of the wrong "type" of contact at a site. Although a particular contact is not specified for measurement, the contacts at a site are of two types, with one type clearly specified (in the picture book) for measurement. Measurement of contacts of both types at a single site confirmed a dimensional difference of approximately one arbitrary unit. Measurement of the wrong contact is a similar issue to measurement of the wrong dimension (with a good set-up) in that it stems from confusion with other processes or layers, but also in that it results from operators mismeasuring because they don't know the correct way to measure and either don't realize that they aren't measuring correctly or don't take the time to check the picture book if they are unsure.

Another issue that arose was the question of orientation. Although contacts are holes, they are not perfectly round and thus measurement of misoriented contacts can add to measurement variability. While orientation should not be a question, in practice it sometimes was if recipes were not exactly correct so that on some layers operators had to, or thought they had to, manually "x-y exchange" the image. The magnitude of the effect of x-y exchange of the image is approximately 1.3 arbitrary units.

All of these types of errors suggest shortcomings in the training system as well as lack of operator diligence, itself arguably a training or system shortcoming.

6.0 REDUCING OPERATOR VARIABILITY

Previous chapters described the significant measurement variability observed for CDSEM measurements of final check contact layers and explored likely causes. This chapter will describe system characteristics and documentation details that relate to CDSEM variability, and will also describe actions taken during the internship to reduce that variability. The actions can be broken down roughly into: improving the picture book; doing informal training updates; and creating an automated feedback system. Development and implementation of the automated feedback system represented a significant part of the internship effort and will be described in Chapter 7.

Training

The primary inputs that a typical SEM operator gets on how to measure a certain structure are: training as they are being prepared for certification (this is verbal/demonstration from a trainer and with the picture book as reference); picture books available at all times for all layers at the SEM; trainers available for questions (since trainers are mostly running WIP themselves); and any updates or specification changes that may occur. Although in theory this array of information seems sufficient, it relies heavily on operator initiative (with respect to picture books and training help beyond certification); on effective communication with operators (with respect to updates); and on nearly blind trust that measurements are being made as they ought to be (as noted already, there are no real checks on individual operator performance). In practice the assumptions above seem not to have been met, leading to shortcomings being addressed in this thesis. Specific instances where the structure of the system seems to be less than optimally effective in terms of reducing mismeasurement and variability are described below.

Currently, there is no individual sign-off of each layer when operators are being certified. One trainer typically has a trainee do two lots per layer (supervised), but this is not documented in the current baseline package/checklist and the practice does not appear to be shared across all trainers. The general assumption made by the baseline package is that each operator will be sufficiently trained; however, without being explicitly named on the checklist there is less assurance that each layer will be specifically demonstrated. Also, naming each layer would mean that operators specifically acknowledge their understanding of and comfort with each layer since MTs must countersign each checklist item. It seems to make sense then, and would not be a large change to the baseline package or general way of training, to require that, e.g., two lots per layer be called out in the baseline package checklist.

Another system issue is that there is no current procedure for recertification or for updating a certification. Informal observation is the only way that an operator's not-quite-correct measurements are likely to be noted and corrected. Having a more formal, scheduled procedure for making sure that operators are current would likely improve consistency among operators.

Two other training system practices also seem to encourage variability. The first is that the relatively large number of trainers, and lack of coordination among those trainers, can reasonably be expected to lead directly to higher variability among operators being trained. Becoming a trainer appears to demonstrate initiative, interest, and desire for additional responsibility on the part of the operator (useful for performance appraisals), but in rewarding those people by making them trainers the system is setting itself up for degraded performance in terms of increased measurement variability. A large numbers of trainers leads to greater variability among trainers, which is then passed on as even greater variability among operators who each add their own variability.

Furthermore, for trainers to be effective and to minimize variability, trainers need to be coordinated and current. This coordination is an investment with respect to the time it takes away from running production and therefore underscores the benefits of reducing the number of trainers, particularly as the trainers are used for "reference" relatively infrequently (operators seemed more likely to ask whoever is sitting at the next SEM). To reduce the number of trainers and at the same time help make the training structure clearer, a goal should be set for the number of trainers per shift (this might be two trainers per shift, in addition to the CTC), which could be accomplished in a reasonable period through attrition.

Another issue is that potential coordination and allocation and acceptance of training responsibilities among MTs in various positions is clouded by the many overlapping formal and informal designations of expertise (trainers, peer trainers, CTCs, coordinators). Who "the" expert is is not clear. Performance of trainers would likely be enhanced by making specific expectations and responsibilities more clear. The ability of yield or process engineering to communicate effectively with the floor on measurement issues would also be significantly improved.

Picture Book

Another system issue relates to the use of the picture books. While the picture books were used, they were not in constant use. It appeared to me that given the relatively short time many of the operators had been running SEM and the number of different layers being run, the picture books were used relatively little. In part, this may relate to reluctance on the part of operators to ever appear unsure. In part it may also relate to being slowed down by a single, relatively large, bulky book that does not fit easily at the SEM. Although several operators asked did not feel that having more picture books available would help, I continue to think that many issues with measuring the wrong dimension could be eliminated by having a constant reference, perhaps just a simple double check for the majority of operators. It might, for example, be beneficial to have a screen that for each lot displayed how the measurements should be done. Then it would not depend on whether the operator wanted to bother with finding the book, leafing through, etc., but rather just on whether the operator looked up at the screen during several seconds where they are not busy – e.g., while the wafer is loading.

A contributing factor to operator variability may have been that existing picture book pages for the contact final check layer, although correct, were not very clear, in part because they were at a lower magnification than was currently being used for production wafers and in part because they did not give clear direction in common areas of misunderstanding. Thus, the picture book in that case did not provide clear guidance for an unsure operator.

An interesting observation, however, was that although many MTs agreed that the (original) picture book pages were not very clear, I never heard of operators having requested clarification. This seems representative of a general reluctance, of people in general, not to do anything that will show (in this case, to other operators, or trainers, including the trainer that certified them, or engineering) that they are unsure about a measurement they may have been doing for a long time and for which in any case they are certified. Part of the system shortfall in this case is that it is difficult to know when something as crucial as the picture book needs improvement if MTs and/or trainers (the people who use it most) are not willing to ask for more or better information.

6.1 Informal Updates on Contact Layer

As was described in Section 4.3.1, a preliminary study was done to assess the variability of final check contact measurements. Following confirmation that there were indeed a large number of erroneous measurements, an informal retraining effort was undertaken. To provide as much consistency as possible the Fab 7 engineering technician who was the Fab 7 "master" SEM trainer and de facto SEM expert on the manufacturing side for both fabs, came to Fab 9 to review correct contact layer measurement with the peer trainer for each shift. Those peer trainers were then to update all other operators on their shift, with a list to engineering as to who had been updated. After training updates were complete, a repeat of the preliminary study would be done to confirm successful retraining.

In fact, the Fab 7 trainer updated three of four peer trainers. Because of constant scheduling problems, a Fab 9 engineer familiar with the issues updated the fourth. Most but not all operators were updated by their peer trainer over a period of several weeks. However, upon repeating the study it was clear that the retraining efforts had not been as successful as hoped. Although the study was not completed, primarily because of a shortage of available SEM time, the first sets of measurements (Figure 8) had many large deviations from baseline and clearly indicated that the training had not been effective. Transfer of a relatively simple update to all SEM operators was much more difficult than anticipated, both with respect to reaching all operators and with respect to effective information transfer.

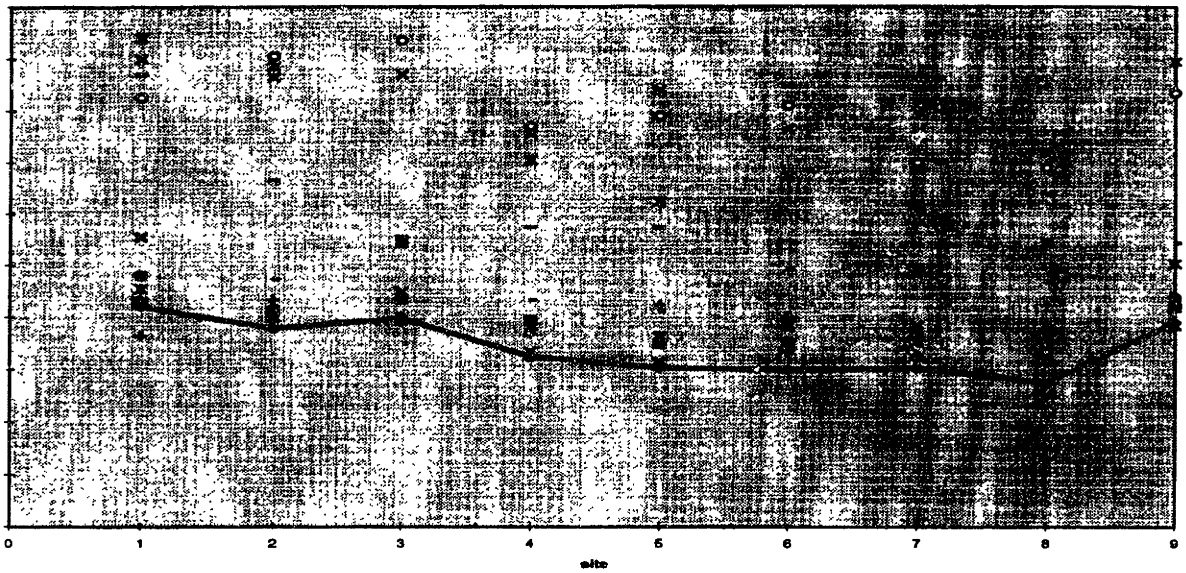


Figure 8 **Repeat of preliminary study measurements in Fab 9 indicated that despite retraining efforts there was still large measurement variability. (Vertical scale identical to plots in Figure 5.)**

Eventually several problems with the retraining effort became clear. First, although nobody said anything at the time, it seemed that the Fab 7 person did not have sufficient credibility with Fab 9 peer trainers (and operators). This meant that if the Fab 7 trainer said something that was different from the usual Fab 9 way, operators were likely to say, 'oh it's just Fab 7 – we don't do it that way' and continue to do it the old, familiar way. Although the Fab 7 person was in many ways an excellent person to do the update -- she was an engineering technician with lots of experience and expertise and had been part of the decision process on how the layer should be measured -- there are also, in hindsight, several reasons why it was not particularly successful. First, it is better to have the "authority" on the update in the same fab so that there is always ready access for follow-up questions to address misunderstandings, and because, although fairly similar, both the SEMs themselves and the automation are slightly different in Fab 7 than in Fab 9. The master trainer thus had credibility with engineering and had significant expertise but did not have sufficient credibility on the Fab 9 floor, mostly, I believe, from lack of exposure. Similarly, although many operators were apparently updated by their peer trainer, there was no acknowledgment of this by the operators themselves, which underscored the informality of the retrain and reduced its credibility (since specification changes, for example, typically require sign off by each operator).

The second problem developed from an issue already referred to – namely the confusion with the via final check layer. Neither engineering nor apparently the Fab 7 trainer had been aware that the confusion between the layers was an issue. Thus, when the retrain addressed only the contact layer, without any mention of or retrain on the via layer, confusion was heightened and there were a few problems with via layer measurements because of this confusion. Thus (similar to the situation of the picture book not being adequate), it was important to know what the source of confusion was so as to be able to address it.

6.2 Picture Book Updates

As noted earlier in this section, the picture book reference was not optimal for the contact and via layers. Improved picture book pages, including clearer pictures and explanatory text, were designed with input from and informal sign-off by the (engineering) layer owners as well as several other process engineers familiar with the layers and issues. The pages were also informally reviewed by the CTCs prior to incorporation to try to ensure that they would be clear to the floor.

After the picture book pages were corrected peer trainers were asked to update all the operators on their shift and vax mail messages were sent to all operators. It was thought that this would be sufficient, especially given the previous informal retrain and the new and improved picture book pages. However, it was clear that whether or not they had all been updated, some operators, and indeed some trainers, continued to measure incorrectly.

To finally try to be sure that all operators actually saw the new pages, operators were asked to review the contact picture book pages for both P1 and a previous process (a page had also been corrected to be consistent, as had several develop check pages) and to then sign off on their review and understanding of the pages. This update appeared to be more successful, with almost all certified operators having signed off. Although it had once seemed unnecessary, including to some in the engineering group, to be so formal for an apparently minor review, it appears that the inertia on the floor, coupled with its lack of coordination in some respects and lack of time for group training, is such that formal mechanisms are the only way to get updates through – and even so it requires significant effort.

Operator Responsibility

While the CTCs ended up being the best route for carrying out training updates, discussing changes to the baseline SEM training package, having operators sign off that they had reviewed new picture book pages, etc., because they meet with each other regularly and training appears to figure prominently in their responsibilities, the situation raises the question of why operators are not more motivated to keep themselves up to date. A system issue, the most obvious answer is that there are neither positive incentives for measuring correctly individually or for achieving a tight, accurate measurement distribution collectively, nor is there negative feedback to an individual's mismeasurements or even to the distribution of the population. With the current system, if an operator's measurements are incorrect, it is unlikely to be noticed other than by chance observation.

6.3 Automated Anonymous Feedback System

An automated system using a standard wafer to provide feedback to operators on their contact layer final check measurements, including their measurements relative to the population, although not ideal, appeared to be potentially useful for reducing variability. If used during training, particularly for the most difficult layers, such a system could also help to promote a more uniform approach to measurement. The system developed and implemented on a pilot basis is described in more detail in Chapter 7.

7.0 AUTOMATED FEEDBACK SYSTEM

This chapter will describe the automated feedback system in some detail, including the original goals of the system and associated constraints for its operation; the process of designing the system itself; how the system works; what the data have shown; and finally, other more qualitative results and observations about use of the system. The system was developed in an effort to reduce operator variability in CDSEM measurement of final check contact layers by giving immediate, anonymous feedback using a standard wafer that all operators measure. Because of the confusion with the via layer described earlier, a via final check wafer was also incorporated into the system.

7.1 Goals and Constraints

The principal goal of the system was to tighten the CDSEM measurement distribution on final check contact layers. An additional benefit would be a current assessment of operator variability on this layer. The principal constraint was that the system not be a significant burden on operators, in terms of time, complexity, or individual performance pressure. A secondary constraint was that the system be set up and maintained with reasonably modest engineering and automation support. The system was therefore designed with input from MTs, engineering, and automation. The backbone design for the system was to have a standard wafer with "known" dimensions that all operators would measure on a periodic basis, with feedback on performance to each operator individually and with population data accumulated and reported anonymously.

7.2 Process for Designing the System

The system was designed from within the litho process engineering group but was developed in concert with both manufacturing and automation. Since the system was being built for MTs it was important to have manufacturing input into its design so a system could be developed that would be practical in terms of both floor operations and MT acceptance. Thus, the supervisor for each shift was asked to choose one SEM operator to be the shift representative for the automated system. Over the course of several meetings, the shift representatives were presented with the preliminary data that was the motivation for the system; were presented with a proposed structure for the system; made suggestions based on the initial structure; made suggestions further along as automation constraints became clearer; and were consulted on major changes. For example, the initial system suggested that MTs measure the wafer once per quarter and that a single SEM be used. The shift representatives suggested, however, that monthly measurements would be better and that it would be better to have more SEMs, so access would be easier, even at the cost of having to accommodate some modest mismatching of SEMs. Based on automation constraints as well as MT suggestions the system was created.

The support of the litho shift supervisors was important for providing legitimacy to the project and because use of the system competed, albeit relatively modestly, with production. Thus at a presentation of the proposed system, supervisors agreed to a pilot for three months, at which time a re-evaluation would be conducted. While not generally concerned about the fine details of the system they did feel that they wanted more feedback by shift and suggested that the alternate IDs assigned to make the system anonymous at least be grouped by shift. In this way they could see if they had a particular problem and needed to devote many more resources to SEM training issues, or if they were in relatively good shape. As there were no objections from shift representatives, the system was in fact implemented that way. It proved to be very useful to have by-shift divisions as it also served to readily highlight which shifts had very high participation and which relatively low, allowing for both direct positive feedback and peer pressure to encourage participation.

The system was also designed with many inputs from the automation group since the anonymity constraint meant that there had to be a significant automation component. Also, because the system is quite different from anything in current use, a number of iterations were required to arrive at a system that was simple enough from the automation side to be able to complete it and get it on the floor in a reasonable length of time while still meeting the key requirements identified by the design team.

7.3 How the System Works

Each operator has an alternate automation system-ID of the form SEM01, SEM02, ..., with designation of SEM#s by shift and with an associated password, for making the measurements. Each operator measures a single final check contact and via layer wafer (the wafer for each layer was kept at SEM and available at any time) once per month using preset programs that go to six sites using standard production conditions. Associated with each site is a "baseline value" for the CD as determined by the automated SEM in Fab 7. A "delta," the operator measurement less the baseline value, is determined for each site and the average and standard deviation of the six deltas are displayed after the measurement is completed. If the delta mean or standard deviation is larger than a certain value (1.25 arbitrary units was used – a nonstatistical value that was agreed to be a reasonable measure of satisfactory measurements and one that was achievable even given the use of the system on several SEMs) then a message is displayed asking the operator to check the picture book or with a trainer and then to remeasure the wafer.

Before introducing the system to operators, shift representatives were first invited to test out the system so they would be familiar with its actual operation. Vax mail messages were then sent to each SEM operator from automation, with their alternate ID and password, and from engineering, explaining how the system worked. In addition, the system was explained in person, either by the shift representative or engineering, to operator groups of various sizes. This combined approach was only somewhat successful as while many operators seemed much more compelled by presentation in person, they didn't remember all the details and were reluctant to use the vax messages for reference.

A graphical month-to-date summary of results for each layer was posted at SEM every (engineering) day to provide feedback to operators on the population as a whole. Physically posting the charts also proved to be a good way to foster informal discussions with operators about the system. At the end of each month summary charts were posted which also showed percent participation by shift to provide feedback on and encourage participation. These summary charts were also distributed to litho shift supervisors and to litho process engineers.

In addition, one week prior to the end of each month, automation reviewed measurements to date for the month and sent a reminder by vax mail to each operator that had not yet measured asking them to please do so before the end of the month. At the end of the month automation provided each shift representative with a list of the operators (but not their alternate IDs) on their shift that had not yet measured so they could particularly encourage participation. Because the system was a pilot (with an undetermined future), measuring was not tied to certification or otherwise enforced; participation was strongly encouraged in as many ways as possible but was ultimately voluntary.

7.4 Costs of the System

The system was developed over several months and was primarily limited by the availability of automation support. The initial costs of the system included: engineering and automation time over a period of weeks to work out the details of the system and to make sure it was up and working correctly; a few hours of the MT shift representatives' time to help design and test the system; a wafer for each layer; engineering time to introduce the system to the MTs; and automation time to set up alternate IDs and passwords for each MT. The ongoing costs of the system include: time for each MT to measure each wafer once per month; time for the shift coordinator to encourage measurement and answer any questions; automation time to add or delete MTs from the system and to send out reminders to operators who hadn't measured; and engineering time to post compiled data regularly on the floor, to interface with the floor as needed, and to provide updates monthly to supervisors and to engineering.

7.5 Results

While participation during the pilot was reasonable (~90% of SEM operators used the system at least once, ~ 67% had participated in at least two months) only two of the three months of the pilot resulted in a sufficient number of data points to be considered representative of the population. Further, the data that there are appear inconclusive. As will be evident below, at face value the data would suggest that the system had actually made the situation worse. However, feedback from and interaction with many operators does not support that conclusion. Qualitative results will be discussed in Section 7.6.

November Contact Layer

Figure 9 shows the results of the automated system measurements for the month of November for the contact layer. On the x-axis, operators are identified by their alternate IDs. Participation was approximately 80%. The y-axis shows deviation from baseline, with limits also marked, beyond which operators are asked to review the measurement procedure and then remeasure. SEM32, for example, at the far right, made an initial measurement that was outside the limits followed by an acceptable measurement. The standard deviation of the first measurements for all operators on the chart (no repeats) is approximately 0.85 arbitrary units.

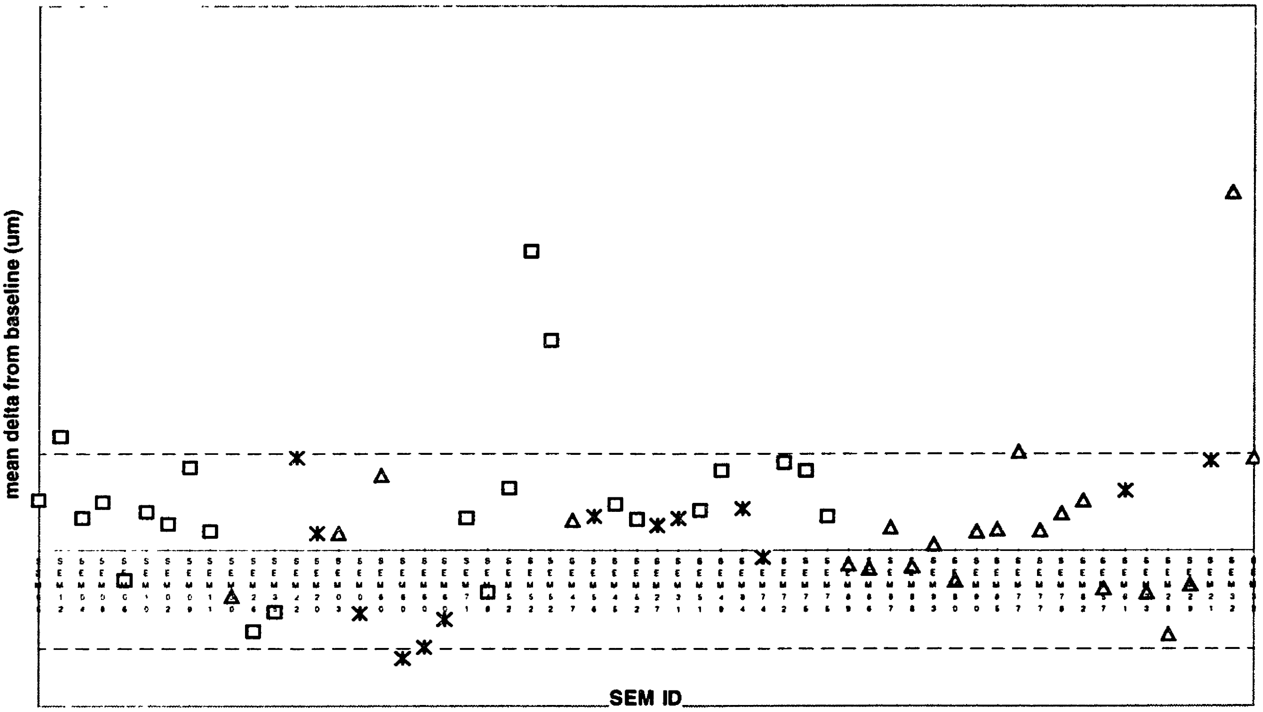


Figure 9 **Automated system data for contact layer measurements in November.**

The data can also be used for P/T estimates. At Intel a P/T ratio smaller than "**a**" is considered to represent good measurement capability, while a P/T greater than **a** but less than "**b**" is considered satisfactory, and P/T greater than **b** is considered poor capability. The November data, using all the first measurements gives a P/T >> **b**, while for first measurements that were within the limits (i.e., variability assumed to be the result of common causes rather than special causes) a P/T > **b**.

For comparison, the preliminary study results had for all first measurements a P/T >>> **b** and for all "good" (within the limits) first measurements a P/T > **a** but < **b**.

December Contact Layer

Figure 10 shows the results for the contact layer in December. Participation is again approximately 80%. Most apparent is the much larger number of initial measurements that are outside the limits. The standard deviation of first measurements is also much larger, as expected (approximately 1.5 arbitrary units). For this data, the first measurements have a P/T >>> **b** but for good first measurements a P/T > **a** but < **b**.

The cause of the apparent increased variability is not clear. The key question is whether the true variability is larger, or whether the apparent variability is a result of the system not adequately controlling for relevant system variables. Using the automated system to estimate the measurement variability in production thus bears a certain resemblance to using production measurements to estimate the process variability – each step adds its own contribution (and potential error) to the final data. For this reason, an ideal measurement of measurement variability would use the production measurements themselves rather than adding an additional layer of measurement for the evaluation.

Returning to the December data, it is possible that the apparent increase in variability accurately reflects a true increase in variability, indicating that operators were actually more confused about how to measure, but other observations do not necessarily support this. My feeling, based on my conversations with a number of operators, is that operators were becoming more knowledgeable and that the apparent increase in variability was due to artifacts of the system. For example, an increasing comfort level with the system appeared to lead to less careful use (at least one operator told me they had not been paying much attention and "suddenly" got a bad set of measurements, which they then went back and redid correctly). It is unclear whether being less careful with the automated system is a more or less accurate approximation of production.

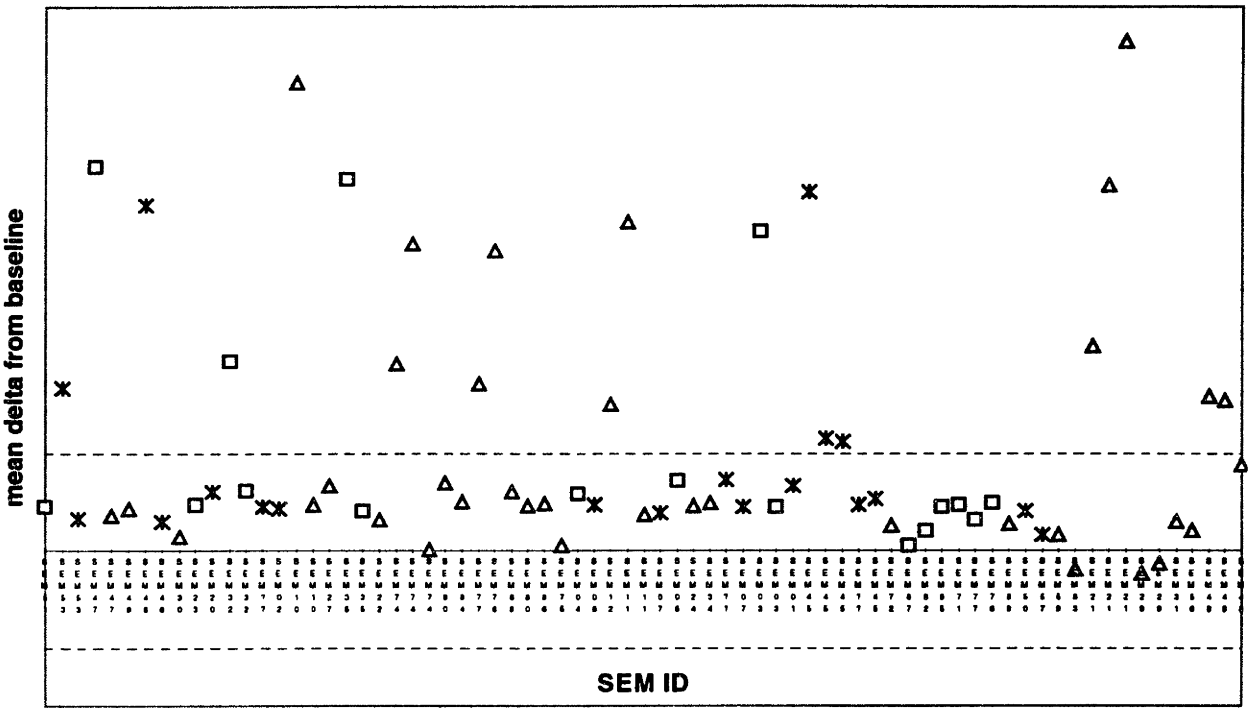


Figure 10 **Automated system data for contact layer measurements in December.**

In addition, greater variability may be the result of fewer automated system measurements being done as one big bunch. Some shifts, and at some times more than others, had operators doing their automated system measurements one after the other on the same SEM. While this was quite effective for increasing participation, it also meant that everyone after the first person came to a system that was at least reasonably well set-up for the measurement. Since, as described earlier, the set-up appears to be very important (with a number of operators likely to measure correctly if the set-up is correct, but not if it is not), coming into a reasonably correct set-up would likely improve the sets of measurements in general. The system would thus be a better training tool if set-ups were done from the beginning, e.g., with screen brightness turned down and the SEM perhaps adjusted to a different set of conditions. It would be difficult to implement this, however, because it would require more time, both at the finish of the previous measurement and at the beginning of the current measurement; it would be different from the way production is run (machines are generally left as they were used for the last lot, with screen brightness only turned down if the SEM will be idle for a time); and because it would have to be done manually by the operators themselves and it would be impossible to confirm that it was actually being done.

January Contact Layer

Relatively few measurements were made in January (participation only about 20%), primarily because of significant organizational changes in the fab around this time. Secondarily, I was on site only through most of December, and thus handed the system off to another litho metrology engineer who agreed to carry on for me but had neither the accumulated relationships nor the time to focus on system participation the way I had. January data are shown in Figure 11, but are not necessarily representative due to the small number of data points. In this case first measurements give a P/T >>> **b** and for good first measurements a P/T > **a** but < **b**.

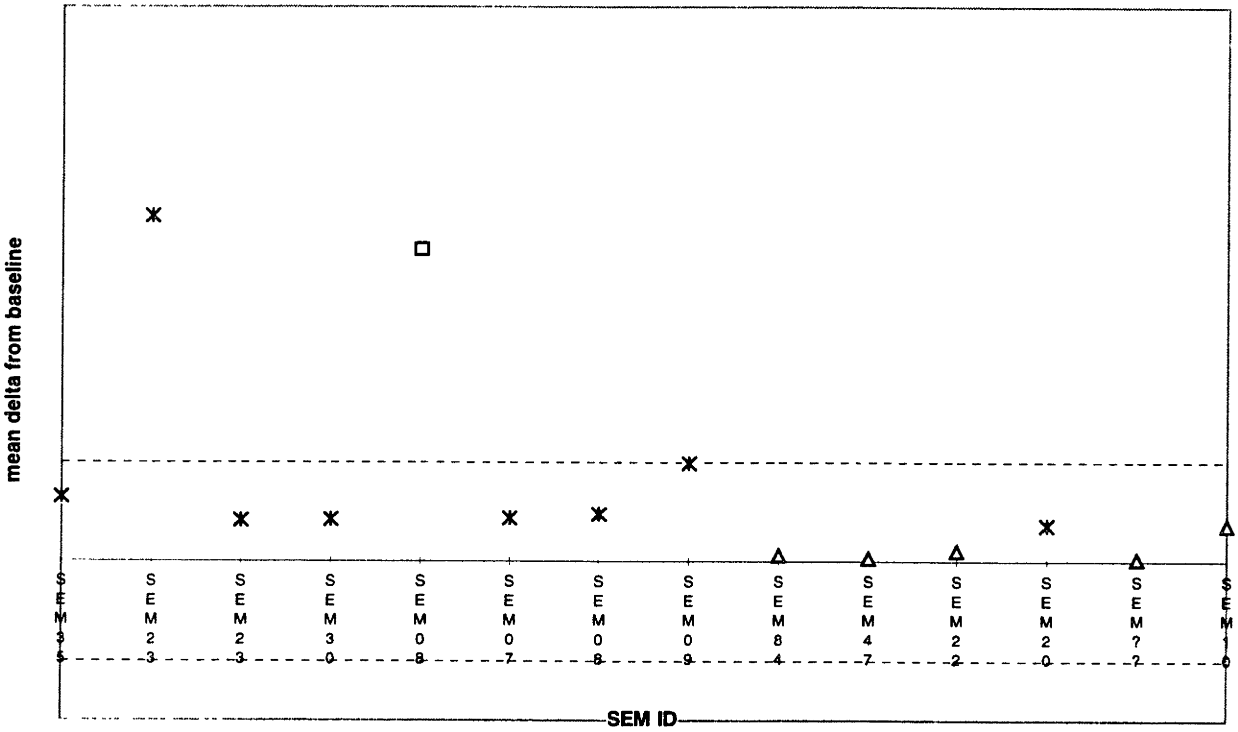


Figure 11 **Automated system data for contact layer measurements in January.**

Via Layer

As mentioned above, a via final check wafer was also part of the system (for the latter two months). The December and January data are shown in Figures 12 and 13. Interestingly, although confusion between the contact and via layers had been a significant issue, the mismeasurements are disproportionately associated with the contact layer. As with the contact layer, January data is insufficient to fairly represent the population.

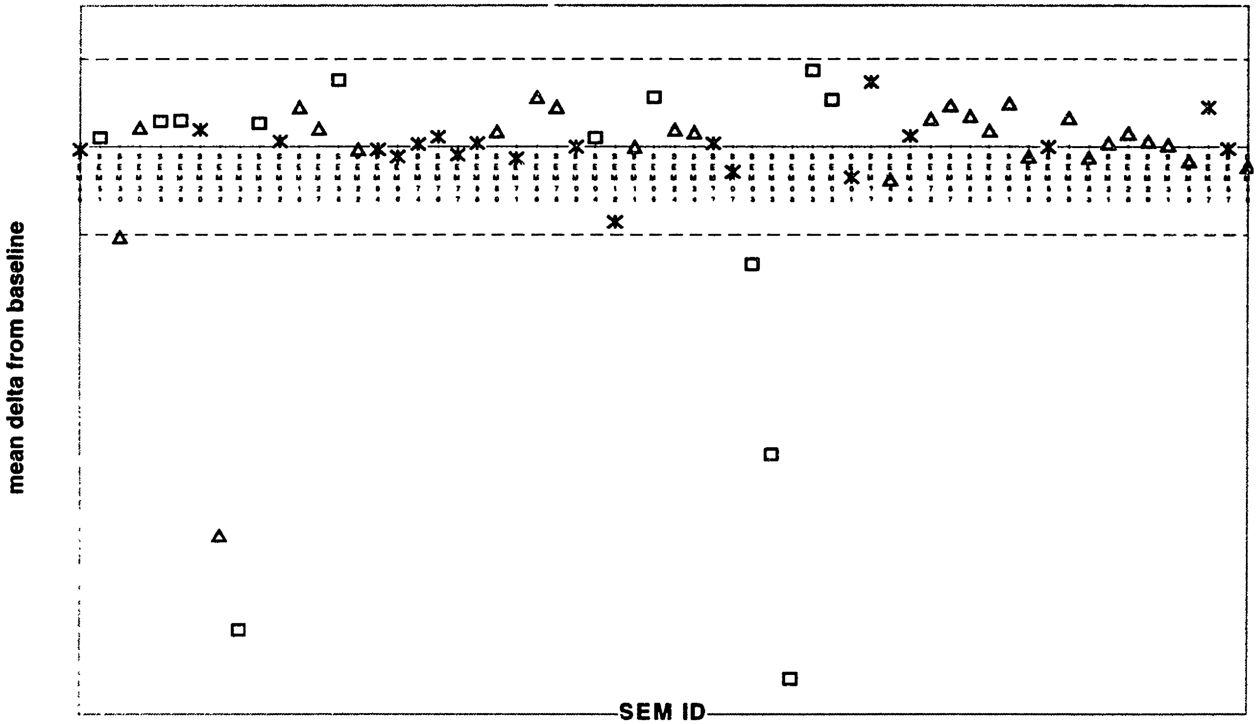


Figure 12 **Automated system data for via layer measurements in December.**

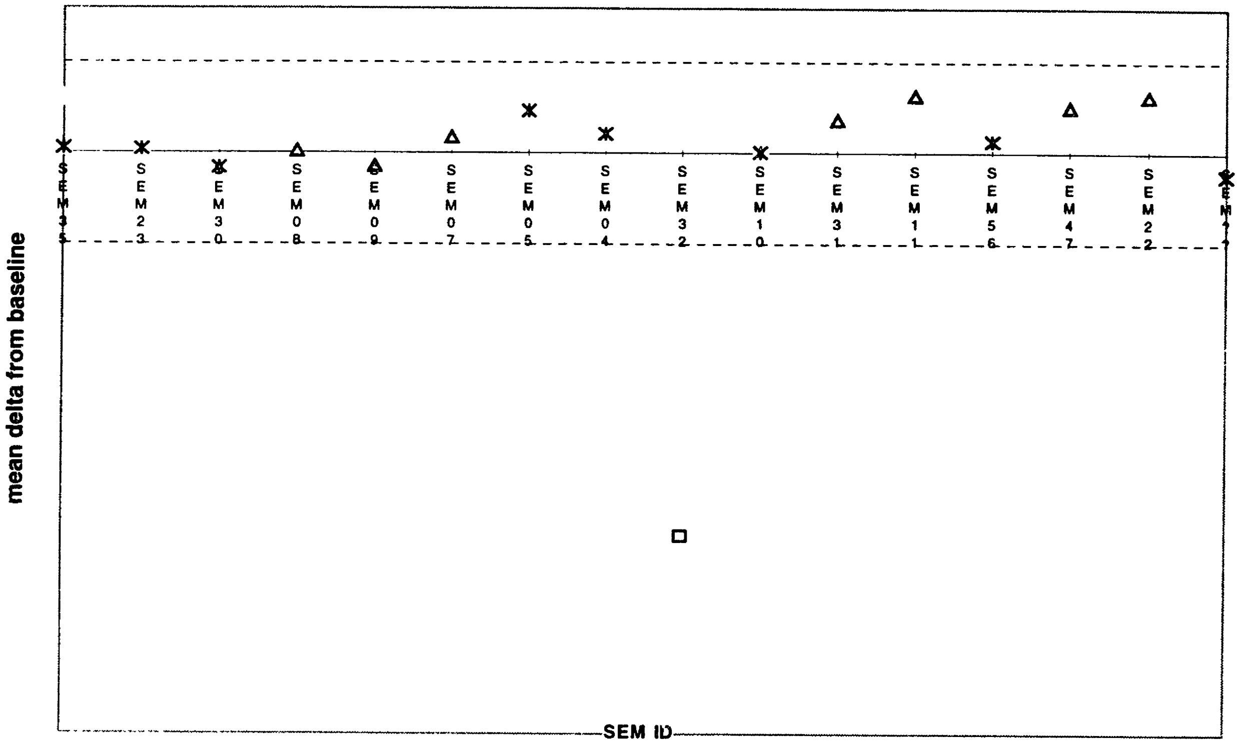


Figure 13 Automated system data for via layer measurements in January.

7.6 Qualitative Results and Observations

While the data from the system are rather ambiguous there were also interesting qualitative results.

Attention to Measurement Quality

It is clear that the automated system pilot generated a significant amount of conversation, among MTs as well as between MTs and engineering, around the correct measurement of final check contact layers (and also to a lesser degree of final check via layers). It some ways it appeared to "legitimize" being unsure about this layer, asking questions, and looking more carefully at the picture book. The automated system also appeared to raise awareness of the importance of high quality measurements generally through attention to the contact and via layers.

Surveys

In the beginning of March, surveys were handed out to as many operators as possible who had used the automated system. Only about a dozen were returned but they are of interest and summarized results will be reviewed briefly. A sample survey is attached as Appendix A. Slightly more than half of respondents felt that the system helped either their own measurements or the measurements of others to be more accurate (both contact and via). Although several respondents suggested that the system would also be useful for other layers, the overwhelming feeling was that the system should not be continued. Many of those who thought it helpful also felt that its purpose had been served and that it was no longer needed, or that it was too time consuming and/or too distracting to production. In general operators liked the anonymous nature of the system, although a few thought it didn't matter.

A few comments that resonate particularly with my experiences follow.

- Do you think you make more accurate [contact] measurements now? **"NO – The picture books work just fine."** Do you think we should continue to use the system? **"No, just make sure people know where the picture book is located."**

In many ways I agree with those comments. I think that if the picture books were very clear and were used frequently and rigorously there would be no training need for the system. At the time, however, measurements indicated that the picture books as used so far had not been sufficient to ensure consistent high quality measurements. In addition, the picture book does not, and can not, provide any indication of the state of the measurement system.

- Do you think we should continue to use the system? **"No, because I think once is enough to get it right."**

In principle I also agree with this, especially if operators take particular care. In fact, I had once thought that after the first time through for each operator there might be no incorrect measurements. Unfortunately the automated system data clearly indicated that this was not the case, probably because the MTs have so many different layers to measure in the course of an average day or week that a single measurement of one layer once a month is not sufficient to change behavior.

- Any other comments and suggestions? **"Let trainers give training updates and additional training on a case by case basis."**

In many ways I think that this comment is a good representation of the prevailing attitude, and was in fact the first approach taken in response to the preliminary study data. As described previously, this approach was not successful in improving measurement quality. Further, I do not think it can work within the current system because there is no way to know who needs more training and no established, effective way to communicate updates.

February measurements

A few operators made measurements in February, on both contact and via layers, although the pilot was clearly finished. This suggests that even if the system is not continued, its availability for use when desired might be a good training tool and help to encourage interest and measurement quality.

Measurement Integrity

As a result of the particular structure of the automation system, it was in fact possible for operators to enter measurement values by hand (type them in), as opposed to having the measurements transferred from the SEM. It appears, based on times between measurements and on the existence of raw data files, that each month a couple of operators chose to simply enter measurement values rather than take the measurements. No action was taken because it appeared to be a fairly isolated problem (frequency stayed quite low), because the instructions had no specific words saying "do not make up values," because I felt that elevating the measurement integrity question to the entire group would be much more negative (highlighting ways around) than positive, and because it was felt to be important to downplay pressure and to increase participation as much as possible.

7.7 Considerations for Formal Adoption of the Automated System

It was agreed at the start that the automated system would have a three-month pilot in Fab 9 followed by a decision about whether and how to implement it officially (if so, would be implementation in both Fab 7 and Fab 9). This decision rests on the costs of not having the system relative to the costs of implementing it. Although the quality of the process data is compromised by large measurement variability, as described previously, the associated product risk is difficult to quantify with certainty.

In the current case, it appears that the extreme shortage (and therefore extremely high "cost") of automation support availability at the present time, combined with the fact that the system constitutes an extra layer in the measurement system (by not using production data the system represents additional work as well as data that is one step removed from production), will make implementation of the system unfeasible. Further, as the specification window for the contact layer is likely to be broadened in the near future, making it less likely that the process would fall out of specification and making process variability a smaller part of the total variability, there is less urgency surrounding measurement variability. However, it remains that without control over measurements, the excursion risk, in the event of a significant process shift, could be significant. While the automated feedback system appears to have been a good training tool, and may be worthwhile to maintain as a training tool, it would be preferable to alter the production measurement system so that control can be verified and problem areas identified using production data.

8.0 CONCLUSIONS AND RECOMMENDATIONS

This project has shown that there can indeed be significant CDSEM measurement variability and that it is not possible to monitor this variability, or mismeasurement, under the current production system. The Type I, Type II, and other errors that are likely to occur given large measurement variability and/or mismeasurement have also been explored. Given the direct link between measurement variability and control of the process being measured, the clear recommendation, at the broadest level, is that some sort of system be implemented that at the very least allows measurement variability to be monitored if not actively controlled. Without such a capability there can be no confidence in the process data. If the process is so stable that it is determined not to need monitoring at all, then that measurement step, and associated control requirements can be eliminated; however, the contact layer does not appear to meet current criteria for elimination of a measurement step.

While the automated feedback system was useful, it is not ideal because it is an extra system layered on top of production and thus adds support as well as time requirements. A better scenario would be one in which production data were used to accomplish the same purpose. If, for example, SEM operators were explicitly responsible for their measurements, with an operator associated with every measurement, it would be possible to do cumulative probability plots of measurements by layer by operator (as is now done by SEM) in order to easily identify outliers. One possible issue with this approach is whether over a reasonable period of time each operator measures each layer frequently enough to make such a chart statistically valid. "Concentrating" operators on certain layers to get sufficient data is not feasible both for ergonomic reasons (e.g., if there were fewer total SEM operators) and for throughput reasons (e.g., if each operator measured only, say, one quarter of the total layers). If the data were sufficient, such a system would also involve a significant change in culture because of the increased level of personal responsibility, and particularly as compared to the rest of litho. For the pressure of this responsibility to be reasonable, operators must be comfortable with their measurements, therefore with their training and their ability to get help and support at any time. Skilled operators would probably not find such an arrangement overly taxing. In addition, such a system provides clear incentive to operators to keep the quality of their measurements very high.

Even if a different system is used to confirm that measurements are in control and have sufficiently low variability for the process they are measuring, incentives to SEM operators to bring the distribution of measurements tighter would be very helpful. The incentives need not be large but would mostly serve to bring attention to the goal of having uniform, consistent measurements, and would recognize progress in that regard. Currently, there can be no such incentive because there is no knowledge of the distribution.

In addition to changing the system in large ways such as just suggested, there are also various more modest recommendations that could help to reduce CDSEM variability but do not help monitor that variability. These include the following, all discussed earlier.

- Reduce the number of trainers (to perhaps two per shift) so there is less variability one level above the operator level and to minimize training costs;
- Have trainers meet periodically so the expertise is uniform within and across shifts;
- Clarify the training structure and responsibilities so there can be more efficient information flows;
- Have a reference at each SEM, preferably an illustrative figure that pops up automatically, to reinforce correct measurement dimensions and locations since operators have to address a large number of layers and structures;
- Remove the measurement values displayed on the CRT at each site to help reduce the conscious or unconscious tendency to "dial-a-CD" and to build operator confidence in their own measurements;
- Modify the baseline training package to have individual sign-off of each layer during certification;
- Add a periodic recertification to make skills more consistent over time.

Appendix A - Survey on Automated System to SEM Operators

Thanks for participating in the SEM Automated Feedback System. Please take a few minutes to fill out this survey, and return it to the coordinator for your shift. The results of this will be used as part of the determination on whether to keep this system, add it to the baseline, etc.

As a result of using the [contact/via] SEM Anonymous Automated Feedback System –

- Your name (optional)
- Do you think you make more accurate [contact] measurements now?
- Do you think you make more accurate [via] measurements now?
- Do you think other people (and/or the population generally) make more accurate [contact] measurements now?
- Do you think other people make more accurate [via] measurements now?
- Do you feel more comfortable with [contact] measurements now?
- Do you feel more comfortable with [via] measurements now?
- What aspects of the system did you like most?
- What did you like least about the system?
- Do you think we should continue to use the system? Why or why not?
- If it is continued, how would you like to see it change? (More, fewer, or different layers? More or less frequent?)
- Do you think the system works best as an anonymous system (as it has been) or would be better without anonymity?
- Any other comments or suggestions?

References

---

[1] Shewhart, Walter, A., Economic Control of Quality of Manufactured Product, Van Nostrand, 1931; repr. ed., American Society for Quality Control, 1980.
[2] Deming, W. Edwards, Out of the Crisis, Massachusetts Institute of Technology, 1986.
[3] Taylor, Wayne A., Optimization and Variation Reduction in Quality, McGraw-Hill, 1991.
[4] Postek, Michael T. and David C. Joy, Submicrometer Microelectronics Dimensional Metrology: Scanning Electron Microscopy, Journal of Research of the National Bureau of Standards, Volume 92, Number 3, p. 205 (1987).
[5] Reimer, L., Image Formation in Low-Voltage Scanning Electron Microscopy, SPIE – The International Society for Optical Engineering, 1993.
[6] Goldratt, Eliyahu M., The Goal, North River Press, 1992.
[7] Menon, Viju, A Constraint-based Systems Approach to Line Yield Improvement in Semiconductor Wafer Fabrication, MIT Master's Thesis, 1994.
[8] Hogg, Robert B. and Johannes Ledolter, Applied Statistics for Engineers and Physical Scientists, Macmillan, 1992.
[9] Ishikawa, Kaoru, Guide to Quality Control, Nordica International Limited, 1976.
[10] RS/1 is a registered trademark of Bolt Beranek and Newman, Inc., Cambridge, MA.

THESIS PROCESSING SLIP

FIXED FIELD: ill. ______ name ______

index ______ biblio ______

▶ COPIES: [✓ Archives] Aero ~~Dewey~~ Eng Hum

Lindgren Music Rotch [✓ Science]

TITLE VARIES: ▶☐ ______

NAME VARIES: ▶☐ ______

IMPRINT: (COPYRIGHT) ______

▶ COLLATION: 79 p.

▶ ADD: DEGREE: ~~M.S.~~ S.M. ▶ DEPT.: Mgt

SUPERVISORS: ______

NOTES: Copy for Dewey Not Available

cat'r: date:

page:

▶ DEPT: MatSci&E ▶J136, 207

▶ YEAR: 1998 ▶ DEGREE: ~~M.S.~~ S.M

▶ NAME: CORTESI, Elisabetta