Evaluating Automated Wafer Measurement Instruments

Technology Transfer # **94112638A-XFR**
SEMATECH
February 28, 1995

Abstract: This document demonstrates a sequential process of evaluating automated wafer instruments and discusses why this approach is useful for studying instruments that have automation features such as loading and focusing mechanisms. The methodology specifies a series of experiments consisting of two or more capability studies followed by a stability study. Each experiment achieves a separate goal, yet combines with the others in providing information needed to assess the usefulness of the instrument. The Analysis of Variance (ANOVA) method of estimating variance components in a nested model is shown to provide information for comparing sources of variation, for assessing the stability of the variation, and for identifying causes of measurement instability.

Keywords: Measurement Capability Analysis, Automated Measurement Instruments, Gauge Studies, Variance Components Analysis, Nested Model, Analysis of Variance

Approvals: Steven A. Eastman, Author
Paul Tobias, Manager, SMG
Peter Lloyd, Director, MCSM
Eugene Woodall, Technical Information Transfer Team Leader

Table of Contents

1 EXECUTIVE SUMMARY 1
2 INTRODUCTION 1
3 PURPOSE 2
4 BACKGROUND INFORMATION 2
4.1 Definitions of Terms 2
4.2 Why is Measurement Capability Important? 4
4.3 Goals of a Measurement Instrument Evaluation 5
4.4 Implications of Automation 5
5 THE METHOD 7
5.1 Organization 7
5.2 Measurement Capability Studies 8
5.2.1 Objectives 8
5.2.2 Capability Experiment Plans 8
5.2.3 Number of Wafers, Types of Wafers, Sites per Wafer 9
5.2.4 Sampling Plan 9
5.2.5 Duration of Sampling 10
5.2.6 Adjusting the Sample Sizes 10
5.2.7 Data Analysis 10
5.3 Stability Study 10
5.3.1 Objectives 10
5.3.2 Stability Experiment Plan 11
5.3.3 Data Analysis 11
5.3.4 Dealing with Wafers that Change Over Time 12
5.3.5 Dealing with Wafer Breakage During the Stability Study 12
5.3.6 Importance of Multiple Measurements per Day in the Stability Study 13
5.3.7 Importance of Constant Sample Size 13
6 SUMMARY OF BENEFITS OF THE METHOD 13
7 RECOMMENDATIONS 14
8 REFERENCES 14
APPENDIX A Calculations of Variance Components Using ANOVA 15
APPENDIX B Control Chart Calculations 17
APPENDIX C Software For Measurement Capability Analysis 19
APPENDIX D Case Study: Surface Particle Counts 23
APPENDIX E Case Study: Thin Film Measurement 29

List of Figures

Figure 1 Crossed Versus Nested Effects Models 6
Figure 2 RS/1 Data Table for Nested Gauge Capability Study 19
Figure 3 RS/1 Output From #VARGRAPH Procedure 20
Figure 4 RS/1 Output From #NESTED Procedure 20
Figure 5 SAS Program for Nested Gauge Capability Data Analysis 21
Figure 6 Output From SAS Program 22
Figure 7 Linearity of Precision on all Oxide Wafers 24
Figure 8a 6100 Å SiO_2 (0.220 µm PSL Spheres)—Total LPD Trend Chart 25
Figure 8b PSL Spheres LPDs (6100 Å SiO_2, 0.220 µm PSL Spheres) 26
Figure 8c Extrapolated LPDs (6100 Å SiO_2, 0.220 µm PSL Spheres) 26
Figure 8d Area Size (6100 Å SiO_2, 0.220 µm PSL Spheres) 27
Figure 9 Dynamic Range Only (Total LPD X-Bar and s Control Charts) 27
Figure 10 Time Order Plot for First Capability Experiment 30
Figure 11 Data from the Second Capability Experiment 31
Figure 12 Data from Third Capability Experiment 32
Figure 13 Data from One Wafer of the Fourth Capability Experiment 33
Figure 14 Data from One Wafer for the Stability Study 34
Figure 15 X-Bar Control Chart for 90 Å Thermal Oxide 35
Figure 16 S-Control Chart of Standard Deviations by Day for 90 Å Thermal Oxide Wafer 35
Figure 17 Components of Variation for 90 Å Thermal Oxide Day 4 of Stability Study 36
Figure 18 Time Order Plot for Day 4 of 90 Å Thermal Oxide Wafer 36

List of Tables

Table 1 Table of Analysis of Variance Calculations 15
Table 2 Precision, Repeatability, Reproducibility, and CV (Short-Term Study) 24
Table 3 Total LPDs Variance Estimates (Long-Term Study) 28
Table 4 Precision, Repeatability, Reproducibility, and CV (Long-Term Study) 28

Acknowledgements

The author would like to thank Kevin Witt for his participation in applying this methodology to the two case studies included herein.

The material in the case studies was extracted from two SEMATECH confidential technical transfer documents.

The author also thanks Mike Bowen, Kurt Johnson, Lori Ellis, and Curt Engelhard of Intel Corporation for their suggestions. Helpful comments were also made by Patrick Spagon, Arnon Hurwitz, and Karen Horrell of SEMATECH.

1 EXECUTIVE SUMMARY

This document describes a new approach for evaluating automated wafer measurement instruments. Although standard methods of evaluating non-automated measurement systems are readily available, those methods are not well-suited for studying automated instruments.

The key features of the new method are as follows:

1. Two or more capability experiments are conducted, followed by a stability study.
2. Sampling plans incorporate a nested series of measurements: repeated measurements within a load cycle, repeated load cycles per day, for several days.
3. Statistical analysis methods are appropriate to the revised sampling plans.
4. Resources, such as wafers and time, are optimized by the use of a sequential series of experiments.
5. Automation effects on measurement variation can be identified and the stability of the system can be measured.

This document describes the recommended experiments and sampling plans and provides rationale for their use. It also contains two case studies in which the method was successfully applied—one on thin film thickness measurement and the other on surface particle detection.

The method provides enough information to estimate the variance components (repeatability, reproducibility, precision), even when there are systematic problems with the instrument. In addition, the method provides information about accuracy and linearity. The stability study provides a way to measure the consistency of the variance components over time. Another benefit of the method is that it provides more data with which to perform detective work when results are suspicious. Also, it can reduce the total amount of work required, by allocating the right amount of resources for achieving each of the separate goals of a measurement instrument evaluation.

Using the method described in this document, engineering teams at SEMATECH have successfully evaluated wafer measurement instruments for usefulness in research and manufacturing.

2 INTRODUCTION

A standard method of evaluating a measurement instrument uses procedure called a Gauge Repeatability and Reproducibility (R&R) Study. The procedure calls for several operators to measure several parts repeatedly, often over several days or weeks. The resulting data is analyzed to determine how well operators repeat their own results on the same part and how well the measurements are reproduced from one operator to another. The total measurement variation is defined to be precision, which is the combined effect of repeatability and reproducibility.

In the semiconductor manufacturing industry, many wafer measurement instruments are highly automated. A cassette of wafers is loaded on the instrument, and then the wafers are automatically loaded, measured, and unloaded. Even though the automation removes the operator influence on measurement variation, the wafer handling mechanism still causes variation. In addition, there can be variation due to environmental influences that show up as time-of-day and day-to-day variation.

There is a need to study the automated measurement instrument using a different method than that used in classical R&R studies. A SEMATECH technical transfer document[5] describes a proposed generic design for gauge studies on wafers. We will describe how that design has been implemented, by showing the evaluation method and case studies. We will show how we have chosen to use a simplified model, and performed analyses on individual wafers rather than on all wafers together. Each study will demonstrate experiment designs that use nested models, and will show how Analysis of Variance (ANOVA) is used to produce estimates of the variance components.

3 PURPOSE

The purpose of this document is to describe a new approach for evaluating automated wafer measurement instruments, to describe the steps in the method, and to provide case studies showing the benefits of using the method to meet the needs of the semiconductor industry. This document does not discuss techniques for matching metrology instruments. This document should be useful to metrology engineers involved in the evaluation of a new instrument, or in the evaluation of the use of an existing instrument for a new purpose (e.g., new wafer type).

4 BACKGROUND INFORMATION

4.1 Definitions of Terms

Repeatability (σ_{RPT}):

Repeatability is the variation of repeated measurements made on the same object under identical conditions.

For automated wafer measurement instruments, this component is evaluated by repeatedly measuring the wafer without removing it from the instrument between each reading. Repeatability is defined to be the variation in repeated measurements without allowing the influence of loading, unloading, repositioning, focusing, or other such factors. Repeated measurements are to be taken in as identical conditions as possible. The variation is solely due to factors such as the instrument's optics and calculation algorithms. This level of variation is considered to be the "best case" instrument performance. Only a change in the structure of the tool, or in the calculation algorithms, affects repeatability.

Reproducibility (σ_{RPD}):

Reproducibility is the variation that results when measurements are made under different conditions.

For automated wafer measurement instruments, this component is evaluated by unloading, loading, and remeasuring the wafer. This allows the introduction of variation due to positioning the wafer and focusing the optics. Although reproducibility could be defined to be variation between two or more instruments, this discussion is limited to the evaluation of a particular instrument. Reproducibility factors are defined as those factors that affect the ability of an instrument to reproduce its own results after the wafer has been completely removed from the instrument.

Precision (σ_{MS})

The total variation in the measurement system is the precision. This is composed of the repeatability and reproducibility. The mathematical model for the variation is the well-known formula:

$$\sigma_{MS} = \sqrt{\sigma^2_{RPT} + \sigma^2_{RPD}}$$

Variance Components

Precision, repeatability, and reproducibility are referred to as variance components. The variation introduced by day-to-day environmental influences is also considered a variance component, which is estimated in the stability study. Part-to-part variation could also be considered a variance component; however, we are not explicitly interested in measuring that variation in an instrument evaluation.

Variance Components Analysis

Variance Components Analysis is the calculation of variance components and the evaluation of the relative magnitudes of the components.

ANOVA

ANalysis **O**f **VA**riance is a mathematical method of calculating variance components such as repeatability, reproducibility, and precision. Montgomery [4] documents the method of using ANOVA to calculate variance components for a nested model.

Accuracy

Accuracy is the difference between the average of repeated measurements made by the tool and the true (standard) value. Bias is a measure of accuracy and can be considered a calibration offset. This definition is a common but not universal definition, as sometimes accuracy is defined as a combination of precision and bias.

Stability

Stability is the degree to which accuracy and precision remain constant and predictable over time. This is assessed with the use of two control charts: an x-bar chart to assess stability of accuracy, and an s chart to assess stability of precision. Control chart calculations are shown in Appendix B. Instability is detected by applying the standard Western Electric Company (WECO) statistical process control (SPC) rules as described by Wheeler and Chambers [11]. Stability can also be measured numerically by estimating the percent of variation that is due to day-to-day variation.

Linearity

Linearity is the degree to which precision and accuracy vary across the measurement range of process variation. This is assessed by estimating precision and average value for several different wafers and by making a plot of precision versus average value and a plot of accuracy versus average value.

Tolerance (T)

Tolerance is the specification range allowed for a process and is defined as:

T = (Upper Process Limit) - (Lower Process limit).

Precision to Tolerance Ratio (P/T)

The ratio of the total measurement variation to the tolerance allowed by the process specification multiplied by 100%. The formula is:

$$P/T = \frac{100\% \times 6\sigma_{MS}}{T}$$

where T is the tolerance. If the tolerance has been arbitrarily defined, P/T may be meaningless. The signal-to-noise ratio (SNR) may be more helpful in assessing the suitability of the measurement tool. P/T should be less than 30%.

Signal-to-Noise Ratio (SNR)

The ratio of the variation of the manufactured product to the precision of the measurement instrument is called the SNR. This is defined to be the ratio

$$SNR = \frac{\sigma_{product}}{\sigma_{MS}}$$

Generally, since measurements are made on the final product with the measurement instrument, it is difficult to directly measure the standard deviation of the product since it also includes variation of the measurement system. Therefore, the SNR can be calculated using the following formula:

$$SNR = \frac{\sqrt{\sigma^2_{Product} - \sigma^2_{MS}}}{\sigma_{MS}}$$

where $\sigma_{Product}$ is the standard deviation of a large number of product measurements obtained using the measurement instrument. It is desirable to have SNR as large as possible: greater than 10 implies that the instrument can be used to distinguish levels of quality of a product. An instrument with SNR less than 3 or 4 would generally be unsuitable for use.

4.2 Why is Measurement Capability Important?

Process Control

The measurement instrument must be able to detect true variation in a process and provide the ability to decide if the process is out of control. If the variability of a measurement tool is too large compared to process variation, the instrument will not be able to efficiently detect process instability.

Product Disposition

If the measurement instrument cannot adequately discriminate between levels of quality of the product, there is increased risk of discarding acceptable product or passing unacceptable product. The use of accurate and precise measurement instruments reduces the need to use guardbands. Guardbands are adjustments to the pass/fail criteria to allow for measurement error. Guardbands

prevent bad product from reaching the customer, at the expense of holding back acceptable product.

4.3 Goals of a Measurement Instrument Evaluation

A measurement capability study can be designed to meet several goals. The following goals are of primary interest:

To estimate capability and stability:

- Quantify the total measurement variation for a given measurement process
- Identify and quantify significant sources of measurement variability
- Determine the stability of the measurement instrument

To obtain useful information for troubleshooting, improvement, and decision-making:

- Minimize the doubt associated with unusual or unexpected results, such as apparent outliers
- Understand the suitability of the equipment to the specific process
- Establish a baseline for improving the measurement process

To obtain information for correct use of the instrument:

- Aid in the development of a sampling plan for monitoring the instrument during production use
- Aid in the development of a calibration strategy
- Establish guardbands for product disposition

4.4 Implications of Automation

The presence of automated wafer-handling mechanisms in wafer measurement instruments affects how the reproducibility component of measurement error is estimated and interpreted. This topic is explored below.

Treatment of Reproducibility as the Operator Effect

Wheeler [1] defines reproducibility as variation introduced by differences between operators. A common way to estimate the effect is to calculate a reproducibility statistic that measures the variation attributable to use by different operators. Wheeler defines the statistic as being the range of the operator averages, R_o, divided by a table constant d_2. In a later reference [2], Wheeler describes the advantages of using control charts to test for significant fixed effects of different operators.

With automated measurement instruments, there is very minimal direct effect of operator technique. The variation among measurements is due to a multitude of automation factors such as loading mechanisms, positioning, and focusing. With each different instrument, there may be few clues as to how the automation might contribute to variability of measurements. Therefore, unless there is prior information, the best strategy is to merely measure the variation due to loading/unloading and call this the reproducibility effect. If the variation is significant, then there may be data to suggest improvement strategies, possibly using experimentation methods described by Montgomery and Runger [6]. If variation is excessive, then the evaluation may be stopped because instrument’s value may be suspect.

Crossed Model or Nested Model?

The traditional model is the crossed-effect model in which the operators are the same ones in each cycle of the measurements (see Figure 1). Each part is measured repeatedly by the same operators. In the crossed-effect model, the operator-by-part interaction is of interest. The interaction term answers the question, "Is the operator-to-operator difference consistent from part to part?"

For automated measurement instruments, the appropriate model is a nested effects model as described by Montgomery [4]. According to this model, repeats are nested within load cycles and load cycles are nested within days. We have found it useful to separately analyze the data for each wafer. The relative contributions of repeats, cycles, and days are reported for each wafer.

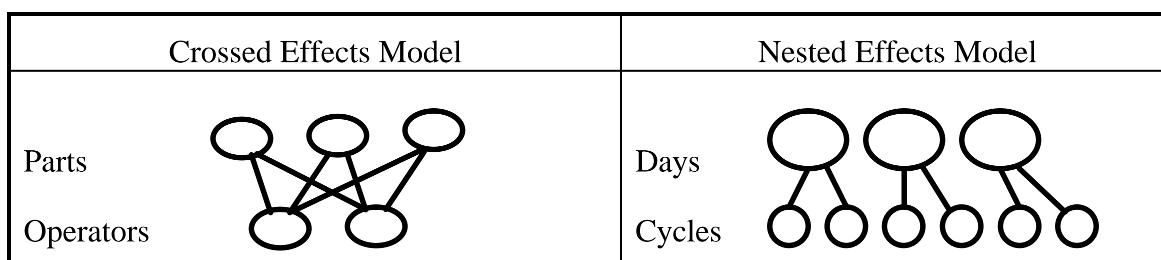

Figure 1 Crossed Versus Nested Effects Models

Analyze using ranges or using ANOVA?

The literature has many references to the use of the "average and range" (A&R) method and the ANOVA method. Wheeler [1,2] documents the average and range method. The Automotive Industry Action Group (AIAG) reference manual [3] describes both methods, and states the advantages and disadvantages of the ANOVA method.

The A&R method is generally considered to be an older, less powerful method than the ANOVA method. Since the data analysis is rarely performed with a hand calculator anymore, it makes sense to use the more powerful ANOVA method with available software. We have chosen to use the ANOVA method here because it produces more accurate estimates of the variances and because software is available to facilitate the calculations for each wafer. Examples of the use of statistical analysis software are given in Appendix C.

Current references on measurement capability [1, 2, 3, 6] describe the use ANOVA model with random effects of parts, operators, interaction between parts and operators, and repeatability. Instead, the nested ANOVA model was used with random effects of days, cycles, and repeatability. The latter model is more appropriate for analyzing the data from the sampling plans. John [5] does show a general nested model that incorporates terms for wafer, placement (load/unload), site, interactions, and repeatability. That generic mixed model has complexity due to the presence of fixed and random effects, and nested and crossed effects. Since wafer and site effects are considered fixed effects unrelated to measurement capability, a reduced nested model was chosen that includes only terms for the days, cycles, and repeats. In the applications used here, automation allows the efficient generation of larger samples, giving enough data for analysis of results for each wafer separately. The variance components were estimated using the general nested

model and ANOVA calculations documented by Montgomery [4] and Beyer [9]. The mathematics is described in Appendix A.

Another Common Strategy

There is another technique in vogue that attempts to address the situation of automated instruments. It calls for a sampling plan as follows:

1. Load the wafer and measure 30 times without unloading. The standard deviation of the measurements is considered to be an estimate of what is referred to as "short term repeatability."
2. Load the wafer, measure it once, unload it. Repeat that cycle 30 times continuously. The standard deviation of the measurements is considered an estimate of reproducibility due to the load/unload effect.
3. Load the wafer, measure it once, unload it. Repeat that cycle once per day for 30 days. The standard deviation of the resulting data points is referred to as "long term repeatability" and is usually interpreted as overall system precision.

There are several problems with such a strategy. One problem is that the reproducibility estimate obtained in step two also has repeatability variation contained in it. It is difficult to justify "subtracting out" the repeatability estimate obtained in the first step since the system may have changed between the steps. Another problem is that it is not easy to identify the cause of unusual data points. Is the strange data due to repeatability, reproducibility, or accuracy problems? Indeed, if unusual data are seen, there is doubt cast upon *all* of the calculated numerical estimates.

The method shown below is one that alleviates some of the problems of evaluating automated measurement instruments.

5 THE METHOD

5.1 Organization

In this method, the measurement instrument evaluation is performed using two or more capability studies followed by a stability study.

Measurement Capability Studies

The purpose of measurement capability studies is to identify and quantify the sources of measurement variation using a series of experiments. These experiments provides data to estimate repeatability, reproducibility, precision, accuracy, and linearity. The experiments are intensive short-term tests of a variety of samples representative of the manufacturing process. The knowledge gained can be used to structure the sampling plan for the stability study.

Measurement Stability Study

If the results of capability studies show that the instrument is capable (i.e., acceptable precision, accuracy, and linearity), then it is possible to go on to perform a stability study. A stability study is used to determine if there are environmental factors that affect how well the instrument maintains that capability over time. It is possible to use the data gathered in the capability studies to establish the number and type of daily measurements required to track the various components of

variation of the measurement tool. *Of central importance to this method is that multiple measurements are made per day so that variation may be estimated for each day.*

5.2 Measurement Capability Studies

5.2.1 Objectives

The goals of the measurement capability studies are as follows:

1. To estimate repeatability, reproducibility, and precision over a short time span
2. To assess linearity of accuracy and precision over several wafer types or thicknesses
3. To decide whether improvements need to be made before proceeding with a stability study

Environmental factors such as humidity and temperature are not expected to influence measurements taken in capability studies but could be influential in stability studies where the sampling is carried out over a longer period of time.

Depending on the goals of the evaluation, it may not be necessary to quantify the specific components of reproducibility. An adequate evaluation can be made by taking enough sample data to allow estimating the total effect of all combined reproducibility components. Factors contributing to reproducibility variation can be individually evaluated later, if needed, using an appropriately designed experiment. See Montgomery and Runger [6] for details.

5.2.2 Capability Experiment Plans

First Capability Experiment Plan

Goal**:** To determine initial rough estimates of repeatability and reproducibility.

Wafers: One or two wafers, representing the most difficult wafer types

Sampling Plan: 2 to 5 load/unload cycles per wafer, with 6 to 15 repeated measurements per cycle. Balance these so as to achieve a total of about 30 measurements per wafer. Examples are: 2 cycles of 15 repeats, 3 cycles of 10 repeats, 4 cycles of 7 repeats, 5 cycles of 6 repeats.

Duration: 1 to 3 Days.

Decision Process: Assess repeatability and accuracy. If acceptable, proceed with second capability experiment. Use the relative magnitudes of repeatability and reproducibility to determine weighting of sample sizes in the next capability experiment.

Second Capability Experiment Plan

Goals: Determine more accurate estimates of repeatability and reproducibility, for a complete range of wafers to be used in production.

Determine if repeatability and reproducibility are linear across the range of measurements.

Wafers:	Many different kinds of wafers, representative of the range of wafer types that will be measured during production. For example: for thickness measurements, choose thick and thin wafers; for particle counts, choose low count and high count wafers. Measure enough sites per wafer to be able to understand how measurement capability differs for different sites.
Sampling Plan:	5 to 20 load/unload cycles, with 3 to 10 repeated measurements per cycle. Aim for a total of about 50 measurements per wafer. Examples: 20 cycles of 3 repeats, 10 cycles of 5 repeats, 5 cycles of 10 repeats. The balance should be based on results of the first experiment.
Duration:	Best if spread over several days.
Decision Process:	Understand whether repeatability and reproducibility are acceptable and comparable for all wafer types. Assess accuracy. If acceptable, continue with stability evaluation. Otherwise, stop the evaluation or perform experiments to isolate and fix the causes of problems. Use estimates of repeatability and reproducibility to determine weighting of sample sizes in stability experiments.

5.2.3 Number of Wafers, Types of Wafers, Sites per Wafer

Since the goal of the first experiment is to understand basic repeatability of the instrument, there is no need to use many different types of wafers. In fact, we suggest using only one wafer each of a few basic types of wafers.

In the second experiment, the goal is more broad—to understand the repeatability and reproducibility of many different types of wafers that might be measured using the instrument. Still, only one of each type of wafer need be used.

Generally, in a capability analysis, much can be learned by measuring a few strategically selected sites. In the wafer thickness case study that follows, one site per wafer was measured. In general though, multiple sites per wafer should be measured. This is to protect against measurement capability inconsistencies across the wafer. Therefore, the data analysis focuses primarily on the variation among repeated measurements on each site. The results for each wafer and wafer site are compared for consistency.

5.2.4 Sampling Plan

One cycle consists of the following steps:

1. Load the wafer.
2. Measure and repeat measuring several times, changing the setup as little as possible.
3. Unload the wafer allowing as much setup change as would be expected in production use.

The objective to is keep the conditions as unchanged as possible inside each cycle, but allow as much variation as possible between each cycle.

Run this cycle two or more times.

The data can be evaluated after the first experiment, in which many repeated measurements are taken to determine basic repeatability within a cycle. If repeatability is immediately seen to be a problem, then the evaluation may be stopped.

After the first experiment is completed, then the number of repeats per cycle can be reduced in the subsequent experiments.

5.2.5 Duration of Sampling

The sampling is performed in a short period of time to minimize the cost of obtaining short-term capability data on all wafer types. Humidity, temperature, and other time-related sources of variation related to reproducibility can be evaluated later in the stability study.

5.2.6 Adjusting the Sample Sizes

It works well to use a sequential process in the selection of the number of measurements taken per cycle and the number of cycles. Start with a small number of wafers, making a larger number of repeated measurements, if little is known about the capability of the instrument. Based on the results, the sample sizes for subsequent experiments on the remaining wafers can be reduced. The number of repeats can be reduced when there is stable and small variation, and number of repeats should be increased when a suspicious pattern or large variation is seen. However, sample sizes should remain relatively constant *within* an experiment.

5.2.7 Data Analysis

The analysis process involves these steps:

1. Plotting the raw data
2. Calculating and tabulating the means and variance components
3. Creating linearity plots
4. Judging the capability

Analysis should be done on each wafer (on each wafer site if multiple sites per wafer are measured), to keep product variation separate from measurement variation. If many wafers and sites per wafer are measured, this can result in a lot of analysis work unless software is employed to perform the repetitive analyses. In the evaluation of a measurement instrument, one goal is to understand linearity of accuracy and linearity of precision. There is generally enough data to accurately estimate instrument capability for each wafer type and even for each site location.

Appendix A shows the ANOVA calculations for estimating the variance components. Appendix C shows how to use RS/1 and SAS software for performing the calculations. A table of variance components can be created that shows results for each wafer or wafer site. See the case studies for detailed examples.

5.3 Stability Study

5.3.1 Objectives

After capability studies have determined the measurement capability of the instrument, there is a need to determine whether the instrument can maintain that capability over time. Examples of factors that might affect the instrument's capability are ambient temperature, humidity, contami-

nation, and drift of the internal electronics of the instrument. In research, as in manufacturing, there is a critical need for a metrology instrument to maintain its capability over time.

5.3.2 Stability Experiment Plan

Goal:

The goal is to determine if repeatability and/or reproducibility change over time.

Choice of Wafers

Use a few representative or challenging wafers. Measure one or a few sites per wafer. It may be reasonable to assume that the level of stability for one wafer type would be the same as that for other wafer types. A representative subset or a challenging subset of wafers can be used in the stability evaluation to determine if any changes occur in the measurement instrument over time. Using a subset of the wafers will substantially reduce the amount of the data collected, easing the burden of computation as well as reducing the amount of time spent acquiring data on a daily basis.

Sampling Plan

There should be a minimum of three cycles per day per wafer, and three repeated measurements per cycle. This provides enough data to adequately estimate reproducibility and repeatability each day. Previous experiments may provide information that suggests increasing these sample sizes.

This sampling plan is necessary to achieve the goal of tracking changes in precision and accuracy over time. If only one measurement is taken per day, there are no estimates of precision and accuracy for that day.

Duration

The stability experiment should consist of least 15 working days, to expose the measurement process to variation-causing environmental effects.

Decision Process

Understand whether repeatability and reproducibility are stable over time. Assess accuracy.

5.3.3 Data Analysis

The data analysis steps for the stability study are as follows:

1. Plotting the raw data
2. Calculating the means and variance components
3. Creating control charts for means and variance components
4. Judging the stability and reviewing the capability

Create a set of charts for each wafer or for each site if there are multiple sites per wafer.

The following four analysis techniques are useful in assessing stability:

Boxplot

This chart shows the distribution of measurements grouped by any grouping variables such as day number, cycle number, or wafer site. The chart will allow the detection of suspicious data that may need to be corrected. Various summary statistics may also be available depending on the software used.

X-bar Chart

This is a chart of the daily average measurements. Control limits can be calculated (see Appendix B), and used to evaluate day-to-day stability.

Calculation of Variance Components

The values are calculated using the ANOVA procedure shown in Appendix A. A table is produced showing the values of repeatability, reproducibility, and precision for each wafer for each day. Values from this table can be used in the s chart mentioned below.

S Chart

This is a chart of the standard deviation of each day's measurements. Control limits can be calculated (see Appendix B) and used to evaluate stability. It is also useful to create a trend chart showing the daily calculated values of repeatability, reproducibility, and precision. See the case studies for an example. A computer program may be needed to avoid an extensive amount of manual calculations.

5.3.4 Dealing with Wafers that Change Over Time

When taking film thickness measurements over time, it is common for thin films to appear to grow thicker due to accumulation of transient surface contamination. With very thin oxides (e.g., 50 Å), the film "grows" quickly with no way to prevent this from happening.

When measuring particle counts, it is common that particles will be added from the environment during the course of the studies.

In both of the above cases, if the data is collected as suggested, the wafer changes can be separated from estimates of precision. The "accumulation" effect will show up as the day-to-day effect in the analysis of the stability data.

5.3.5 Dealing with Wafer Breakage During the Stability Study

There is always a risk of wafer breakage during the study. One of the benefits of multiple measurements per day is that *independent daily estimates* of the components of measurement variation are obtained, not necessarily dependent upon a particular wafer. Therefore, if a wafer is broken during the course of data collection, a similar wafer may be substituted and the experiment carried on. The difference in thickness of the two samples can be observed at the data analysis stage. The stability of precision could still be assessed using the s-control chart. Stability of accuracy could be assessed by looking at each wafer's x-bar chart separately. The underlying assumption, that the response of the measurement instrument is essentially the same for films of the same material having similar thicknesses, can be checked. If extra wafers are prepared in the

same process prior to the start of experiments, then an extra wafer will be available for substitution for a broken wafer.

5.3.6 Importance of Multiple Measurements per Day in the Stability Study

As with the capability study, the stability study step makes use of multiple measurements per wafer. With several cycles each day with multiple measurements per cycle, this method allows for the separation of daily precision into the components of repeatability and reproducibility. When combined with the daily average thickness, this information will be used to assess the stability of precision and accuracy of the instrument.

The investigator will undoubtedly need to balance a number of factors in the stability study: the number of wafers, the number of sites per wafer, the number of cycles and the number of repeats per cycle. The main idea is to design the plan so that it is easier to do the detective work if strange results are seen. With a few repeats per cycle, and a few cycles per wafer per day, there is ample data for identification of problems.

5.3.7 Importance of Constant Sample Size

SAS PROC NESTED produces unbiased estimates of the variance components, even if the sample sizes are not constant (day to day or cycle to cycle). However, PROC NESTED does not produce F statistics for tests of significance in this case, because of the more complicated statistical method involved. Although the F statistics can be useful, they are not absolutely critical. In addition, the computation of control limits is easier if the sample size is constant. To summarize, it is generally a good idea to keep the sample size constant during an experiment.

6 SUMMARY OF BENEFITS OF THE METHOD

Most modern fabs have wafer measurement instruments incorporating automated wafer handling mechanisms. Using the method described in this document, engineering teams at SEMATECH have successfully evaluated such instruments for usefulness in research and manufacturing.

As can be seen from the case studies, the method provides enough information to estimate the variance components (repeatability, reproducibility, precision), even when there are systematic problems with the instrument. In addition, the method provides information about accuracy and linearity. The stability study provides a way to measure the consistency of the variance components over time. Another benefit of the method is that it provides more data with which to perform detective work when results are suspicious. Also, it can reduce the total amount of work required, by allocating the optimal amount of resources for achieving each of the separate goals of a measurement instrument evaluation.

7 RECOMMENDATIONS

It is recommended that software be enhanced to provide the following features:

1. Control charts for variance components in a nested model.
2. Confidence intervals for variance components for a nested model.
3. Statistical comparison of variance components of different instruments to facilitate instrument matching studies.
4. Ability to automatically perform the generation of statistics and graphs for multiple wafers and/or wafer sites.

Although software exists for performing the basic calculations (see Appendix C), these enhancements would make the analysis process much more powerful and much more efficient.

8 REFERENCES

[1] Donald J. Wheeler, and Richard W. Lyday, *Evaluating the Measurement Process*, SPC Press Inc., 2nd Ed, (1989) pp. 35-47.

[2] Donald J. Wheeler, *Problems with Gauge R&R Studies*, 46th Annual Quality Congress, ASQC, (1992) pp. 4-5.

[3] Automotive Industry Action Group (AIAG), *Measurement Systems Analysis Reference Manual*, ASQC (1990), pp. 38-46, 65-70,

[4] Douglas C. Montgomery, *Design of Experiments*, Wiley, 3rd Edition, (1991) pp. 439-450.

[5] Peter M. John, *Alternative Models for Gauge Studies*, SEMATECH Technical Transfer document 93081755A-TR (1993), pp. 41-44.

[6] Douglas C. Montgomery, George C. Runger, *Gauge Capability and Designed Experiments, Parts I and II*, Arizona State University, Publication Numbers 92-11, 92-12.

[7] George A. Milliken, and Dallas E. Johnson, *Analysis of Messy Data* (Vol. 1: Designed Experiments), Reinhold, 1992, Chapter 20.

[8] *SAS/STAT User's Guide, Volume 2*, GLM-VARCOMP, (The NESTED Procedure: Chapter 28), SAS Institute, 1990.

[9] W. H. Beyer, *CRC Handbook of Tables for Probability and Statistics*, 2nd ed., CRC Press, Inc, 1988, pp. 28,31 and 451-454.

[10] Harrison M. Wadsworth, *Handbook of Statistical Methods for Engineers and Scientists*, McGraw-Hill Publishing, 1990.

[11] Donald J. Wheeler, and David S. Chambers, *Understanding Statistical Process Control*, Statistical Process Controls, Inc., Knoxville Tennessee, 1986.

APPENDIX A
Calculations of Variance Components Using ANOVA

A nested model is used to analyze the data for each experiment. Estimates of repeatability, reproducibility and total precision are obtained from the analysis. In the stability study, these estimates can be calculated for each day of the study. Two sources of the information contained herein are: Chapter 13 of Montgomery [4], and page 31 of Beyer [9].

The calculations are best described in the ANOVA table as follows. The notation used is that of Montgomery (except for replacing *y* by *x*):

Table 1 Table of Analysis of Variance Calculations

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	Expected Mean Squares
Day	a-1	$SS_d = bn\sum_{i=1}^{a}(\bar{x}_{i..} - \bar{x}_{...})^2$	$MS_d = \frac{SS_d}{a-1}$	$E(MS_d) = \sigma_e^2 + n\sigma_c^2 + bn\sigma_d^2$
Cycle within Day	a(b-1)	$SS_{c(d)} = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{ij.} - \bar{x}_{i..})^2$	$MS_{c(d)} = \frac{SS_{c(d)}}{a(b-1)}$	$E(MS_{c(d)}) = \sigma_e^2 + n\sigma_c^2$
Error (repeats)	ab(n-1)	$SS_e = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{ij.})^2$	$MS_e = \frac{SS_e}{ab(n-1)}$	$E(MS_e) = \sigma_e^2$
Total	abn-1	$SS_t = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{...})^2$		

The experiment involves:

i=1 to a days

j=1 to b cycles per day

k=1 to n repeats per cycle

x_{ijk} = i^{th} measurement of the jth cycle on the kth day.

$\bar{x}_{ij.}$ = average of n measurements in the jth cycle on the ith day.

$\bar{x}_{i..}$ = average of b x n measurements on the ith day.

$\bar{x}_{...}$ = grand average of a x b x n measurements taken in the experiment

The total variation in the measurements, σ_{MS}^2, is composed of the variance components as follows:

$$\sigma_{MS}^2 = \sigma_d^2 + \sigma_c^2 + \sigma_e^2$$

where σ_e^2 is the variance component for error (repeatability), σ_c^2 is the variance component for cycle (reproducibility), and σ_d^2 is the variance component for day (stability). The estimates of the variance components are calculated in the following way (the carets over the sigmas indicate "estimate of"):

$$repeatability = \hat{\sigma}_{rpt} = \hat{\sigma}_e = \sqrt{MS_e}$$

$$reproducibility = \hat{\sigma}_c = \sqrt{\frac{MS_{c(d)} - MS_e}{n}}$$

$$stability = \hat{\sigma}_d = \sqrt{\frac{MS_d - MS_{c(d)}}{bn}}$$

The precision of the measurement instrument, often called gauge R&R, is calculated from the variance components for repeatability and reproducibility:

$$\begin{aligned} precision &= \hat{\sigma}^2_{precision} \\ &= \hat{\sigma}^2_{MS} \\ &= repeatability + reproducibility \\ &= \sqrt{\hat{\sigma}^2_e + \hat{\sigma}^2_c} \end{aligned}$$

A negative variance component estimate usually indicates a non-significant variance component. If this occurs, the variance component is usually set to zero. The estimates of the variance components will be biased, so the results should be interpreted with caution.

In a capability study carried out during one day, the day component is omitted from the model. In a stability study, the day component is a measure of how much variation is due to the day-to-day influences. Still, precision is defined to be the total contribution of repeatability effects plus load/unload effects.

APPENDIX B
Control Chart Calculations

X-bar (Averages) Control Charts:

"X-bar" refers to an X with a bar over it, as in $\overline{X}$. The bar indicates "average", and X-bar control charts are control charts for the averages of groups of measurements. Refer to Figure 9 for an example on surface particle counts. For *each wafer type*, control charts can be created using the following procedure:

1. Calculate the daily X-bar for the wafer of interest. For example, X-bar could be the average daily particle count.
2. Calculate the average of the daily X-bars, denoted $\overline{\overline{X}}$. The statistic $\overline{\overline{X}}$ is used as the center of the X-bar control chart.
3. Calculate the upper and lower control limits for the X-bar control chart as follows. *This method treats the averages as individual data points, and the control chart is the same as a control chart on individuals.* The reason for using this technique is that the standard deviation of the daily averages is the best statistic for calculating an estimate of the expected in-control day-to-day variation of the averages.

First calculate the standard deviation of the X-bars:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (\overline{X}_i - \overline{\overline{X}})^2}{n-1}}$$

where:

n = number of days

$\overline{X}_i$ = average of measurements for the ith day

$\overline{\overline{X}}$ = average of the daily averages

Then calculate the control limits:

upper control limit (UCL) = $\overline{\overline{X}} + 3s$

lower control limit (LCL) = $\overline{\overline{X}} - 3s$

4. Plot $\overline{X}_i$, UCL, LCL, and center of control chart versus day.

S (Standard Deviation) Control Charts:

S control charts for each wafer type are constructed by the following procedure, from page 452 of Beyer[9].

1. Calculate s_i, the sample standard deviation of all the measurements for day *i*. Do this for each day's measurements.

$$s_i = \sqrt{\frac{\sum_{j=1}^{b}\sum_{k=1}^{n}(X_{ijk} - \overline{X}_i)^2}{bn-1}}$$

where:

n = number of measurements per cycle

b = number of cycles per day

X_{ijk} = k^{th} measurement in the jth cycle on the ith day.

$\overline{\overline{X}}_i$ = average of all measurements on the ith day.

2. Calculate the center line, and control limits for the S control chart as follows:

   Center Line $= c_2' s_p$

   UCL $= B_4' s_p$

   LCL $= B_2' s_p$

   where c_2', B_4' and B_2' are constants obtained from the control chart constant tables in Beyer[9], where the sample size is the total number of measurements in a day.

3. Plot s_i, versus day, and add the lines for the center line and control limits.

APPENDIX C
Software For Measurement Capability Analysis

ANALYSIS USING RS/1 SOFTWARE

Procedures #NESTED and #VARGRAPH were written at SEMATECH in the RPL language of RS/1.

```
# DISPLAY TABLE1
```

TABLE 1 30R x 4C Page 1

Data from RS/1 table #MCA_EX_3

0	SEQUENCE	1 DAY	2 LOAD	3 REPEAT	4 MEASUREMENT
1	1	1	1	1	34
2	2	1	1	2	45
3	3	1	2	1	43
4	4	1	2	2	32
5	5	1	3	1	35
6	6	1	3	2	26
7	7	2	1	1	56
8	8	2	1	2	44
9	9	2	2	1	49
10	10	2	2	2	37

Figure 2 RS/1 Data Table for Nested Gauge Capability Study

This is a partial listing. There are two measurements per load cycle, three load cycles per day over five days. The following graph was produced by the SEMATECH procedure #VARGRAPH.

```
# DISPLAY TABLE1_GRAPH
```

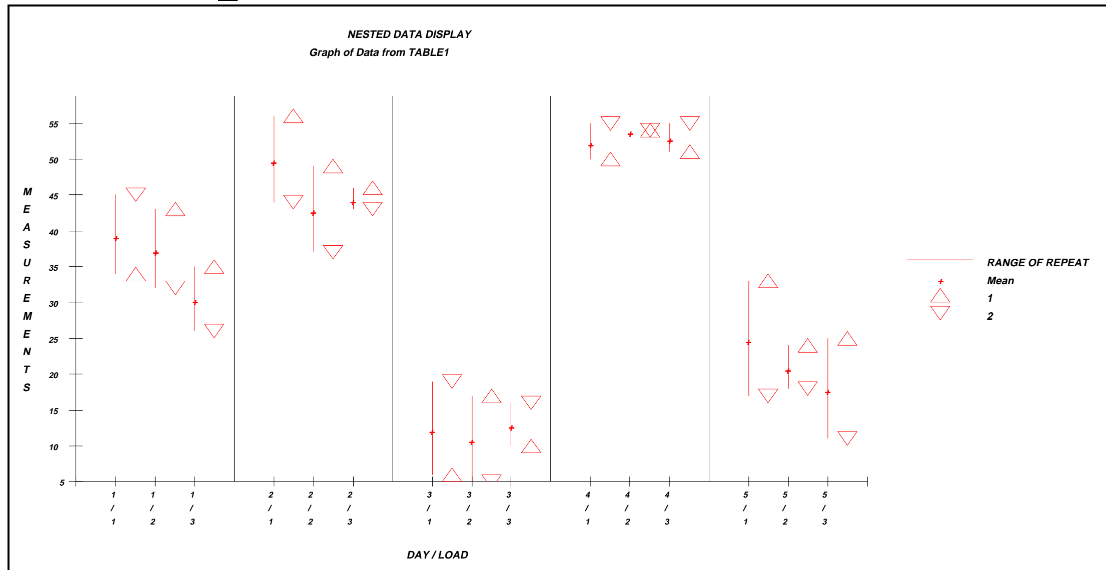

Figure 3 RS/1 Output From #VARGRAPH Procedure

It is apparent from Figure 3 that the day-to-day variation is the dominating effect. Also, the cycle-to-cycle effect is less than the repeatability within cycle. Numerical analysis will now be used to quantify the sources of variation. The RS/1 dialog is

CALL #NESTED

Variance Components from Nested Designs

Tableportion (press enter for help): COLS 1,2,3,4 OF TABLE1

Working...

DOC_NESTED 5R x 5C Page 1

Variance Components Results for MEASUREMENT

from Data in Table @DOC@TABLE1

0	1 Variance Source	2 d.f	3 Standard Deviation	4 Variance Component	5 % of Total Variance
1	DAY	4	16.837623	283.505556	85.0
2	LOAD	10	0.000000	-14.983333	0.0
3	REPEAT	15	7.066352	49.933333	15.0
4					
5	TOTAL	29	18.260309	333.438889	100.0

NOTE: Negative variance components are set to zero for all further calculations.

Figure 4 RS/1 Output From #NESTED Procedure

Since 85% of the variation is due to the day-to-day instability, there are problems with stability for the instrument. The source of the instability should be found and remedied.

Since the load cycle component is negative, the standard deviation for LOAD (reproducibility) is set to zero. It can be assumed that the load cycle effect is not significant compared to the day effect and the repeatability.

Precision is composed primarily of repeatability, since the LOAD effect is assumed to be zero. The value of the repeatability (1 sigma) is 7.1, the rounded value of 7.066352.

For capability studies:

If the capability study is conducted in one day, then the DAY component does not exist. Estimate only repeatability, reproducibility, and precision.

If the capability study spans several days, handle the analysis in the same way as for a stability study.

For stability studies:

Perform the analysis with the DAY component included in the design, and also perform separate analyses for each day, and create a graph of the variance components versus the day number. This process is simplified by using the WHERE clause in the specification of the table portion in the #NESTED procedure. For example, the text to type for estimating the components for day 1 would be:

CALL #NESTED

Variance Components from Nested Designs

Tableportion (press enter for help):

COLS 2,3,4 OF TABLE1 WHERE COL 1=1

ANALYSIS USING SAS SOFTWARE

The SAS program for producing the variance components analysis is as follows:

```
OPTIONS LS=64;
DATA TABLE1;
  INFILE 'TABLE1.DAT';
  INPUT SEQNCE DAY LOAD REPEAT MEAS;
RUN;

PROC NESTED;
     CLASSES DAY LOAD;
     VAR MEAS;
RUN;
```

Figure 5 SAS Program for Nested Gauge Capability Data Analysis

The output from the program is as follows:

Nested Random Effects Analysis of Variance for Variable MEAS

Variance Source	Degrees of Freedom	Sum of Squares	F Value	Pr > F	Error Term
TOTAL	29	7832.666667			
DAY	4	6884.000000	86.194	0.0000	LOAD
LOAD	10	199.666667	0.400	0.9264	ERROR
ERROR	15	749.000000			

Variance Source	Mean Square	Variance Component	Percent of Total
TOTAL	270.091954	333.438889	100.0000
DAY	1721.000000	283.505556	85.0247
LOAD	19.966667	-14.983333	0.0000
ERROR	49.933333	49.933333	14.9753

Mean	33.66666667
Standard error of mean	7.57407860

Figure 6 Output From SAS Program

The variance components report provides the same information as does the RS/1 analysis, except that the standard deviations for each variance component must still be calculated. For repeatability, the estimate of sigma is the square root of 49.9, or about 7.1. For total precision, the estimate of sigma is 7.1 also, since the reproducibility variance component is negative (its sigma is assumed to be zero).

Note that since the SAS procedure allows the use of the BY variable, it is possible to easily obtain multiple analyses. For example, it is reasonable in a multiple day study to set the BY variable to be the DAY variable and thereby receive estimates of repeatability and reproducibility for each day of the study. Another example is the use of the wafer number or site number as the BY variable, resulting in separate analysis for each wafer or site.

APPENDIX D
Case Study: Surface Particle Counts

Source of Case Study

This case study is based on the evaluation of a surface particle counting instrument evaluated in a joint project between SEMATECH and Sandia National Laboratories.

Objectives

The instrument was studied to understand its capability to measure particles greater than .2 microns in diameter on several specific types of wafers.

There was an experimentation phase in which calibration was determined and recipes were developed, After that phase was completed, a short-term capability study and a long-term stability study were performed.

Measurements Taken

A cycle consisted of loading the wafer and measuring it multiple times before unloading the wafer. Measurements consisted of counts of light point defects (LPD's), with counts tallied in bins according to size of defect.

The Capability Study (Short Term)

Objective

The purpose of the first capability experiment was to obtain initial data on precision, repeatability, reproducibility, linearity, and coefficient of variation.

Sampling Plan

Wafers Used:	10 wafers: 2 Bare Silicon, 2 6100 Å Thermal Oxide, 2 4300 Å Thermal Oxide, 4 4300 Å CMP Oxide. Polystyrene latex (PSL) spheres were deposited on some of the wafers.
Sampling Plan:	3 measurements per cycle, 5 load-unload cycles.

Results

Reproducibility variation is very low, as indicated in Table 2, except for two wafers. When measurements were repeated on these wafers, the reproducibility variation dropped down to the same level as the other wafers. The coefficient of variation (CV), which is the ratio of σ_{MS} to the average count, is also reported.

Repeatability variation is higher for polished wafers than for unpolished oxides and bare silicon wafers. The repeatability for PSL wafers is consistent to that for similar non-PSL wafers.

Table 2 Precision, Repeatability, Reproducibility, and CV (Short-Term Study)

Sample #	Wafer Type (slot #)	PSLS Size (µm)	Average LPDs	σ_{MS}	σ_{RPT}	σ_{RPD}	CV (%)
1	Bare Silicon (5)	None	177	4.63	4.63	0.0	2.6
2	Bare Silicon (7)	None	209	7.08	7.08	0.0	3.4
3	6100 Å Thermal Oxide (9)	0.220	2236	5.18	5.18	0.0	0.2
4	6100 Å Thermal Oxide (10)	0.220	2873	8.43	8.43	0.0	0.3
5	4300 Å Thermal Oxide (9)	0.220	4228	17.2	7.4	15.6	0.4
6	4300 Å Thermal Oxide (10)	0.220	2847	12.1	12.1	0.0	0.4
7	4300 Å CMP Oxide (8)	None	2122	16.6	16.6	0.0	0.8
8	4300 Å CMP Oxide (9)	None	2181	15.6	15.6	0.0	0.7
9	4300 Å CMP Oxide (21)	0.269	5429	15.9	15.8	2.0	0.3
10	4300 Å CMP Oxide (22)	0.269	5916	16.2	12.8	10.0	0.1
Repeat Data:							
5	4300 Å Thermal Oxide (9)	0.220	4315	10.5	10.5	0.0	0.2
10	4300 Å CMP Oxide (22)	0.269	5902	15.0	15.0	0.0	0.3

Because the number of particles varied from wafer to wafer, it was possible to assess linearity of precision. The linearity graph is shown in Figure 7.

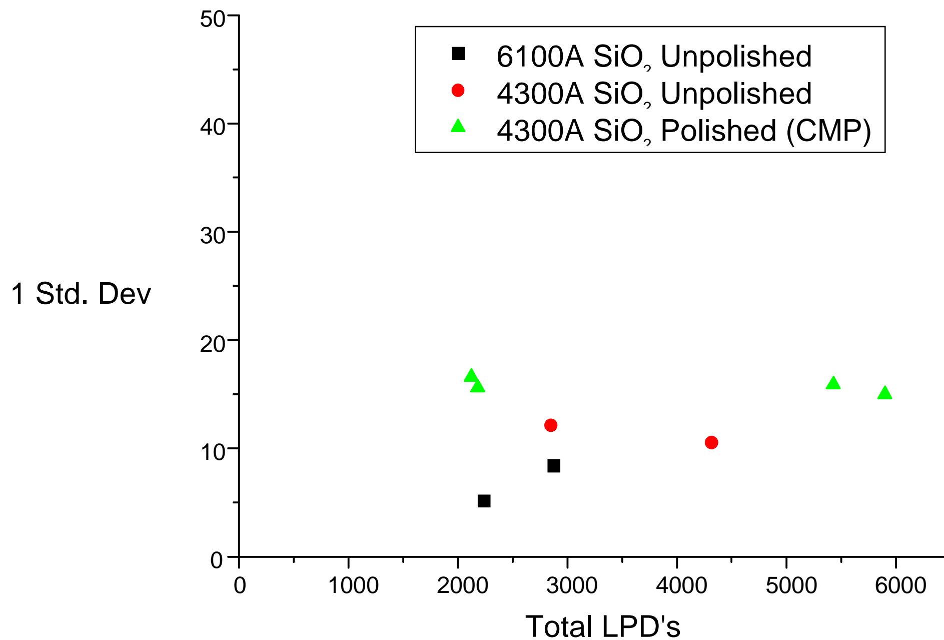

Figure 7 Linearity of Precision on all Oxide Wafers

The Stability Study (Long Term)

Objective

The purpose of the stability study was to determine if the precision and accuracy change over time.

Sampling Plan

Wafers Used:	5 wafers total, 1 wafer of each type that was used in capability experiment.
Number of Days:	15
Number of Cycles:	3 per day, with the order of the wafers varied each day.
Measurements per Cycle	3

Initial Graphical Analyses using Scatterplot Graphs

Trend charts were created, for each wafer type. These were useful for determining stability for each wafer type.

In the case of the 6100 Å SiO_2 wafer with 0.22 µm PSL spheres, the scatterplot in Figure 8a shows that there was a step function change observed in total light point defects (LPD's).

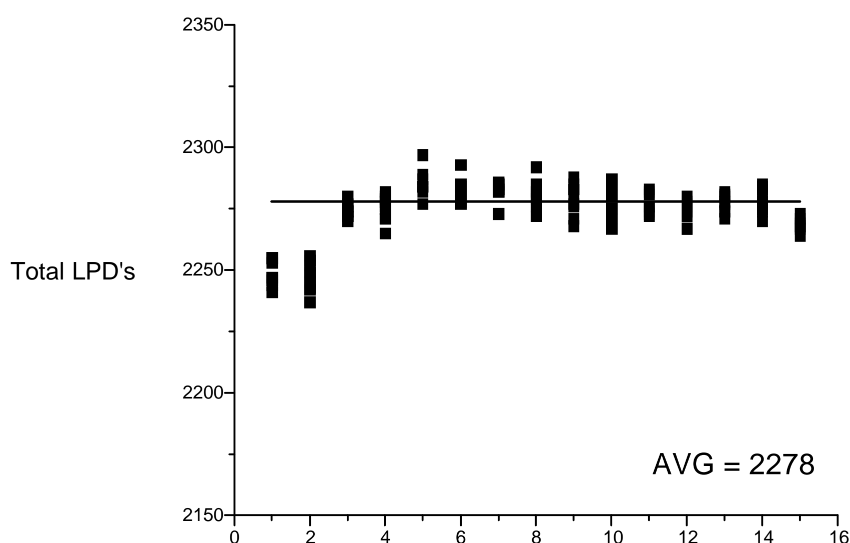

Figure 8a 6100 Å SiO_2 (0.220 µm PSL Spheres)—Total LPD Trend Chart

Since LPD counts are binned according to size of defect, it was useful to create trend charts for different bins. Figures 8b, 8c, and 8d show the trends for PLS spheres, extrapolated LPDs, and area size, respectively. These charts reveal a lower count on day 1 and 2 of extrapolated LPDs, which are included in total LPDs. There is a corresponding increase in area size defects, which are not included in total LPDs. The conclusion was that the drop in total LPDs was due to wafer orientation, which shifted counts from extrapolated LPD bin to the area size bin. This shift could be accounted for by a scratch, with the count being recorded differently depending on the orientation of the wafer.

Because the sampling plan included multiple measurements per day, the investigators found it easy to determine that there was a step function in total counts, with precision remaining relatively the same from day to day.

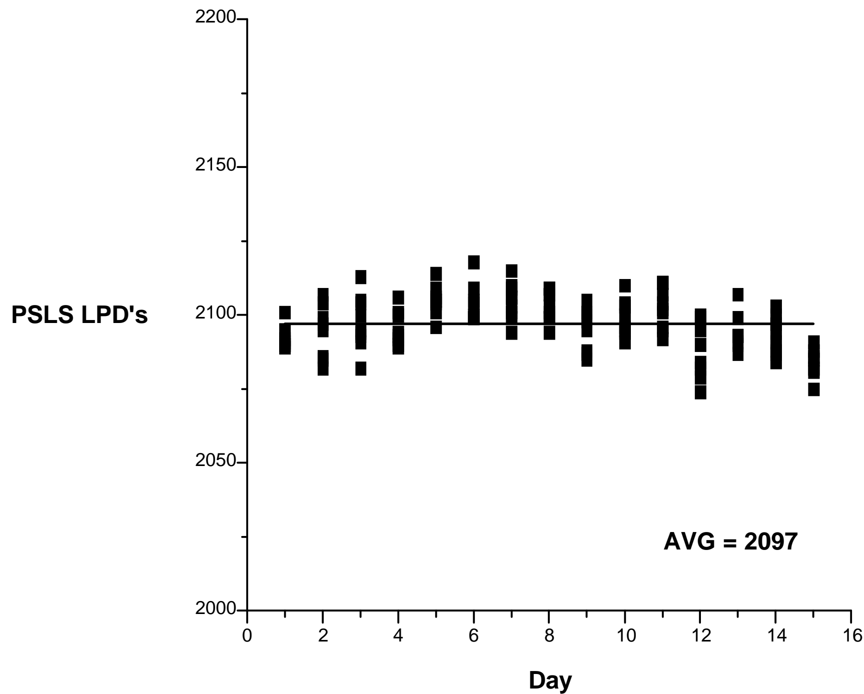

Figure 8b PSL Spheres LPDs (6100 Å SiO_2, 0.220 µm PSL Spheres)

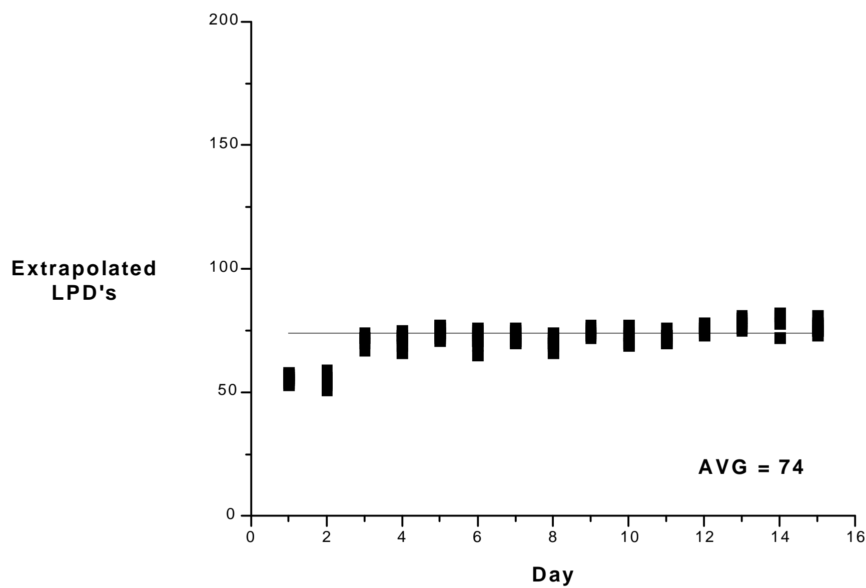

Figure 8c Extrapolated LPDs (6100 Å SiO_2, 0.220 µm PSL Spheres)

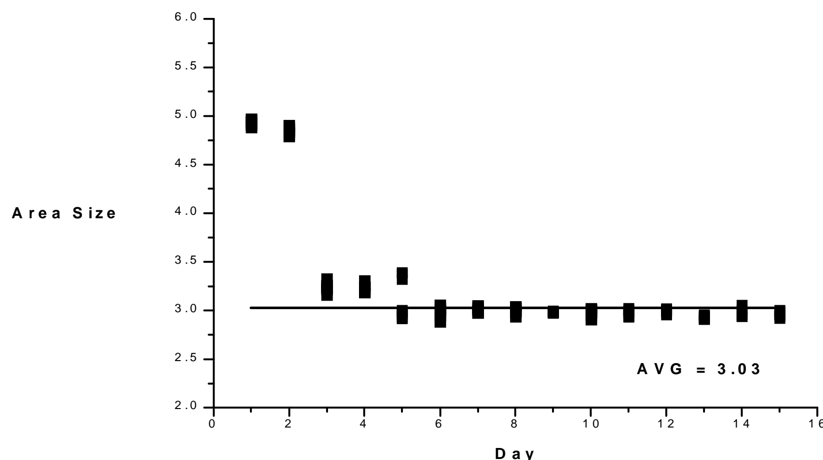

Figure 8d Area Size (6100 Å SiO_2, 0.220 µm PSL Spheres)

Stability Assessment Using Control Charts

After the data was checked using scatterplot graphs, the next step was to check the stability of the variance components using control charts. The control charts demonstrated that the system was stable for all wafer types. A sample control chart is shown in Figure 9. Control limit formulas are given in Appendix B.

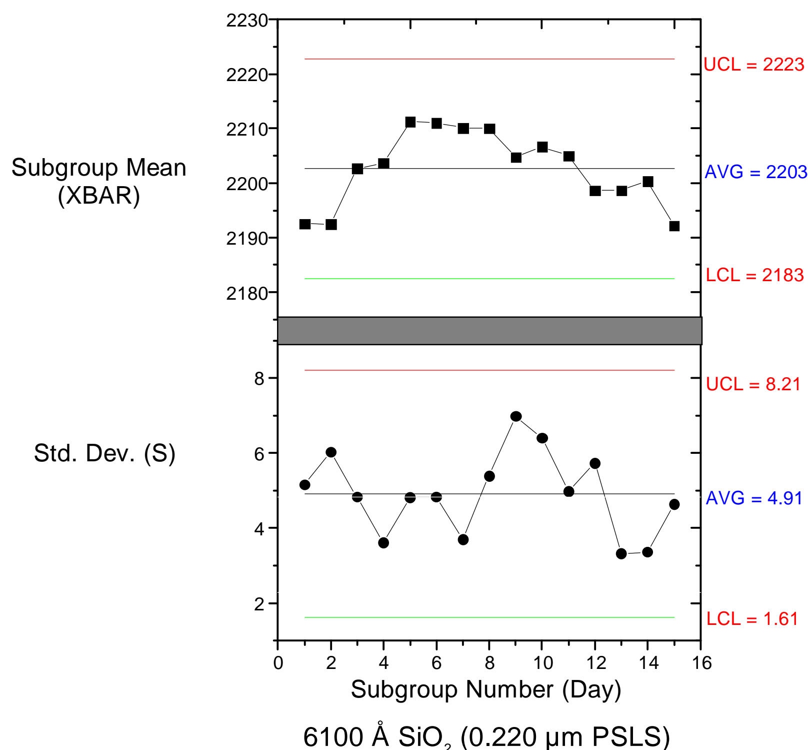

Figure 9 Dynamic Range Only (Total LPD X-Bar and s Control Charts)

Variance Components Analysis

The next step was to summarize the variance components in a table, for easy comparison and reference.

Table 3 contains results of variance components analysis performed for each wafer, calculating the REPEAT, CYCLE, DAY variance components and the total variance. Using these numbers, the sigmas for measurement error (precision), reproducibility (day and cycle combined), and repeatability were calculated and displayed in able 4

Table 3 Total LPDs Variance Estimates (Long-Term Study)

Variance Component	Silicon No PSL Spheres Sample 1	6100 Å** SiO_2 0.220 µm PSL Spheres Sample 3	4300 Å SiO_2, 0.220 µm PSL Spheres Sample 6	4300 Å CMP 0.269 µm PSL Spheres Sample 9	4300 Å CMP No PSL Spheres Sample 7
Rep	51.3 (47)*	22.3 (32)	106.3 (73)	335.7 (76)	136.6 (81)
Cycle	0.0 (0)	4.2 (6)	16.7 (11)	66.9 (15)	10.4 (6)
Day	56.8 (53)	42.9 (62)	22.8 (16)	40.2 (9)	22.1 (13)
Total	108.1 (100)	69.4 (100)	145.8 (100)	442.9 (100)	169.1 (100)

* ( ) % of variance estimate due to the respective variance component for each wafer.
** Variance estimates for 6100 Å SiO_2 wafer are calculated using total LPDs in Bins 1−7 only.

Table 4 Precision, Repeatability, Reproducibility, and CV (Long-Term Study)

Sample #	Wafer Type (Slot #)	PSL Size µm	Average LPDs	σ_{MS}	σ_{RPT}	σ_{RPD}	CV (%)
1	Bare Silicon (5)	None	187	10.4	7.20	7.5	5.6
3	6100 Å Thermal Oxide (9)**	0.220	2203	8.30	4.70	6.90	0.4
6	4300 Å Thermal Oxide (10)	0.220	2834	12.1	10.3	6.30	0.4
7	4300 Å CMP Oxide (8)	None	2125	13.0	11.7	5.70	0.6
9	4300 Å CMP Oxide (21)	0.269	5414	21.0	18.3	10.4	0.4

** Results for 6100 Å SiO_2 wafer are calculated using total LPDs in Bins 1−7 only.

Conclusions

Precision (1 sigma) is less than 1% of total number of defects on the oxide wafers. The instrument appeared to be stable over the 17 days.

There was sufficient data to support detective work that suggested reasons for the identified shift in the measurements for the 6100Å wafer.

Further conclusions addressed the inability of the instrument to detect particles as small as .2 micron, and also discussed the possibility that the instrument is sensitive to scratches on the wafer.

APPENDIX E
Case Study: Thin Film Measurement

Source of Case Study

The information in this case study was extracted from a SEMATECH evaluation of a thin film measurement instrument.

Objectives

The thin film measurement instrument was evaluated for short-term capability and long-term stability. There were five experiments in this evaluation: four capability studies and one stability study. The number of experiments and the sampling plans in each experiment were determined by the unusual results seen as the experimentation progressed.

Measurements Taken

A cycle consists of loading the wafer and measuring one site (in the center of the wafer) multiple times before unloading the wafer. In the second capability experiment, however, the wafer was not unloaded between cycles; instead, the engineer changed the optical objective, causing the instrument to refocus. This was done to determine if the thermal effect was caused by the loading process or by the refocusing process.

First Capability Experiment

The purpose of the first capability experiment was to obtain initial data on repeatability and understand the balance between reproducibility (cycle-to-cycle) and repeatability (within cycle).

Sampling Plan

Wafers Used:	1 wafer, 90 Å oxide on silicon
Number of Cycles:	5
Measurements per Cycle:	30

Results

As can be seen in Figure 10, a strong cyclical trend was observed. Also note that the pattern of measurements in the first cycle is different from succeeding cycles. It is not clear that measurements have stabilized by the end of each cycle. This data led to a hypothesis that wafer load/unload introduces a thermal effect on measurements.

Conclusions

A decision was made to conduct a second capability experiment with more readings per cycle and with only a refocus (not a reload) between cycles.

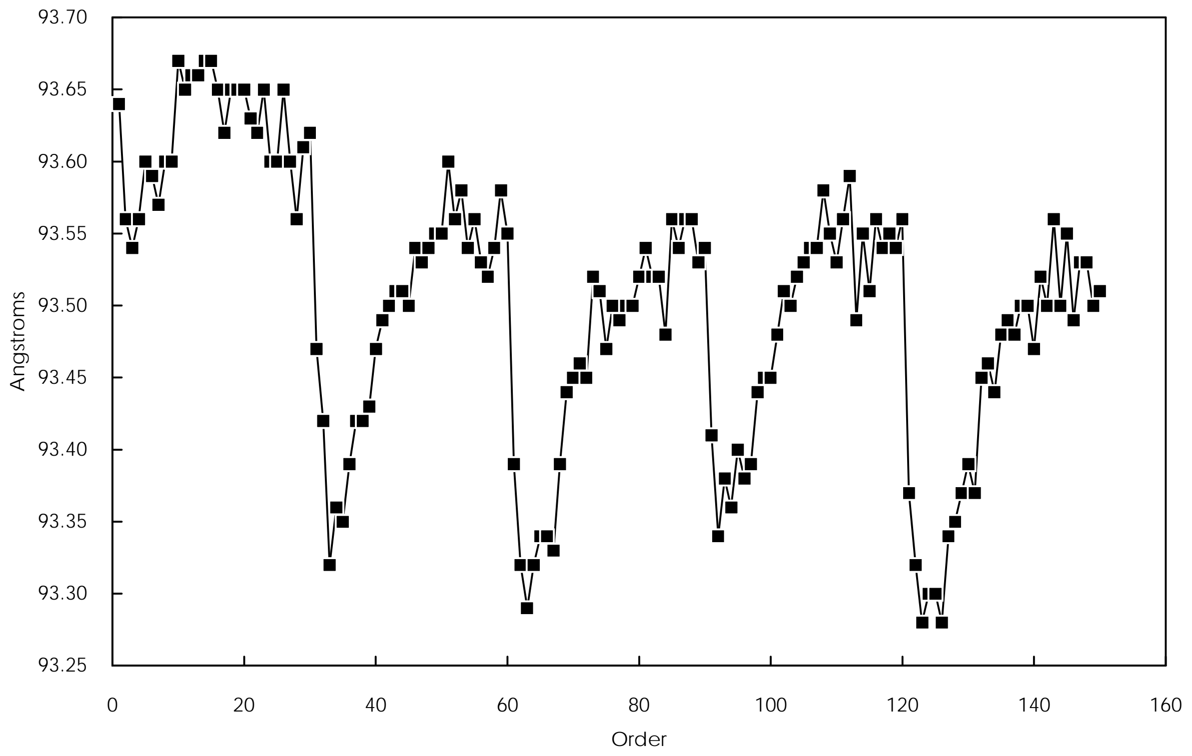

Figure 10 Time Order Plot for First Capability Experiment

Second Capability Experiment

The purpose of this experiment was to understand when the thermal transient stabilizes.

Sampling Plan

Wafers Used	1 wafer, 90 Å oxide on silicon (same wafer as before)
Number of Cycles	3 (refocusing between cycles, but not reloading wafer between cycles)
Measurements per Cycle:	100

Results

Figure 11 shows that, again, the pattern of measurements in the first cycle is different from succeeding cycles. The circles indicate the first measurement taken of each subset of data. It appears that measurement values stabilize after approximately 15 readings (30 sec). The thermal effect is again observed in every cycle. Cycle-to-cycle variation appears to dominate within-cycle measurements during the stabilized portion.

Conclusion

It was decided to conduct a third capability experiment with more cycles to understand cycle-to-cycle variation.

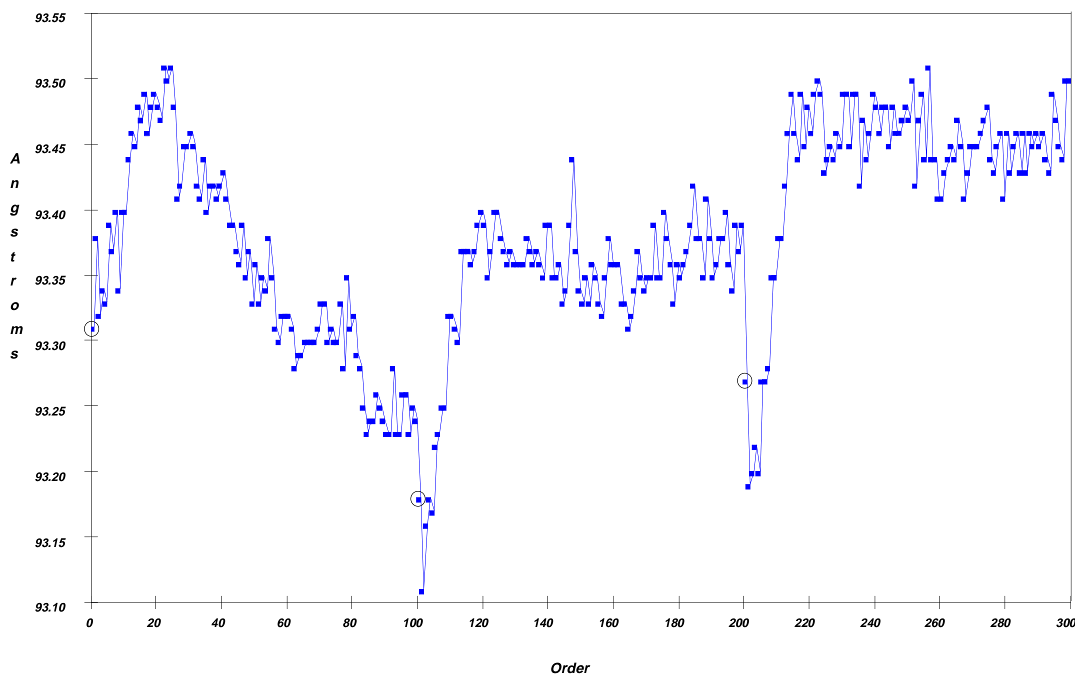

Figure 11 Data from the Second Capability Experiment

Third Capability Experiment

The purpose of this experiment was to obtain better estimate of cycle-to-cycle variation.

Sampling Plan

Wafers Used	1 wafer, 90 Å oxide on silicon (same wafer as before)
Number of Cycles	30
Measurements per Cycle:	15

Results

The pattern of measurements for the first cycle is again different from succeeding cycles. Figure 12 shows that the general upward trend is apparent in the first few cycles. The circled data points are the first readings for each cycle. The suspicious dip in measurements in cycles 6 and 7 is possibly attributable to an automatic calibration done by instrument.

Conclusion

Although interesting patterns of measurements have been seen, the results looked good, so the engineer decided to go ahead and perform capability assessments for a wider range of wafers.

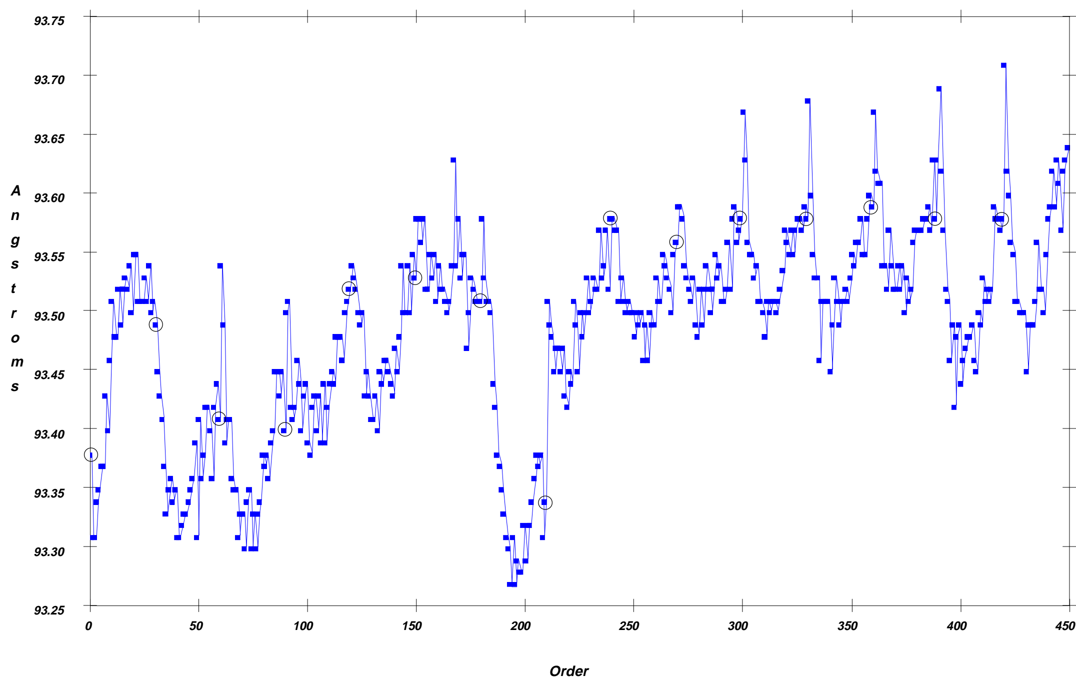

Figure 12 Data from Third Capability Experiment

Fourth Capability Experiment

The purpose of the fourth capability experiment was to estimate capability for a variety of wafers.

Sampling Plan

Wafers Used	10 different wafer types
Types of Measurements	Oxide thickness, poly thickness, top oxide thickness resist thickness, nitride thickness, TEOS thickness (not all done on all wafers)
Number of Cycles	25
Measurements per Cycle:	49 (to continue to monitor the within cycle pattern of variation for various wafer types)

Results

See Figure 13. The same patterns appear, although overall variation is quite small compared to the requirements. No particular problems were seen with any of the wafers.

Conclusion

The decision was made to proceed with the stability study.

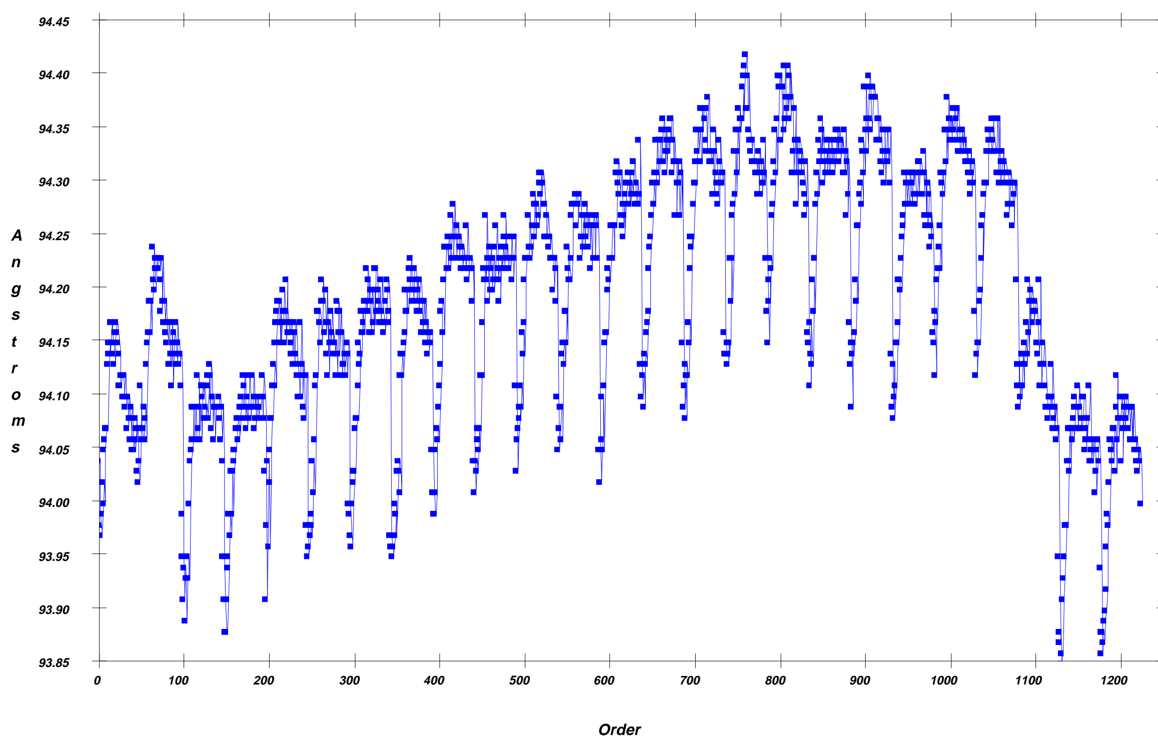

Figure 13 Data from One Wafer of the Fourth Capability Experiment

Stability Study

The purpose of the stability study was to evaluate stability of a representative sample of wafer types.

Sampling Plan:

Wafers Used:	Several representative wafers.
Types of Measurements:	Several measurement types
Number of Days	15
Number of Cycles	3 per day
Measurements per Cycle	15 per cycle

Results:

A set of charts are available for each wafer. Refer to the source document for details.

Cycle-to-cycle variation is the major contributor to overall precision. Day-to-day trend charts of the variance components show that reproducibility fluctuates more than repeatability.

Figure 14 shows a time-order plot that indicates the short-term thermal effects continuing as well as a possible overall upward trend of measurements. Figure 15 shows the x-bar chart for the same wafer, further suggesting the overall upward trend of measurements. Figure 16 is the s chart, indicating a change in precision on day 4, but otherwise stable. Figure 17 is a chart of variance components, with the day 4 data showing that the precision was influenced by cycle-to-cycle differences (reproducibility). Figure 18 is a time-order plot for day 4, confirming a shift between cycles, but relatively stable variation within cycles.

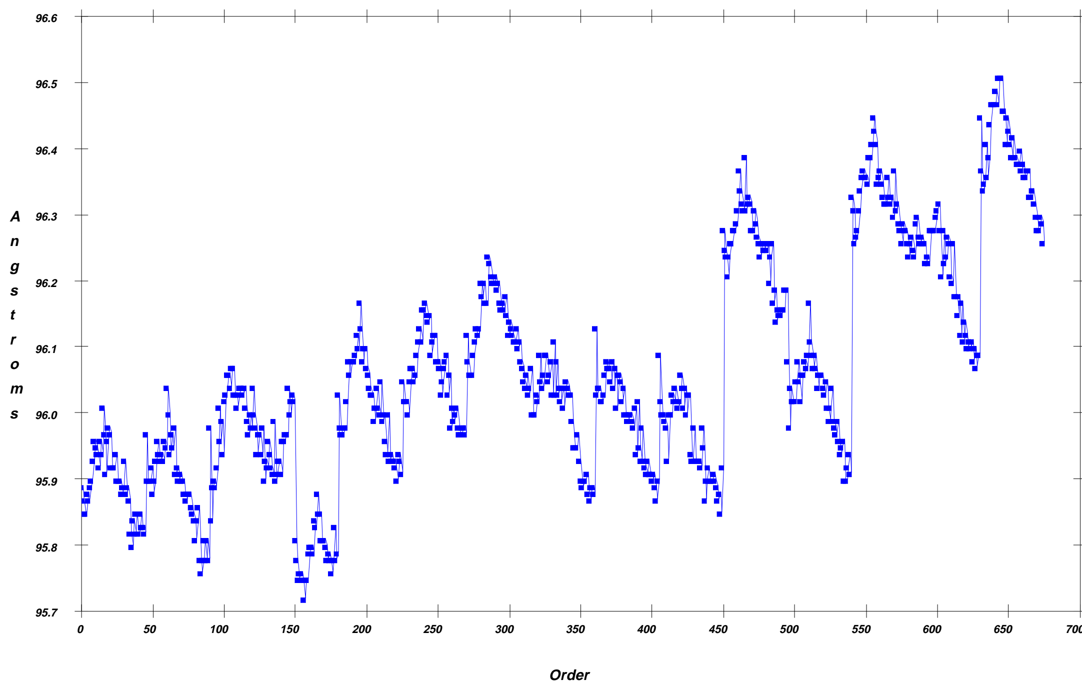

Figure 14　　Data from One Wafer for the Stability Study

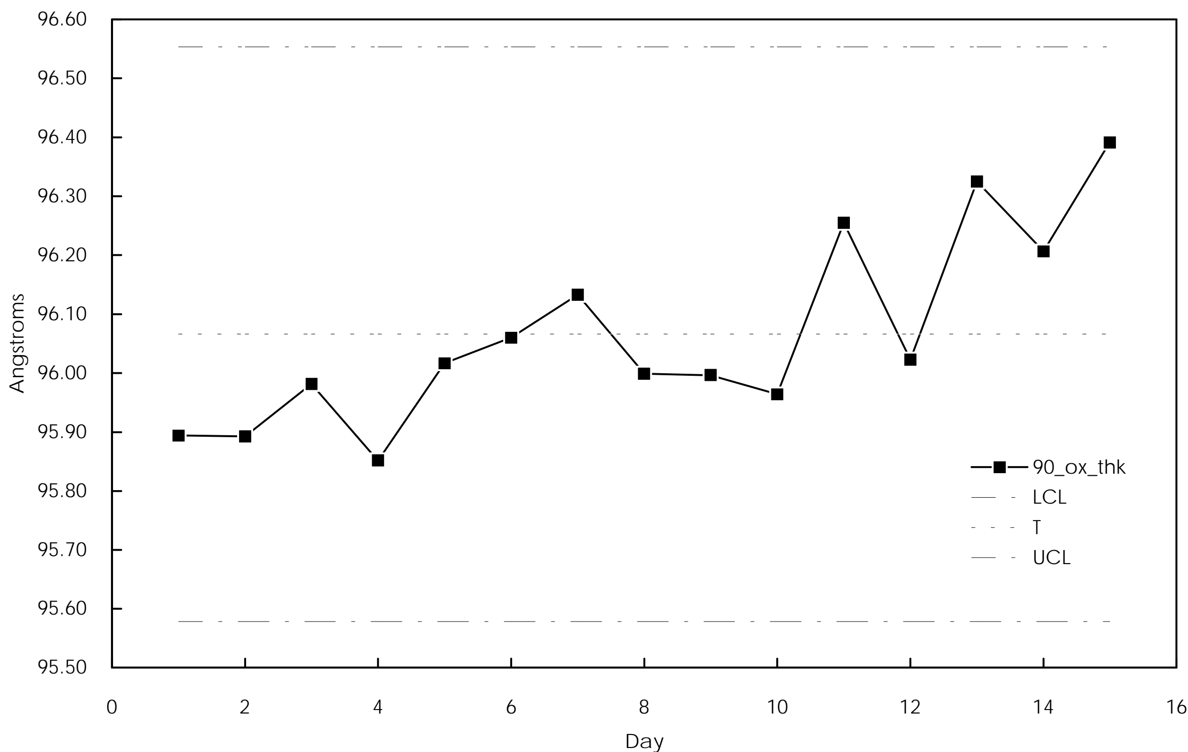

Figure 15 X-Bar Control Chart for 90 Å Thermal Oxide

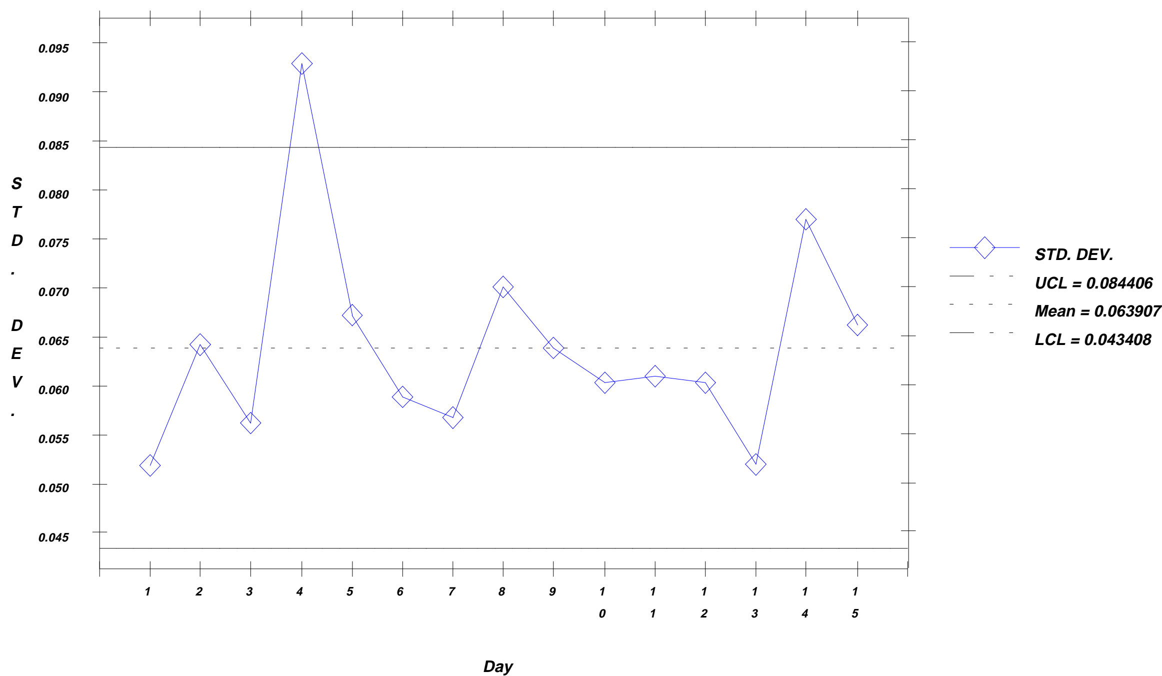

Figure 16 S-Control Chart of Standard Deviations by Day for 90 Å Thermal Oxide Wafer

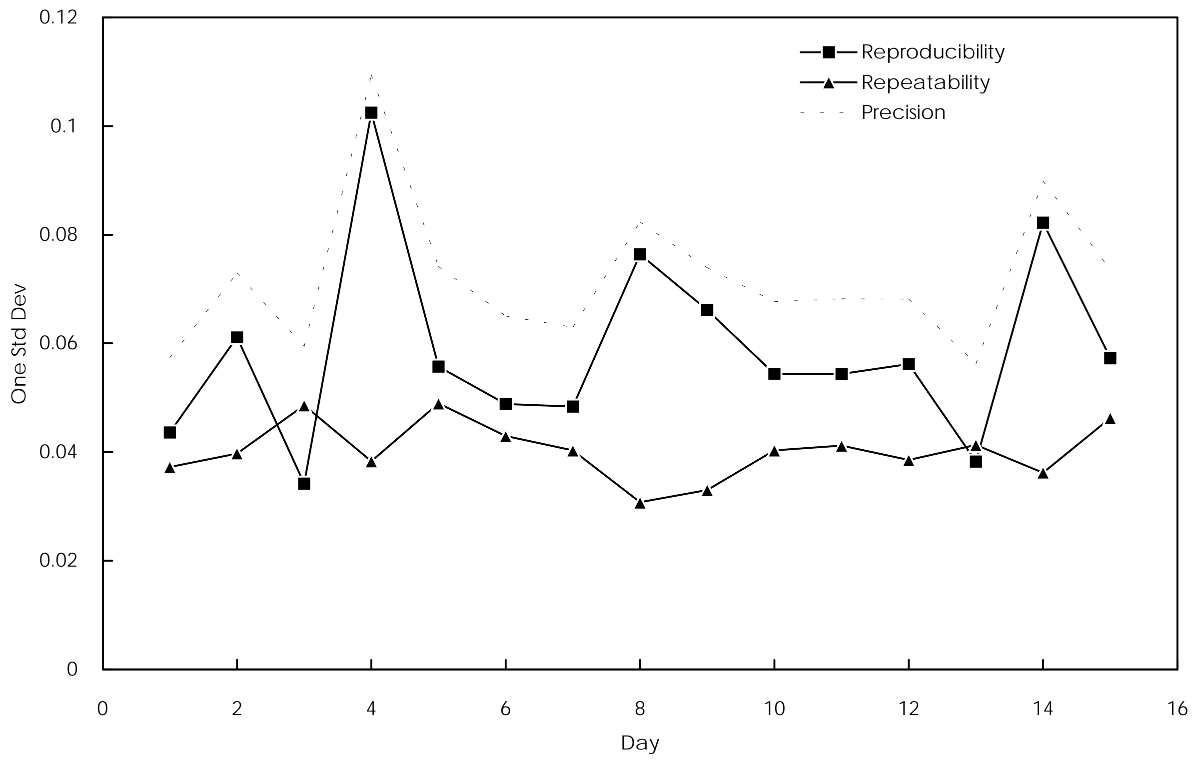

Figure 17 Components of Variation for 90 Å Thermal Oxide Day 4 of Stability Study

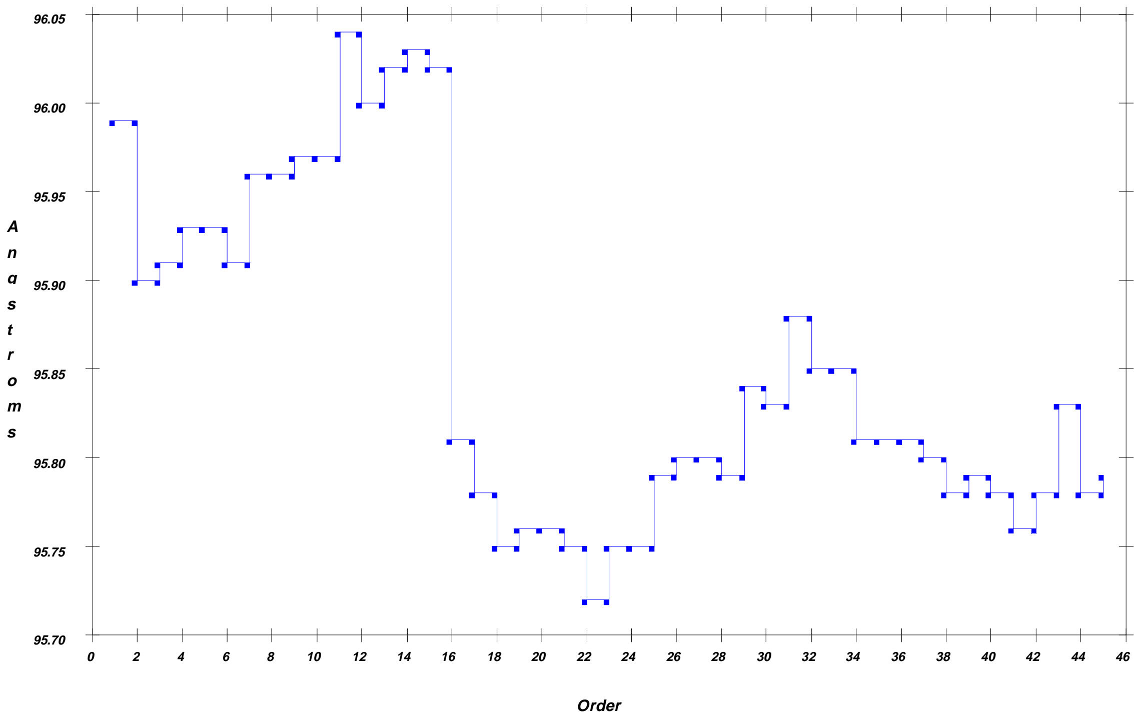

Figure 18 Time Order Plot for Day 4 of 90 Å Thermal Oxide Wafer

Conclusions

The evaluation of the instrument involved four capability experiments, two more than planned, because of the thermal effects seen in the first experiment. Had those effects not been seen, the variation might have appeared to be random and the experiment would have proceeded directly to the multi-wafer capability study.

Improvement strategies were proposed, for eliminating the thermal effects. Overall precision is acceptable considering the application requirements. Accuracy was not assessed due to the unavailability of a suitable set of standard wafers.

The structure of the experiments, with multiple readings per cycle and multiple cycles helped determine the likely sources of the patterns of variation seen in the data.

The stability study provided enough data to determine that suspected instability points were due to cycle-to-cycle effects rather than within-cycle effects.

The experiments provided useful information for improvement of the instrument as well as information useful for planning future capability experiments on this type of instrument.

SEMATECH Technology Transfer
2706 Montopolis Drive
Austin, TX 78741

http://www.sematech.org