Variation Reduction in a Wafer Fabrication Line through Inspection Optimization

by

John W. Bean

B.S. Manufacturing Engineering
Brigham Young University, 1994

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

Master of Science in Material Science and Engineering
and
Master of Science in Management

at the
Massachusetts Institute of Technology
June 1997

©1997 Massachusetts Institute of Technology
All rights reserved

Signature of Author ____________________
Sloan School of Management
Material Science and Engineering
May 9, 1997

Certified by ____________________
Tom Eagar, Department Head
Professor of Material Science and Engineering, Thesis Advisor

Certified by ____________________
Roy Welsch
Professor of Statistics and Management, Thesis Advisor

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

Accepted by ____________________
Jeffrey A. Barks, Associate Dean
Sloan School of Management Masters and Bachelor Programs

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUN 16 1997
Dewey
LIBRARIES

Variation Reduction in a Wafer Fabrication Line through Inspection Optimization

by
John W. Bean

Submitted to the Department of Material Science & Engineering
and the Sloan School of Management on May 9, 1997
in Partial Fulfillment of the Requirements for the Degrees of Master of Science in Material Science & Engineering and Master of Science in Management

ABSTRACT:

Semiconductor manufacturing is a complex and dynamic process. The semiconductor manufacturing process changes as production volumes increase and defect levels vary during the life of the process. One of the limits to output is unexpected defects which lower the yield of the fab until the problem can be identified and resolved. The in-line inspection system has the responsibility of identifying the defect close to its source and reacting appropriately. The location and sampling frequency used to inspect the product must be adjusted as the process changes. The goal of this thesis is to develop a methodology and supporting tools to optimize the allocation of inspection resources in a wafer fabrication line. This approach was developed in conjunction with the Defect Reduction group at Intel's Fab 9 in Albuquerque, New Mexico.

A scenario analysis tool and an integer program were developed to support this analysis. The primary focus of this document will be to explain the general formulation used by both tools. The recommended methodology, which incorporates both tools, will also be explained. A sample analysis is provided to more clearly illustrate this recommended methodology.

Proper in-line inspection optimization can reduce the impact of special cause defects and reduce variation within the fab. The variation reduction in one process line resulted in a projected gain of more than $837,000 each week by optimizing in-line inspections. Similar results would be expected in other process lines and during the lifecycle of each process.

Thesis advisor	Tom Eagar	Professor Material Science and Engineering
Thesis advisor	Roy Welsch	Professor of Statistics and Management

Acknowledgments

The work presented in this thesis was performed under the auspices of the Leaders for Manufacturing Program (LFM), a partnership between MIT and major US manufacturing companies. I am grateful to LFM for the opportunity to associate with my fellow students. My participation in the LFM program has been my goal for over 6 years. I am grateful to be able to graduate and become part of the corps of Manufacturers who will improve the world economy.

I want to thank the Defect Reduction group at Intel's Fab 9 for allowing me to "empty" their brains and being a filter for my insights. Many thanks to Dave Miller, Karen Sibbett, Debra Fenner, Cheryl McLaughlin, Willie Onwo, Dilip Patel, Dina Blair, Cy Hudson and Pei Chiu. Many thanks to Mark Isenberger, my supervisor at Intel, who served as the voice of the customer for this project.

I would like to thank Prof. Steve Graves of the LFM program and Keely Croxton of the MIT Operations Research Center for their assistance in developing a solvable optimization routine.

I would like to thank my MIT thesis advisors, Tom Eagar and Roy Welsch, whose guidance and challenging perspective made this project more valuable to myself and others.

Finally, eternal thanks to my wife, Serene, and my two wonderful boys, Maxwell and Adam, who motivated me to work quickly by making it so nice to come home. Their support has been key to all of my life's decisions.

I dedicate this work to my parents, Brent and Vanita Bean, who made me work hard as a boy so I would be prepared to always work hard.

Table of Contents

Introduction....9
Background....9
Fab Environment....12
Project Goal....16
Model Formulation....17
MAR-WIP....23
MAR-PLAN....25
Optimization....29
Scenario Analysis....31
Rule-based Optimization....33
Model Assumptions....35
Additional uses of the models....37
Example of Inspection Plan Analysis....38
Conclusion....43
Bibliography....44
Appendix....45

List of Figures

Figure 1. General process pattern. 11
Figure 2. The KLA system compares multiple dice and identifies any discrepancies as defects. 13
Figure 3. The Tencor system identifies defects through irregular diffraction patterns.....14
Figure 4. Venn diagram of Comparison of % Defects Captured. 14
Figure 5. Flow of excursionary material and information. 17
Figure 6. Probability of excursion in module. 19
Figure 7. Example of excursion calculation for a module. 20
Figure 8. Probability of either excursionary wafer or die in a module. 21
Figure 9. Typical Yield Histogram. 21
Figure 10. Normal Probability Plot of good die per wafer. 22
Figure 11. Example of effect of excursion cut-off on detection probabilities. 23
Figure 12. Formula for MAR-WIP. 24
Figure 13. Formula for MAR-PLAN. 25
Figure 14. Decision tree of excursion detection. 26
Figure 15. Equation for Pr[looking at an excursionary wafer]. 26
Figure 16. Monitor effectiveness matrix. 28
Figure 17. Probability of recognizing source of excursion as more information is collected. 29
Figure 18. Optimization formulation. 31
Figure 19. Column chart of MAR at each module with proposed sampling plan. 33
Figure 20. MAR chart for current inspection plan. 39
Figure 21. Optimum inspection plan at current production levels. 40
Figure 22. Implementation Options for first modification. 40
Figure 23. Implementations options for fourth modification. 41
Figure 24. The optimum sampling plan when the process has matured. 42

Introduction

The potential profits realized by successful semiconductor manufacturing motivate companies to master this complex and dynamic fabrication process. The process changes as the company strives to correct processing problems and reduce overall defect levels. In-line inspections are used to identify yield-threatening situations. These inspections are performed on a percentage of the product at certain steps in the process. The goal of in-line inspections is to reduce the impact of special cause defects through early identification and appropriate corrective action.

The sampling plan used to inspect the product must be as dynamic as the processing conditions themselves. The goal of this thesis is to develop a methodology and support tools to optimize the allocation of inspection resources as determined by a risk assessment at each step in the process. The model formulation is used by scenario analysis and optimization programs to identify the value-added by in-line inspection and to discover opportunities for improvement in terms of reduced defective material.

The main emphasis of this thesis will be an explanation of the model formulation. The thesis will begin with a background discussion of the fab environment and the types of inspections. The general model formulation will be explained. The value of the scenario analysis program and the optimization routine in maximizing inspection allocation will then be explained. An example of the type of inspection plan analysis will also be provided.

Background

Semiconductor manufacturing has the same basic operations as are used in other industries. In the manufacturing of components for automobiles, material is removed through cutting operations or added through physical assembly or welding. This pattern of assembly and

removal is designed to produce a functional part to be used in the final assembly. In the semiconductor industry, material is removed or added at a molecular level. Films are deposited through spraying of metal molecules and then removed in certain regions by corrosive etchants. This process of addition and removal is designed to produce an integrated circuit which can be assembled in a computer or controller.

The semiconductor manufacturing process is highly complex. Although much is understood about the physics of each step, the molecular scope of the process makes process control and complete understanding quite difficult. This complexity, coupled with high production volumes, makes data analysis difficult as well. Thousands of data points can be collected each day. Out of control points can be a one time occurrence or signals of a catastrophic process failure. In this environment it is essential to prevent large-scale defects while overlooking the defects which affect only a small amount of the product.

A typical process consists of repeating general process patterns. A thin film is deposited before being patterned in a photolithography step. This pattern serves as a mask during etching as regions of the deposited film are removed, or implanting as the conductivity of the underlying layers is altered through ion bombardment. This general pattern -- deposition, patterning and then etching or implanting -- will then be repeated 10 or 20 times before the process is complete. This general repeated pattern is shown in Figure 1. Product may pass through the same equipment many times while in the fab. Hence, if a machine is not operating properly it can affect product at several different stages of the process.

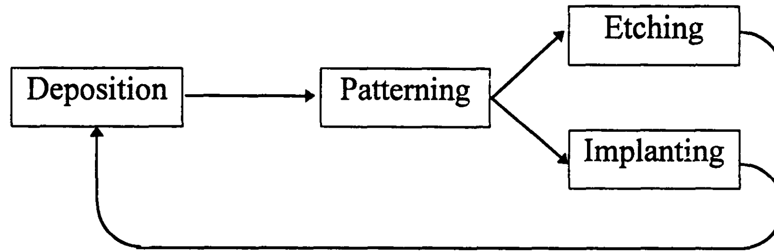


Figure 1. General process pattern.

Output of a wafer fab can be limited by the three factors listed below.

1. Capacity constraints - The high cost of equipment and floor space in this industry rewards companies which can maximize the utilization of all resources. A typical process line will be balanced to maximize the utilization of each piece of equipment. Hence, to increase capacity, a company must make a large investment to purchase equipment for several steps in the process.
2. Normal variation - Each process step has the potential to affect product quality according to its own natural variation. Wafer yields are distributed as some wafers escape defects while others on the lower end of the distribution are significantly impacted. The product affected by this limiting factor is referred to as *baseline.*
3. Sporadic problems - Each process step also has the potential to affect product in a more catastrophic manner. This impact on yield can be assigned to a specific cause which ceases when the fault is corrected. This product affected by large scale problems is referred to as *excursionary*, or product affected by an *excursion.*

A significant monetary commitment is required to reduce capacity constraints and minimize normal variation. Capacity is increased by purchasing equipment and adding floor space -- both of which are capital intensive. Equipment and manufacturing processes are optimized to reduce normal variation. Excursions, on the other hand, are typically caused by a preventable condition which can be corrected through better system design and maintenance.

System improvements are cost effective until a reasonable risk level is achieved. A company must live with some level of excursion risk below which it is uneconomical or impossible to achieve improvement. At this point, the impact of excursions can only be limited by improved detection.

The impact of an excursion depends on its severity and how much product is affected before it is identified and corrected. Excursions can be traced to certain higher risk process steps. These high risk steps include

- metal deposition steps where metal particulates fall off the walls of the chamber,
- photolithography steps where particles under the wafer can prevent it from being properly focused,
- etchant steps where residues remain on the wafer, or
- film deposition steps where faulty process conditions result in layers with electrical properties outside of allowable operating limits.

A lot is made up of 25 or so wafers which are usually processed together. Each wafer contains hundreds of dice which will be separated in the final processing steps. Each excursion has a different impact on the lots, wafers and dice it affects. Some excursions are intermittent while others are consistent. One excursion may impact all wafers but only a few dice per wafer. Another may only impact one out of three wafers but every die on these wafers. The different levels of impact are referred to as the excursion characteristic.

Fab Environment

Effort is taken to eliminate excursions through the use of automation and process control. Most operations are automated to some degree in order to improve consistency by eliminating operator handling. Process control is maintained through the use of statistical control charts to track process and product parameters. A rigorous maintenance schedule is

followed to maintain equipment and to stop excursions caused by equipment failure before they occur. The impact of excursions is limited through detection by in-line inspections.

There are two different types of inspection monitors used to inspect product while in the fab. These monitors are referred to by the company which designed them. KLA and Tencor design equipment use different inspection technology. The system produced by KLA uses optical pattern recognition algorithms to compare multiple dice in order to identify discrepancies. By comparing the pattern from three different dice it is possible to identify particulate or photolithography defects (see Figure 2). These discrepancies, which are viewed as defects, are identified by size and position. The optical algorithm enables this type of monitor to examine patterned product with topography from several process layers. This system distinguishes between the complex patterning of the circuitry and anomalous particles.

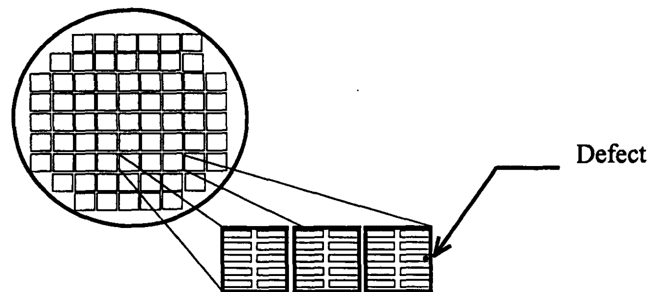


Figure 2. The KLA system compares multiple dice and identifies any discrepancies as defects.

The system produced by Tencor uses laser diffraction to identify the presence of particles. A particle on the surface of the wafer causes the laser light to diffract in a non-uniform manner (see Figure 3). The diffraction patterns are detected by the Tencor system and identified as different size defects. The algorithm and method used by the Tencor system makes it much faster than the KLA system, but not as capable of recognizing defects from previous process layers. For this reason, the Tencor system is generally used after a film deposition step because of the more even surface. The KLA system is used to inspect through film layers or over

different topographies. The KLA is generally more effective at identifying previous defects but requires more time per wafer inspection.

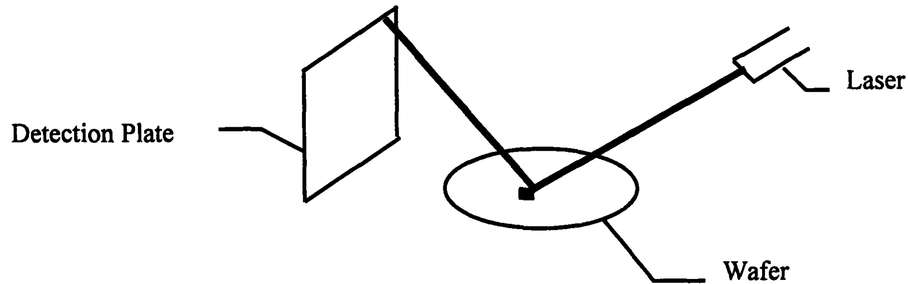
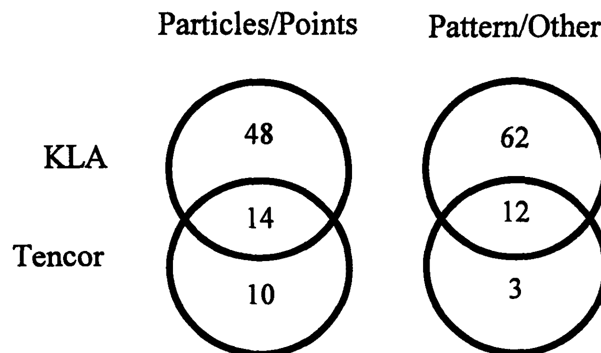


Figure 3. The Tencor system identifies defects through irregular diffraction patterns.

A comparison of the capabilities of these two monitors was conducted by Motorola [1]. According to the article, the KLA system was able to capture 62% of the particle defects and 74% of the pattern defects which could be identified by any current monitoring equipment. The Tencor system could identify 24% and 15% respectively but only 10% and 3% of the defects which were not also identified by the KLA system. A Venn diagram of the capabilities of the monitors is shown in Figure 4. Some defects were identified by only one of the monitors and other defects could be identified by both monitors. (Another monitor, called the OSI system which uses Fourier filtering was also used in the study. This monitor type was not considered in this optimization analysis since it is not used in the wafer fabrication line that was optimized.)



(Values shown are percentage of total observable defects.)

Figure 4. Venn diagram of Comparison of % Defects Captured.

The current procedure calls for a manual review of some fraction of the defects identified with either monitor. Defects are reviewed and classified to recognize shifts in defect trends by type and to better understand the source of most defects. If defect counts are encountered which are above normal limits, the technician notifies the group which has responsibility for excursion control of that process. The yield department has the responsibility to work with the technicians to recognize and control excursions. There is also a process engineering group in charge of the equipment who may be notified if a particular machine is suspect. Manufacturing is notified if the equipment needs maintenance or if production should be halted until the excursion can be rectified.

The product volumes in a typical semiconductor fab limit the percentage of product which can be economically inspected in-line. A "skip-lot" inspection coverage is used to limit the amount of inspection resources required to support production. If 15 lots are tagged for inspection with a 150 lot per week start rate, then the lot coverage is 1:10. It is assumed that the lots and wafers inspected are representative of the product being processed at that time. Research has been conducted to determine if a die can be skipped if defective dice around it indicate it will be defective as well [2], but current procedures call for complete inspection of each wafer.

Product is inspected at the end of the process to determine the final functionality of each die. This end-of-line (EOL) inspection is referred to as sort. Sort data is used to develop yield models and to track trends. Since the yield impact of a defect identified in-line is uncertain, product is usually not classified as an excursion until after it has been sorted. It is a constant effort to correlate in-line information to sort data [3] but the most accurate method of identifying excursions is from EOL data. Product which yields below the forecasted value is investigated to determine if it was affected by an assignable defect. In-line documentation is examined to

identify any probable cause for the excursion. The product may also be examined in the laboratory with a scanning electron microscope in order to identify the excursion source.

Project Goal

The impact of excursions can be reduced through improved allocation of inspection resources. In-line inspection optimization promises significant returns with minimal input and should be considered during process design and throughout the life of a process.

The promising returns from inspection optimization are due to the dynamic nature of semiconductor processes. In-line inspection plans are developed during the initial process development and may or may not be adjusted as the process is modified. Hence, current inspection plans may be artifacts of past processing conditions. An optimized inspection plan would be tailored to current process conditions and production volumes.

There is also a tendency to view inspection plan modifications in isolation. Inspections may be added, moved or removed based on data collected at a single area in the process. This data may indicate the need to modify a specific inspection step, but a holistic view of the process could uncover additional opportunities for improvement.

The dynamic nature of process improvement and the interactions between inspection locations justifies the development of a method for optimizing in-line inspections. This analysis should require minimal data manipulation to enable at least biannual re-calculation. The number of re-design iterations performed each year depends on how often the process changes through improvement efforts, the emergence of different excursions and production volume modifications. The goal is to reduce excursion impact by developing a methodology and supporting tools to optimize allocation of inspection resources.

The impact of excursions is minimized by reducing the time between the beginning of the excursion and its detection. Proper excursion control will consider both the informational

and production loops shown in Figure 5. The solid line represents the flow of product and the dotted line represents the flow of information. The goal of this thesis is to develop the analysis tools needed to design an inspection plan which will reduce the material impacted from the time the excursion starts until its source is identified.

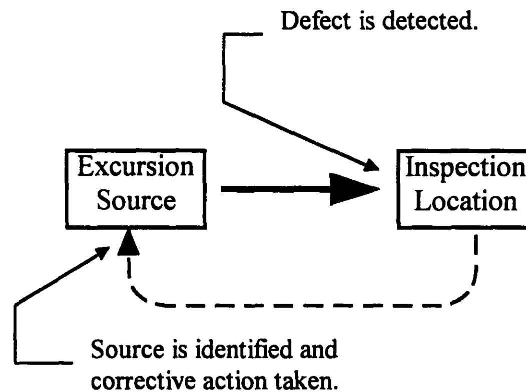


Figure 5. Flow of excursionary material and information.

The analysis tools which were developed include the formulation for an optimization routine and a scenario analysis tool. Both tools are built from the same general formulation which mathematically represents the nature of excursion occurrences and the typical procedures followed by the defect reduction group. This formulation can serve as a model for a similar analysis at other fabs.

Model Formulation

The model formulation begins by breaking the several hundred process steps down into modules based on the most probable inspection locations. Typically, the top 30-50 possible inspection locations are used to divide the process into modules. Process steps are sectioned into modules according to the following guidelines.

- Queue limitations -- Queue limits between steps prevent any additional material handling or inspection. The most common example is the queue restriction following each cleaning

operation. Product must be processed immediately following a cleaning operation to prevent re-contamination of the wafer surface.

- Process limitations -- Processing conditions make certain locations unfavorable for inspection. Photo-resist residue, deposited films or etching operations may reduce the sensitivity of an inspection below acceptable levels.
- Risk assessment -- Certain steps can be grouped together because of low probability of excursion occurrence, Pr[excursion].

A risk assessment is performed on each module to determine the likelihood of an excursion occurring at steps within the module. The assessment should begin with an examination of the relevant historical process data. In a wafer fab, this data can be collected from EOL statistics for each lot and wafer. Ideally, each lot will be grouped into either a baseline or excursionary category according to sort data. Lots are grouped into excursionary categories according to the source of the yield loss. If the yield loss for several lots can be traced to a particular excursion, these lots would be grouped under a title that described the location, source or defect type of that excursion.

In order to determine the risk assessment of each process module, each excursion category is assigned to the process steps which could have been the source for that excursion. For example, if there are three steps which use the same type of equipment, an excursion caused by that equipment which affected 6% of the lots processed during that time frame would be divided into a Pr[excursion] of 2% at each step. It may be appropriate to portion the Pr[excursion] value unevenly if certain steps have a higher risk than others. For example, the 3 steps in the previous example may be assigned Pr[excursion] values of 3%, 2% and 1% if there is a different risk level at each step. The statistical relationship used to determine the Pr[excursion] for each module is shown in Figure 6. In the examples given above,

Pr[excursion]=6%, N_E=3, $F_{E(S)}$=1 in the first part of the example and $F_{E(S)}$ =1.5, 1 and 0.5 when the risk level was not all equal.

$$\Pr[\text{excursion}]_M = \sum_E \frac{\Pr[\text{excursion}]_E}{N_E} \cdot F_{E(S)}, \text{ where } \frac{1}{N_E}\sum_S F_{E(S)} = 1$$

where M = module
E = excursions which could have occurred in module, M
N_E = total number of steps with same potential source of excursion
$F_{E(S)}$ = risk factor assigned to each step with same source of excursion

Figure 6. Probability of excursion in module.

The Pr[excursion] forms the basis for determining the risk assessment of each process module. This probability should be increased if experience and analysis suggest a higher probability than was seen in the sampled historical data. The assigned risk for excursions which were recognized and stopped through the efforts of in-line inspection should be increased from historical data to allow equal comparison with excursions which occurred at steps which are not current inspection locations. The effect of an excursion which impacted past production but which has been eliminated with a systems or process change should be removed from the risk assessment for that module. It is also important to remove excursions which cannot be detected with current inspection equipment. For example, material property excursions such as undesirable electrical conductivity should not be considered when optimizing particulate and pattern defect inspection.

Excursions have different lot, wafer and dice impacts. It is important to accurately characterize the expected excursion impact at each level because of the interaction between inspection coverage and excursion characteristic. Multiple excursions may have the same impact but an excursion which affects a few dice on each wafer will be easier to identify than one which affects many dice on only one wafer. Hence, each excursion is characterized with three variables:

1. Pr[lot] -- the probability that a lot will be excursionary. This value is calculated by determining the ratio of excursionary lots to the total number of lots processed during the relevant time period.
2. Pr[wafer | lot] -- the probability that a wafer will be excursionary given that the lot is excursionary. This value is calculated by determining the ratio of excursionary wafers to the total number of wafers in the excursionary lots.
3. Pr[die | wafer] -- the probability that a die will be excursionary given that the wafer is excursionary. This value is determined by calculating the ratio of actual die yield below baseline average to the total possible die yield for the product.

The values for each excursion must be accurately partitioned to each module which uses equipment or process conditions similar to the step which caused of the excursion. The value of Pr[wafer | lot] and Pr[die | wafer] for each module is the weighted average for each process step in that module. An example calculation is presented in Figure 7 and the statistical formulas are in Figure 8.

	Excursion 1	Excursion 2	Excursion 3	Module
$Pr[lot]_E$	5	4	1	**10**
$Pr[wafer \mid lot]_E$	40	60	30	**47**
$Pr[die \mid wafer]_E$	9	5	15	**8**

Figure 7. Example of excursion calculation for a module.

$$Pr[wafer|lot]_M = \frac{\sum_E Pr[lot]_E \cdot Pr[wafer|lot]_E}{\sum_E Pr[lot]_E}$$

$$Pr[die|wafer]_M = \frac{\sum_E Pr[lot]_E \cdot Pr[die|wafer]_E}{\sum_E Pr[lot]_E}$$

E = excursions in module M

Figure 8. Probability of either excursionary wafer or die in a module.

In order to facilitate data collection, a yield value is selected to distinguish between baseline and excursionary product. The typical shape of wafer yield histogram is shown in Figure 9. This *cut-off* value or yield point can be determined from a probability plot of the yield data. An example of a probability plot can be found in Figure 10. The cut-off value is not necessarily the same for both Pr[lot] and Pr[wafer | lot].

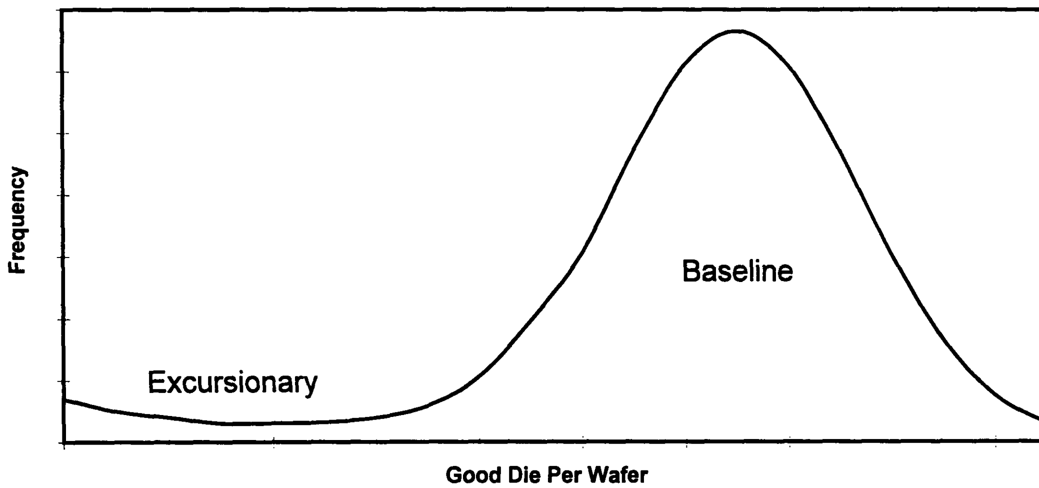
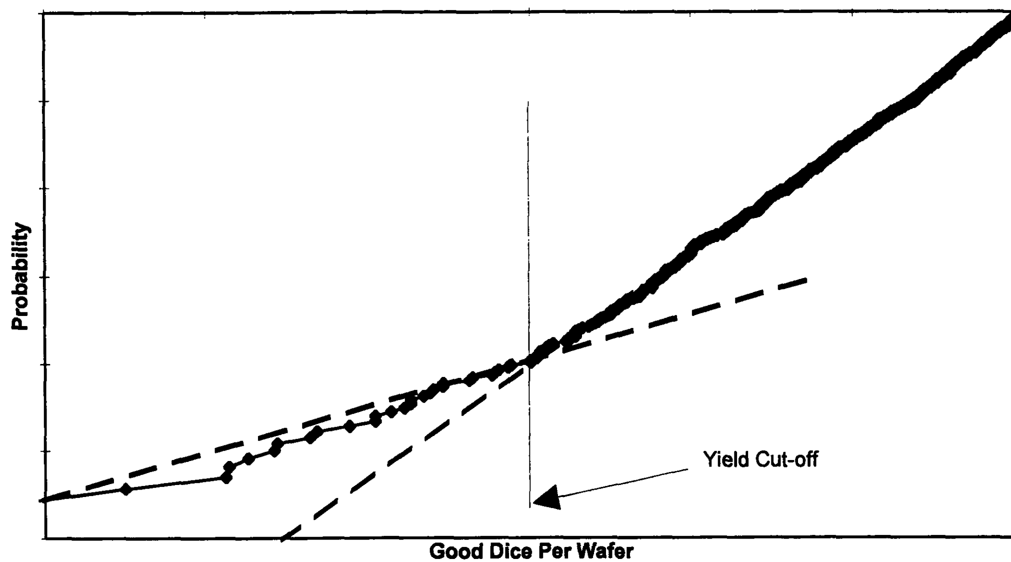


Figure 9. Typical Yield Histogram.



The dotted lines are drawn along the two distributions. The inflection point is selected as the value for the yield cut-off.

Figure 10. Normal Probability Plot of good die per wafer.

If two distinct distributions, baseline and excursionary, are not evident then a yield point must be selected with the following caveat. The value selected will have an effect on the recommended sampling plan. A yield point that is too low will reduce Pr[lot] but increase Pr[wafer | lot] since most wafers in the selected lots will be impacted. This will reduce the influence of the inspection coverage on the total expected defective material because an excursionary lot will differ dramatically from baseline product in terms of its wafer impact. Using the data in Figure 11 as an example, the sensitivity to the cut-off point can be examined. If a yield of 75% is selected as the cut-off, then only 20% of the lots are excursionary but 80% of the wafers are excursionary. With this cut-off value, the wafer level inspection coverage would have little or no influence on the probability of detecting the defect. On the other hand, if 80% is selected as the cut-off then only 60% of the lots and 57% of the wafers are excursionary. This

second cut-off accurately represents the influence of the inspection coverage on the potential impact of an excursion.

	Lot Average	Wafer 1	Wafer 2	Wafer 3	Wafer 4	Wafer 5
Lot 1	83%	71%	86%	80%	97%	78%
Lot 2	86%	98%	89%	100%	58%	87%
Lot 3	79%	81%	92%	98%	55%	69%
Lot 4	82%	88%	82%	90%	54%	96%
Lot 5	91%	88%	96%	97%	84%	91%
Lot 6	77%	83%	90%	60%	81%	73%
Lot 7	76%	55%	87%	89%	71%	81%
Lot 8	80%	84%	60%	80%	78%	97%
Lot 9	71%	60%	74%	69%	97%	53%
Lot 10	66%	51%	74%	70%	77%	57%

cut-off	Pr[lot]	Pr[wafer | lot]
75%	**20%**	**80%**
80%	**60%**	**57%**

Figure 11. Example of effect of excursion cut-off on detection probabilities.

The values for Pr[lot], Pr[wafer | lot] and Pr[die | wafer] can be used to calculate the expected defective material, or material at risk (MAR). The total MAR is a function of the excursion probabilities and the volume of product being processed. The total MAR is the sum of the MAR due to the work in process (WIP) between the excursion source and the inspection step plus the MAR due to product which passed the inspection step because of the plan and monitor effectiveness (PLAN).

MAR-WIP

MAR-WIP is the expected defective material at each step due to the previous process modules. It is a function of the total probability of a defective die and the WIP between a process module and the next subsequent inspection step (see formula in Figure 12). The total probability of a defective die, D_i, is the product of Pr[lot], Pr[wafer | lot] and Pr[die | wafer]. If either the WIP between steps or the probability of a defective die decreases then MAR-WIP will

also decrease. A large separation between defect sources and inspection steps will result in large MAR-WIP values. Hence, the value of this term can be reduced by locating inspection resources close to high risk process steps or by decreasing the excursion probabilities through process improvement.

$$MAR_{WIP} = V_{i,j} \cdot D_i$$

where $V_{i,j}$, = product volume between source i and inspection step j
D_i = Pr[defective die]
= Pr[*lot*] • Pr[*wafer* | *lot*] • Pr[*die* | *wafer*]

Since all wafers are part of the lots being considered and all dice are found on the wafers being considered, the conditional probabilities reduce to a single probability. In other words, a wafer is not considered to be excursionary unless it is part of an excursionary lot.

$$\begin{aligned} \Pr[\text{defective die}] &= \Pr[lot] \bullet \Pr[wafer \mid lot] \bullet \Pr[die \mid wafer] \\ &= \Pr[lot] \cdot \frac{\Pr[wafer \cap lot]}{\Pr[lot]} \cdot \frac{\Pr[die \cap wafer]}{\Pr[wafer]} \\ &= \Pr[lot] \cdot \frac{\Pr[wafer]}{\Pr[lot]} \cdot \frac{\Pr[die]}{\Pr[wafer]} \\ &= \Pr[die] \end{aligned}$$

Figure 12. Formula for MAR-WIP.

The grouping of individual steps into process modules will tend to overestimate the amount of MAR-WIP. This formulation assumes that excursion sources are located at the beginning of each process module which means that all the WIP in that module will be at risk. This becomes more of an issue for modules containing many process steps. If a module has only one or two steps at the end of the module with a significant risk of excursion, then the WIP amounts should be adjusted to remove this overestimate. The scenario analysis tool can be used to determine the impact of this overestimation.

MAR-PLAN

MAR-PLAN is determined by the ability of an inspection to detect an excursion once it has begun. The three events which must occur before an excursion can be detected are listed in the three sections below. The formula for MAR-PLAN, which represents the effects of these three events, in shown in Figure 13. Each term in the formula in Figure 13 will be explained below in the three sections.

$$MAR_{PLAN} = \sum_{n} \Pr[looking] \cdot \exp\{k \cdot n\} \cdot L \cdot D_i$$

where Pr[looking] = probability of looking at excursionary wafers [see Figure 15]
k = ln{1-monitor effectiveness • Pr[looking] }
monitor effectiveness depends on the ranking assigned [see item 3 below]
L = lot coverage, i.e., for 1:12 lots inspected, L=12
D_i = Pr[defective die]

Figure 13. Formula for MAR-PLAN.

<u>Section 1</u> The excursionary material must be inspected.

Since a skip-lot inspection is used, not all product will be inspected. Hence, it is possible for excursionary wafers to "slip" through and avoid detection. The interaction between the excursion characteristic and the inspection coverage affects the probability of actually inspecting one of the excursionary wafers in the lot. The derivation of the Pr[looking at an excursionary wafer] must represent the inspection procedures for the process line being modeled. The wafer fab being modeled for this project has two inspection phases. If only one wafer is found with excessive defect levels in the first inspection phase then a second phase is conducted. Corrective action is taken if more than one excursionary wafer is discovered in the first inspection phase or if one excursionary wafer is found in the first phase and at least one is found in the second phase. This procedure is diagrammed in the decision tree shown in Figure 14. The two branches in a box correspond to the options which result in corrective action. Using the

excursion characteristics, it is possible to calculate the probability of excursion detection using the formula shown below in Figure 15.

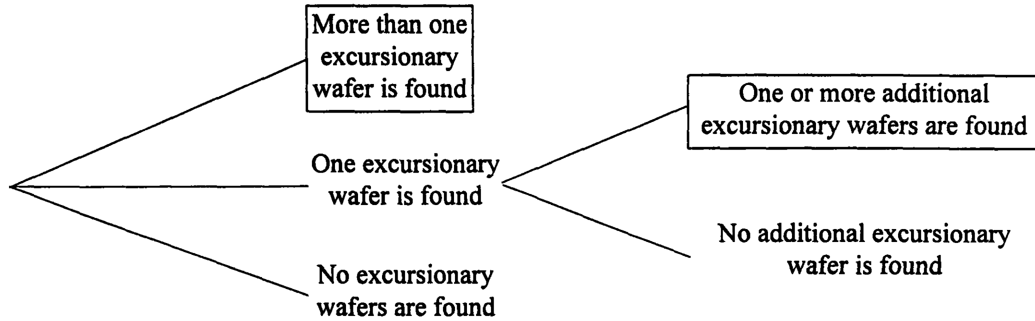


Figure 14. Decision tree of excursion detection.

Pr[*looking*]= Pr[more than one excursionary wafer] + Pr[exactly one excursionary wafer]•
Pr[at least one excursionary wafer]

$$\Pr[looking] = \left[1 - \sum_{y=0}^{1} \binom{W}{y} E_i^{y} (1-E_i)^{W-y}\right] + \left[\binom{W}{1} E_i (1-E_i)^{W-1} \left\{1 - \binom{W_s}{0} (1-E_i)^{W_s}\right\}\right]$$

W = number of wafers inspected per lot
E_i = Pr[wafer | lot] for i^{th} excursion
W_s = number of wafers inspected in second phase of inspection

Figure 15. Equation for Pr[looking at an excursionary wafer].

<u>Section 2</u> The defect, or symptom of the excursion, must be detected.

Before the source can be recognized, the defect must be identified and traced back to its source. The effectiveness of the monitor determines the probability of detecting the defect and its source. In-line process monitors are not 100% effective at detecting defects, especially as the processing delay between the source and the following inspection increases. Typically, a monitor will be highly effective in identifying a defect if located within two or three modules of the source but decrease in effectiveness until it is completely ineffective at about 10 modules away. Monitor effectiveness is also a function of which monitor is used at each location. Each monitor type is better suited to inspecting product at different stages of the process. For

example, the Tencor system, which inspects with laser diffraction, will be most effective following steps where films are deposited and less effective at steps with patterns and topography variations.

The monitor effectiveness must be determined for each process module at each subsequent inspection location. For example, if a metal deposition step is being considered as a possible inspection location, then the effectiveness of the KLA and the Tencor systems at detecting excursions from all previous steps should be determined. An excursion may be undetectable and unrecognizable after only a few steps or it may leave signs which can be seen at several following steps. The manufacturing process itself also affects effectiveness by the deposition or removal of layers which mask or uncover defects.

The values for monitor effectiveness were determined through discussion with the engineers and technicians in the relevant departments. Effectiveness was rated either high, medium, low or zero. The monitor effectiveness depends on factors such as

- Size of particles/defects - Obviously, the larger the particle the more likely the monitor will detect it.
- Type of defect - Information collected on past excursions increases the probability of identifying the source of the defect should it occur again.
- Pattern of defect - The pattern of defective dice on the wafer can indicate the type of phenomena which caused the excursion.

Typically, the monitor effectiveness drops from high to zero within 10 modules for the KLA system and within 5 modules for the Tencor system. An example of the monitor effectiveness matrix is shown in Figure 16. A ranking of the ability of a monitor to recognize an excursion from each previous process module was decided by the defect reduction group.

	Mod 1	Mod 2	Mod 3	Mod 4	Mod 5	Mod 6	Mod 7	Mod 8
Module 1	**H**	**H**	**M**	**M**	**M**	**L**		
Module 2		**H**	**M**	**H**	**M**	**L**	**L**	
Module 3			**H**	**H**	**H**	**L**		
Module 4				**H**	**H**	**M**	**M**	**M**
Module 5					**H**	**L**	**L**	

Figure 16. Monitor effectiveness matrix.

<u>Section 3</u> The source of the defect must be recognized.

The defect will be traced back to its source if enough information can be collected about the excursion. If an excursion only impacts one lot before a routine adjustment of the equipment corrects its source, then it is rare that this low yielding lot can be traced to a particular source. Ironically, this leads the engineers to hope for more defective material so they can identify the source of an excursion. As more information is collected about an excursion, the probability of recognizing its source increases.

A learning curve model best represents the exponentially increasing value of collected information (see Figure 17). Hence, the probability of recognizing an excursion source increases along a exponential curve as more lots are inspected. Each inspection performed on potentially excursionary material helps point to potential sources. The probability of recognizing the source of an excursion starts out at a value which depends on the probability of detecting the defect and the probability of looking at an excursionary wafer. It then increases at an exponential rate depending on the probability of recognizing the source of the defect. For example, at a location far from the excursion source the probability of seeing a defect and the probability of sorting through the many intermediate steps to identify the source of the excursion will both be low. Hence, the probability of recognizing an excursion in this case will start low and increase at a slower rate than for an inspection with a higher monitor effectiveness.

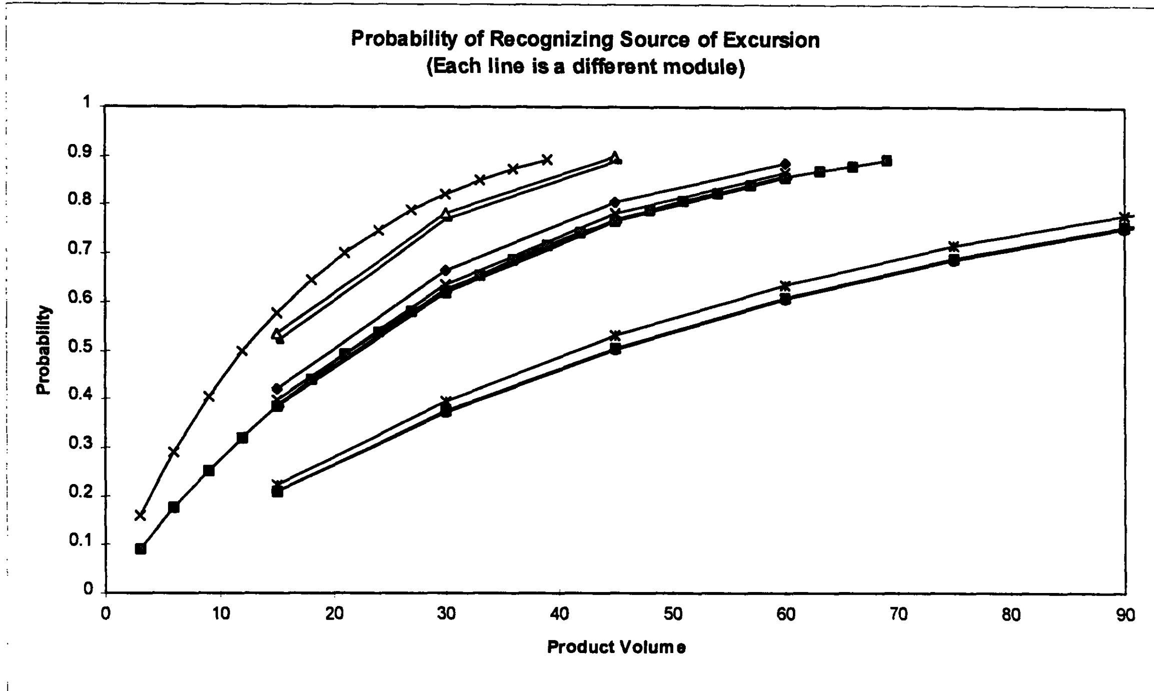


Figure 17. Probability of recognizing source of excursion as more information is collected.

The MAR from each inspection is added until either enough information is collected to assign a source to the defect or enough product has been processed. It is assumed that enough information has been collected to recognize the excursion source when the probability rises above some confidence level. The assumption is that excursions with a probability of detection above that confidence level will be detected. An upper product limit is also used to represent the influence of periodic maintenance and operator intervention on correcting the problem even before the defect is traced to its source. The periodicity of maintenance limits the maximum amount of product which can be impacted by an excursion.

Optimization

This general formulation is used by both methods of solution -- optimization and scenario analysis. Both methods contribute different aspects to the problem solution. The optimization determines the target value to be used with the scenario analysis. The integer

program calculates the minimum MAR possible with a given set of inspection resource constraints. The target value can be used to evaluate the potential return from inspection plan redesign. A comparison of the current inspection plan with the optimum target value will indicate the opportunities for greatest improvement.

The complete optimization formulation is shown below in Figure 18. This includes the equations and a brief explanation of each equation's purpose. The optional constraint should only be used to maintain agreement with the scenario analysis calculation. This constraint removes solutions which allow a module to skip an inspection before being tested, i.e., module 1 is inspected at module 5 while modules 2-4 are inspected at module 4. This formulation requires a large scale solver to accommodate the variables needed for each sampling plan considered. For example, to consider two types of monitors will require 899 integer variables for the 29 module process considered for this project. An answer was achieved in less than 10 seconds while working with a GAMS interface to a CPLEX solver engine. CPLEX is an example of a large scale solver which will solve an integer program of this size. There are other programs which will work equally as well. The input and output for this particular solver are included in the appendix.

i = potential excursion source
j = potential inspection location
s = sampling plan
m = monitor type - different sampling plans use the same monitor
$Y_{j,s}$ = binary variable signifying inspection with plan s at step j
$L_{i,j,s}$ = binary variable signifying inspection of source i at step j with plan s
C_m = maximum number of inspections with monitor m
$M_{i,j,s}$ = material at risk (MAR) from source i if inspected at step j with plan s

$$\sum_{s \forall m} \sum_{j} Y_{j,s} \leq C_m$$

constraint on number of inspection locations for each type of monitor

$L_{i,j,s} \leq Y_{j,s}$ for each i, s and j ($j \geq i$)

assigns $Y_{j,s} = 1$ if plan s is used at step j

$$\sum_{s} Y_{j,s} \leq 1 \text{ for each j}$$

assures only one sampling plan is used at each step

$$\sum_{j>i}\sum_{s} L_{i,j,s} = 1 \text{ for each I}$$

assures each source is inspected at only one location

$$\sum_{a<i}\sum_{b>j}\sum_{s} L_{a,b,s} < (i-1)\cdot\left(1-\sum_{s} L_{i,j,s}\right) \text{ for each i and j} \qquad \text{[optional constraint]}$$

prevents a source from being inspected at a step following a previous inspection location

$$\sum_{i}\sum_{j\geq i}\sum_{s} L_{i,j,s}\cdot M_{i,j,s}$$

objective function for total MAR

Figure 18. Optimization formulation.

Scenario Analysis

To complement the ideal solution provided by the optimization, the scenario analysis tool provides flexibility and additional evaluation capability. The scenario analysis increases the benefit of this approach for several reasons.

- It allows for gradual change toward the ideal. The complexity of the wafer fabrication process makes change control an essential part of maximizing yields. An optimum solution alone does not allow the company to consider the impact of each individual inspection modification. The company will most likely want to make several small adjustments towards the optimum instead of one radical change.
- It allows consideration of the effect of inspection modification costs. Each adjustment of the inspection location or frequency is accompanied by a cost. This includes the cost of documentation and qualification of any adjustments as well as any additional data collection needed to justify the change. Each adjustment also creates complexity and an opportunity

for misprocessing. By allowing stepwise implementation, each improvement can be considered in light of the associated cost.

- It will determine the largest opportunity for improvement. With the use of the scenario analysis tool, it is possible to determine which adjustment toward the optimum will result in the largest improvement. If a company is limited in the number of modifications possible because of complexity, time or cost, the biggest weakness in the current plan can be identified and corrected.
- It reduces the possibility of unexpected shifts in excursion probabilities. A gradual adjustment towards the optimum will allow time for the process to stabilize before each subsequent move. This will reduce the possible impact of excursions not considered in the previous analysis which were controlled by the old inspection plan.

The scenario analysis tool was designed to facilitate a rule-based optimization. This tool is run from a spreadsheet interface by visual basic macros. A copy of the macros can be found in the appendix. The software routine within the analysis program calculates the values of MAR-WIP, MAR-PLAN and MAR-Total for each scenario. The graph in Figure 19 displays a sample of these values. The columns on the bottom are the values of MAR-WIP, -PLAN and -TOTAL. The columns descending from the top of the graph indicate the modules which are followed by an inspection, either with the KLA or the Tencor system. This graphical method allows the user to recognize the weaknesses in each scenario and to formulate steps for improvement. The software routine determines the sensitivity of each solution to changes in production volume, monitor effectiveness and inspection coverage. The number of wafers inspected per week is calculated to allow accurate comparison between inspection plans with different lot and wafer coverage.

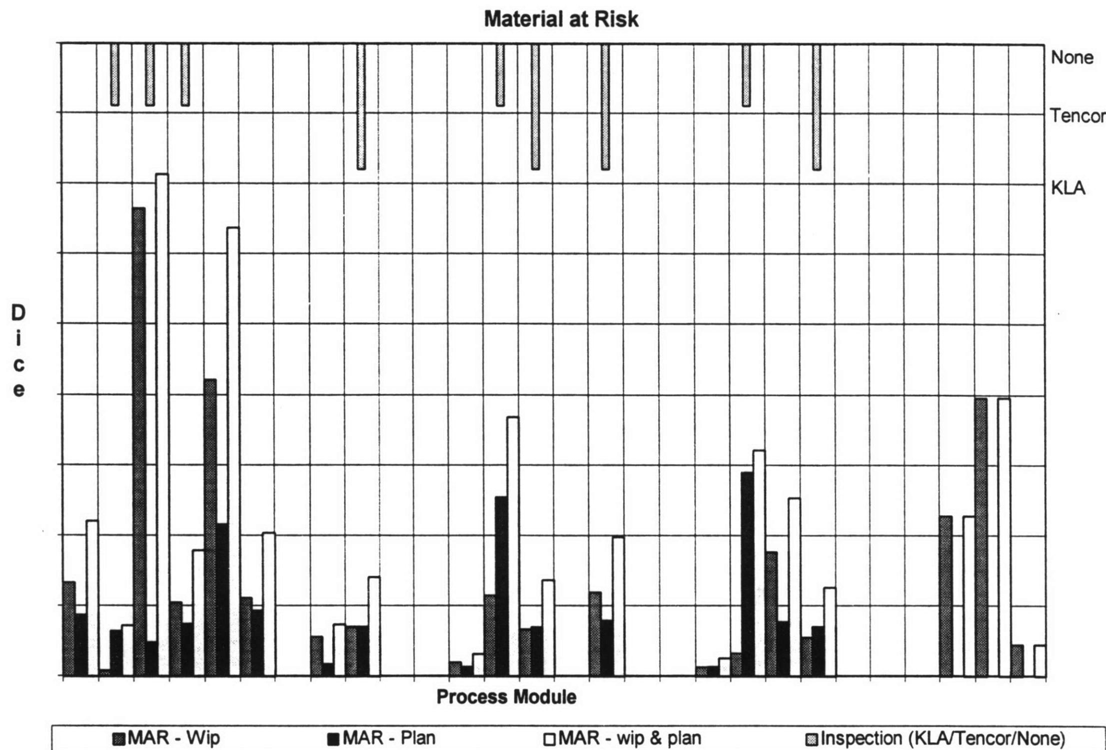


Figure 19. Column chart of MAR at each module with proposed sampling plan.

Rule-based Optimization

By following several simple rules and using the scenario analysis tool, the inspection plan can be optimized to a point relatively close to the global optimum (as determined by the integer program). In the process modeled in this project, the rule-based optimization was within 5% of the global optimum identified by the integer program.

The rule-based optimization is begun by removing all in-line inspections from the analysis. This provides an upper bound on the MAR to quantify in relative terms the added value of in-line inspection. This step also illustrates which step has the highest potential of causing an impact on yield. To reduce the MAR value for that module, an inspection should be located after that module. An inspection at that location will increase the probability of recognizing excursions from that module and reduce their impact. After the first inspection

placement, there are two options for further improving the sampling scheme depending on the next highest MAR value. If the highest impact on yield is another location, then an inspection should be added at that step. If the highest contributor is a location which already has an inspection and most of the MAR comes from MAR-PLAN, then the sampling plan at that process step should be adjusted to increase inspection coverage. If most of the MAR is due to MAR-WIP at that step then the overall risk can only be reduced by considering an additional inspection step within that module. This cycle of adding locations or increasing inspection coverage should continue until all available inspection resources are assigned. There are several other guidelines in allocating inspection resources.

- Assign the proper type of inspection equipment. If the analysis includes inspection equipment with different effectiveness, start by assigning the more sensitive equipment until the number of possible inspection locations are maximized. Then consider the effect of using the less sensitive equipment at each location to determine which steps will receive the larger benefit from the more sensitive monitor. (The more sensitive equipment usually requires more inspection time or is associated with greater variable cost so it is prudent to maximize its usage in terms of its benefit.) Assign the more effective monitor to inspection locations preceded by other modules with a high risk of yield impact. It may be necessary to move the inspection location a few modules in each direction to discover the optimum inspection location.
- Consider the excursion characteristic in assigning inspection coverage. The excursion characteristic, or how many lots versus wafers were impacted has an effect on the probability of recognizing the source of an excursion. An excursion with higher lot than wafer impact will require an inspection coverage with higher lot coverage to achieve the highest benefit from inspection resources.

- Consider sensitivities. The effect on MAR from changes in volume, inspection resources or monitor effectiveness should be examined to determine the effect of a possible redesign. A plan which is more sensitive to changes in volume should not be selected for a process about to be ramped for higher production. If a process has only recently begun production, select a plan which is relatively insensitive to monitor effectiveness errors to protect against limited process understanding.
- Risk of moving inspection locations. While data conditioning is intended to remove the effects of current in-line inspections, it is wise to keep this effect in mind when considering radical inspection plan redesign. Moving an inspection to target a different excursion source may result in a "see-saw" effect as excursions pop up in the now vacant inspection location. The benefits of the inspection modification should be balanced with the risks associated with a redesign.
- Cost of inspection movement. There is a cost associated with each inspection resource re-allocation. This cost includes the extra investment needed to perform the adjustment and the cost of lost information. When an inspection is moved, the value of information pertaining to excursion behavior and yield impact at that step is decreased. While any significant inspection plan improvement will justify the added costs and loss of information, these costs should be considered before each inspection plan modification.

Model Assumptions

The general formulation is based on several assumptions which must be understood to leverage the value of this model. There are also several factors which are not considered in this formulation but are opportunities for future improvement.

- Independence of inspection steps. The model assumes there is only one inspection location for each defect source. While it is possible to detect an excursion which passed undetected through a previous inspection, these effects are not considered under this formulation.
- Multiple machines at a process step. The effects of multiple machines on the likelihood of recognizing the source of the excursion and the value of collected information are too complex to be considered under this formulation. The defect reduction department is much more comfortable assigning an excursion to one tool of many at a step if all the defective material flowed through that equipment. If only one tool is used, it is statistically insignificant if it processed all the defective material. For example, if a group of excursionary lots are all processed at a particular step with one machine of the 10 possible, it is likely that the one machine in the cause of the excursion. But if there were only one machine at that step, there is nothing unique about it processing all the excursionary product. Since the excursionary lots have the machine in common by force and not by chance, the common machine may or may not have been the source.
- No consideration of inspection costs. The inspection costs associated with the probability of accepting defective material (alpha risk) or rejecting acceptable material (beta risk) were not considered.
- Assumes no disposition of in-line material. In-line rejection of material is a limited practice because of the informational value of fully processed material and the potential for acceptable product. Each lot which is processed and inspected increases the amount of information available to characterize the yield impact of defects originating at each step. Since the exact impact of each excursion on final yield cannot be determined beforehand, each wafer is processed fully to maximize output.

- Assumes independence of excursions. The probability of product being impacted by more than one excursion is considered to be negligible. This assumption is supported by assigning lots in the historical data to only one excursion.

Additional uses of the models

Investigation of each process line with the optimization and scenario analysis tools will enable a company to recognize opportunities for improvement in excursion control. These tools can be used in additional ways to consider possible changes to the process.

- Lifecycle patterns. A typical process follows a distinct lifecycle curve. Yields are exponentially increasing as improvements increase yield until reaching a plateau. Production volumes follow a bell curve as production ramps up, peaks, then ramps down. These changes effect the amount and location of optimum inspection resources and should be considered while designing an inspection plan.
- Effect of additional process. It is important to consider the influence of other process lines, especially in light of lifecycle patterns. Proper allocation of inspection resources between different processes should consider the influence of products in different lifecycle stages and varying profit margins. When should resources be shifted from the mature high-margin process to the emerging process line? What process improvements must be reached before bringing in additional process lines?
- Benefit of process improvement. One of the potential benefits of process improvement is a decreased need for excursion control. A complete analysis of proposed improvements will include the impact to inspection resources. How much improvement must be achieved to significantly reduce inspection requirements?
- Benefit of additional inspection resources. A decision about a monitor upgrade or headcount increase is reached only after balancing the costs with the projected improvement. The

optimization and scenario analysis tools will allow a company to consider the hypothetical impact of additional or improved inspection equipment or systems.

Example of Inspection Plan Analysis

This example is included to demonstrate how to use both support tools in evaluating the allocation of inspection resources. In this fictitious example there are two stages of analysis that must be done. The first analysis is to optimize the current inspection plan. The plan currently being used is based on the plan passed down from the development fab during initial development . The process has only been modified incrementally since then. The second analysis must be performed to anticipate the changes associated with a maturing of the process. Specifically, these changes include an increase in production volume and a decrease in defect levels.

A graph of the current material at risk (MAR) is shown below in Figure 20. The current expected loss of 26.2 wafers is well below the loss of 62.7 wafers if there were no inspections but there is still significant opportunity for improvement. The optimization tool identifies 20.6 wafers as the lowest possible wafer loss.

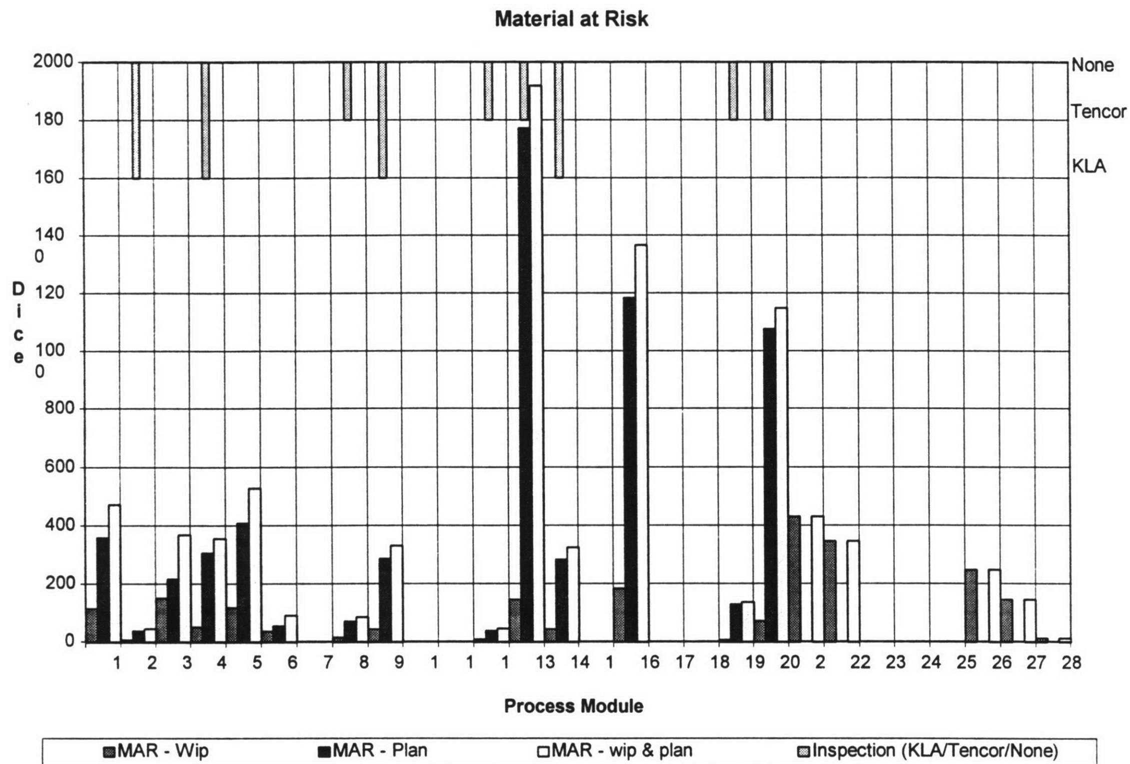


Figure 20. MAR chart for current inspection plan.

The goal for implementation is to work towards the optimum in gradual steps. The optimum inspection plan as determined by the CPLEX optimization program is shown in Figure 21. Each step must be as simple as possible and create a reduction in expected defective material for each change. Using the scenario analysis tool it is possible to investigate the options at each step. Each step in the gradual implementation is explained below.

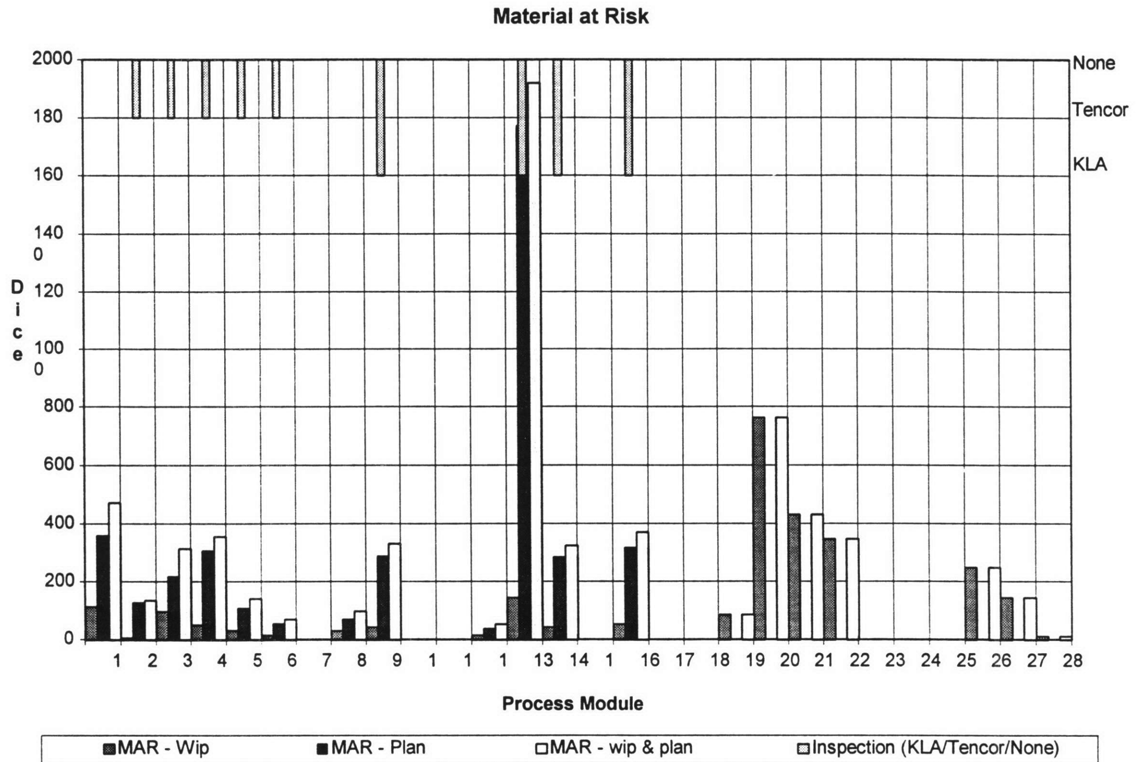


Figure 21. Optimum inspection plan at current production levels.

1. There are 6 options available for the first step. Each option and the associated MAR in dice lost are shown in Figure 22 below. The option shown in bold is the best option. The inspection at step 20 is late enough in the process so that the 100% inspection at Sort will identify any excursions. This inspection resource could be better utilized earlier in the process, i.e., step 6

Current MAR	8395
Move Tencor at step 19 to 6	8191
Move Tencor at 20 to 6	**7639**
Move Tencor at 8 to 6	8041
Move Tencor at 12 to 6	8124
Exchange the KLA at 13 & the Tencor at 2	8485
Exchange the KLA at 13 & the Tencor at 4	8996

Figure 22. Implementation Options for first modification.

2. A similar calculation is done on the remaining options above to determine the second step. Reallocating the inspection at step 19 to step 5 results in a MAR-TOTAL of 6854.
3. The Tencor inspection at step 18 should be moved to step 3 to lower the MAR-TOTAL to 6814. This modification only results in a decrease of 40 dice so it may be worthwhile to consider the cost of inspection modifications mentioned in the section entitled "Rule-based Optimization." After this modification, there are no more simple inspection modifications available. A more complicated modification must be considered.
4. There are four options available at this phase of the modification as shown in Figure 23. The Tencor inspection at step 12 should be moved to step 4 and the KLA inspection at step 4 should be moved to step 16. This results in a MAR of 6615 dice.

Move KLA inspection at step 12 to 2 & shift KLA inspection at 2 to 16	6705
Move KLA inspection at step 12 to 4 & shift KLA inspection at 4 to 16	**6615**
Move KLA inspection at step 13 to 2 & shift KLA inspection at 2 to 16	6757
Move KLA inspection at step 13 to 4 & shift KLA inspection at 4 to 16	6666

Figure 23. Implementations options for fourth modification.

5. The integer program recommends exchanging the Tencor inspection at step 13 with the KLA inspection at step 2, but the MAR value (6615 dice) achieved by the previous modification is the same as the target. This modification will result in no significant reduction of MAR, i.e., less than 1 die improvement.

The proposed sampling plan is more robust to variations in monitor effectiveness and only slightly more sensitive to changes in production volumes. If the proposed solution had been dramatically less robust then the current sampling plan, then a trade-off would need to be made between the anticipated improvements and the higher risk caused by this sensitivity.

As the process matures the defect probabilities are expected to decrease because of improvement efforts. Production volumes are expected to increase from 150 lots per week to over 275 lots per week. It is desirable to anticipate any modifications to the sampling plan that

will be necessary to optimize in-line inspection as the process matures. The proposed sampling plan is shown in Figure 24.

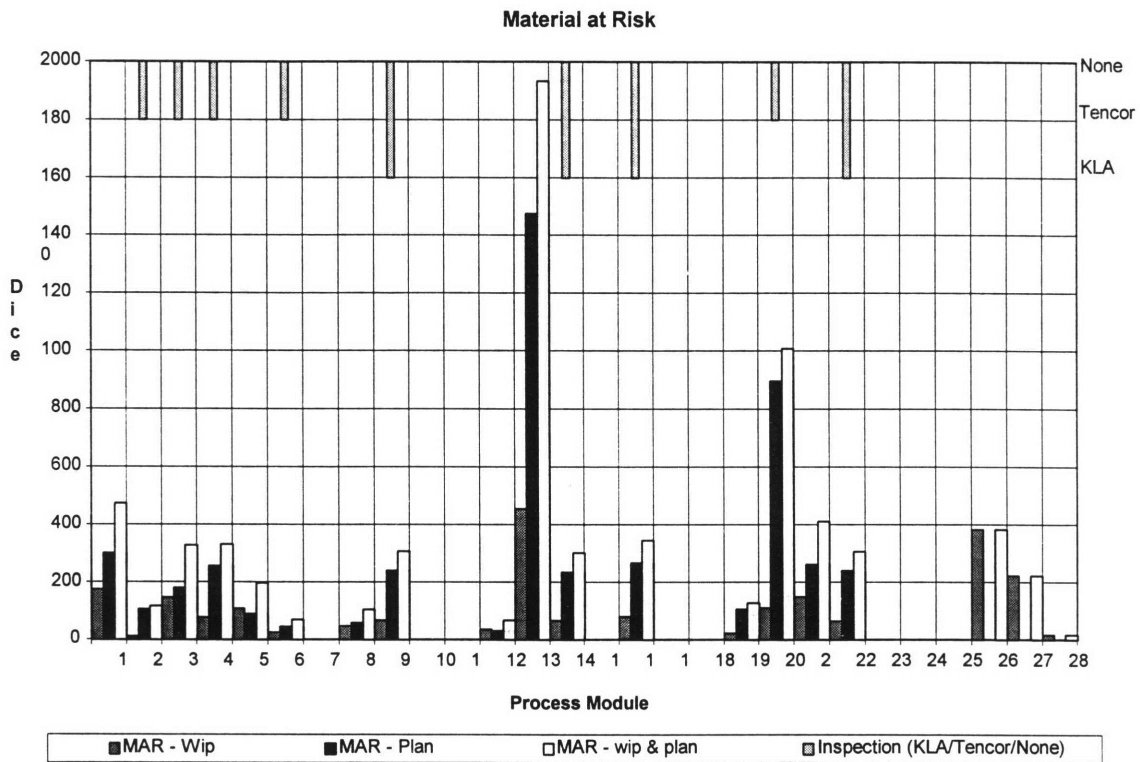


Figure 24. The optimum sampling plan when the process has matured.

There are only two options to be considered to adapt to the maturing process. The first option is to move the Tencor inspection from step 5 to step 20 which results in a MAR of 7288. The second option is to first move the KLA inspection at step 13 to step 22 which results in a MAR of 7256. The MAR increased from 6615 in the first part of the example to 7401 because of the 77 lots per week increase in production volume and the 20% across-the-board decrease in defect levels. The second option (move the KLA) results in a bigger improvement so it should be implemented first if a choice must be made. After both changes the MAR is 7048 dice with close to 500 wafer inspections per week. It may be more realistic to reduce the wafer or lot inspection coverage or even eliminate inspections to reduce the number of inspections. Only 270

wafers per week are inspected at the current production volumes before the increase considered in the second half of the example.

Conclusion

The reduction in variation achieved through inspection optimization makes it an essential part of process design for companies with a serial process line provided there is some flexibility in the location of in-line inspections. The results achieved in this study indicate significant opportunities for improvement. In the process line examined as part of this project more than $837,000 each week could be gained by reducing the impact of excursions through inspection resource optimization.

The two programs support a gradual implementation towards the global optimum. It is recommended that both programs (integer program and scenario analysis) be used to improve in-line inspections.

Bibliography

1. Wang, Ping, Fourmun Lee, K. Michael Chan, Ray Goodner and Ron Ceton. 1996. Yield Enhancement in a High-Volume 8-in. Wafer Fab, *Semiconductor International* June, pp. 221-226.

2. Longtin, Mark, Lawrence M. Wein, and Roy E. Welsch. 1996. Sequential Screening in Semiconductor Manufacturing, I: Exploiting Spatial Dependence, *Operations Research* 44, pp. 173-195.

3. Sharma, Vikas. 1996. Doctorate Thesis, Massachusetts Institute of Technology.

4. Pappu, Suguna. 1989. A Dual Ascent Algorithm for Finding the Optimal Test Strategy for an Assembly Sequence. M.S. Thesis, MIT, Cambridge, MA.

5. Schmidt, Dennis N. and David Trestain. 1996. Silicon Wafer Sleuthing, *The Electrochemical Society Interface* Summer, pp. 24-28.

Appendix

Visual Basic Program used by Scenario Analysis tool

```
'macro which runs the multiple scenarios and the sensitivity analysis

Sub multiple_scenario()

    Dim cnt_steps As Integer

    Set inpt = Worksheets("inputs")
    Set outpt = Worksheets("output")

    'counts how many scenarios are being considered by the number of headings
    cnt_trial = 0
    Do While inpt.Cells(5, 12 + cnt_trial).Value <> ""
        cnt_trial = cnt_trial + 1
    Loop

    'counts the number of steps - counts till it finds 'SORT'
    cnt_steps = 0
    Do While UCase(inpt.Cells(7 + cnt_steps, 3).Value) <> "SORT"
        cnt_steps = cnt_steps + 1
    Loop

    'copies step names from inputs sheet to output sheet if they aren't there already
    For i = 1 To cnt_steps + 1
        If outpt.Cells(1, 1 + i).Value <> inpt.Cells(6 + i, 3).Value Then
            outpt.Cells(1, 1 + i).Value = inpt.Cells(6 + i, 3).Value
        End If
    Next i

    'copies scenario plan into target cells
    For trial = 1 To cnt_trial
        'puts volume in target cell
        inpt.Cells(3, 2) = inpt.Cells(6, 11 + trial).Value

        'puts sampling plan in target cells
        For i = 1 To cnt_steps
            inpt.Cells(6 + i, 4) = inpt.Cells(6 + i, 11 + trial).Value
        Next i

        'runs sensitivity
        If UCase(inpt.Cells(4, 11 + trial).Value) = "Y" Then
            For s = 0 To 6
                Call scenario(cnt_steps, (s), (trial))
            Next s
        Else
            Call scenario(cnt_steps, 0, (trial))
        End If

        'output
```

```
    outpt.Cells(14 + trial, 1).Value = inpt.Cells(5, 11 + trial).Value
    outpt.Cells(14 + trial, 3).Value = Left(outpt.Range("B12").Value, Len(outpt.Range("B12").Value) -
12)
    outpt.Cells(14 + trial, 2).Value = outpt.Range("B7").Value

  Next trial

End Sub

Option Base 1
'macro which computes MAR-WIP, -PLAN for proposed inspection plan

Sub scenario(cnt_steps As Integer, s As Integer, trial As Integer)

  Const alpha = 0.05

  Dim sens(7) As Integer

  mx = cnt_steps + 1   'number of process steps

  ReDim iij(mx, mx, 2) As Double
  ReDim vij(mx, mx) As Double
  ReDim di(mx) As Double
  ReDim theta(mx) As Double
  ReDim mw(mx) As Double
  ReDim mp(mx) As Double
  ReDim mar(mx) As Double
  ReDim insp(mx) As String
  Dim yj, el, ew, ed, vj As Range

  Set inpt = Worksheets("inputs")
  Set outpt = Worksheets("output")
  Set yj = inpt.Range("D7:D" & 6 + mx)   'y/n inspection at that step
  Set el = inpt.Range("e7:e" & 6 + mx)   'pr(exc-lot)
  Set ew = inpt.Range("f7:f" & 6 + mx)   'pr(exc-wafer|exc-lot)
  Set ed = inpt.Range("g7:g" & 6 + mx)   'pr(exc-dice|exc-wafer)
  Set vj = inpt.Range("h7:h" & 6 + mx)   'volume tween steps
  Set effk = Worksheets("kla_eff").Range("D3:ae30")     'effectiveness of monitors
  Set efft = Worksheets("tncr_eff").Range("D3:ae30")

  dpw = inpt.Range("B1")
  wpl = inpt.Range("B2")

  If s <> 0 Then sens(s) = 1  'sensitivity calculation

  'adjusts volume by 1% for sensitivity calculation
  If s = 3 Then
    inpt.Range("B3") = inpt.Range("B3") * 1.01
  ElseIf s = 4 Then
    inpt.Range("B3") = inpt.Range("B3") / 1.01
  End If

  h_see = 0.68 * (1 + 0.01 * sens(4)) 'probability of seeing defect
  m_see = 0.4375 * (1 + 0.01 * sens(5))
```

```
l_see = 0.195 * (1 + 0.01 * sens(6))
z_see = 0

h_rec = 0.5   'probability of recognizing source of defect
m_rec = 0.2
l_rec = 0.1
z_rec = 0

p_max = 200

'determines what inspection plan will be used to inspect each defect source
For st = mx - 1 To 1 Step -1
   If yj.Cells(st).Value <> "" Then temp3 = yj.Cells(st).Value
   insp(st) = temp3
Next st

If s = 0 Then Worksheets("exp").Range("A1:BB30").ClearContents

'computes theta vector: effectiveness due to sampling plan
'binomial function based on excursion=2 exc-wafers/lot definition
For i = 1 To mx - 1
   di(i) = el(i) * ew(i) * ed(i)
   If insp(i) <> "" Then
      W = Right(insp(i), 1) + sens(2)
      outpt.Range("A1") = "=1-BINOMDIST(1," & W & "," & ew(i) & _
         ",TRUE)+ BINOMDIST(1," & W & "," & ew(i) & ",FALSE)*(1-BINOMDIST(0,2," & _
         ew(i) & ",TRUE))"
      theta(i) = outpt.Range("A1")
      'outputs theta value to output sheet
      If s = 0 Then Worksheets("exp").Cells(1, i + 1).Value = theta(i)
   End If
   If i = mx - 1 Then outpt.Range("A1").Value = ""
Next i

'makes vij matrix from vj vector  [volume between steps]
For r = 1 To mx
   For c = mx To 1 Step -1
      If r = c Then vij(r, c) = vj(c)
      If c < r Then vij(r, c) = vij(r, c + 1) + vj(c)
   Next c
Next r

temp = mx   'temp = last inspection step
For j = mx To 1 Step -1 'loops from the bottom to the top
   temp2 = 0
   'determines what monitor will be used & adds up number of scans per monitor
   Select Case Left(yj.Cells(j).Value, 1)
      Case "k", "K"   'kla matrix
         d3 = 2
         temp2 = 1
         'k_scans=sum[lpw*wafer coverage/lot coverage]
         If j <> mx Then k_scans = inpt.Cells(6, 11 + trial) * (Right(yj.Cells(j).Value, 1) + _
            sens(2)) / (Left(Right(yj.Cells(j).Value, Len(yj.Cells(j).Value) - 2), _
            Len(yj.Cells(j).Value) - 4) + sens(1)) + k_scans
```

```
    Case "t", "T"   'tencor matrix
      d3 = 1
      temp2 = 1
      If j <> mx Then t_scans = inpt.Cells(6, 11 + trial) * (Right(yj.Cells(j).Value, 1) + _
        sens(2)) / (Left(Right(yj.Cells(j).Value, Len(yj.Cells(j).Value) - 2), _
        Len(yj.Cells(j).Value) - 4) + sens(1)) + t_scans
  End Select
  If temp2 = 1 Then
    outpt.Cells(5, j + 1) = d3
  Else
    outpt.Cells(5, j + 1) = ""
  End If

  'sets up mx x mx x 2 matrix for what source, what inspection step, what monitor
  For i = j To 1 Step -1        'loops from jth column back to first
    iij(j, i, d3) = temp2      'sets it equal to temp2 (0/1)
    If yj.Cells(j).Value <> "" Then
      For r = j + 1 To temp   'loops down to set cells below it to zero
        For c = 1 To j     'loops across to jth column
          For m = 1 To 2  'repeats for both monitor types
            iij(r, c, m) = 0
          Next m
        Next c
      Next r
      temp = j    'temp=last inspection step
    End If
  Next i
Next j

'calculates MAR from Iij matrix
For m = 1 To 2
  For r = 1 To mx
    For c = 1 To r
      If iij(r, c, m) = 1 Then
        If m = 1 Then       'm=2 for kla, m=1 for tencor
          effm_ch = efft(c, r)
        Else
          effm_ch = effk(c, r)
        End If
        Select Case effm_ch    'converts h,m,l rating to numbers
          Case "H", "h"
            effm_see = h_see
            effm_rec = h_rec
          Case "M", "m"
            effm_see = m_see
            effm_rec = m_rec
          Case "L", "l"
            effm_see = l_see
            effm_rec = l_rec
          Case Else
            effm_see = z_see
            effm_rec = z_rec
        End Select
```

```
            'calculates MAR, except for sort since it has 100% inspection
            If r <> mx Then
              mw(c) = di(c) * vij(r, c)
              'If c = 14 Then Stop
              k = Log(1 - theta(c) * effm_see * effm_rec) 'coefficient
              leng = Len(insp(c)) 'what monitor, lot & wafer coverage
              If leng > 0 Then
                L = Right(Left(insp(c), leng - 2), leng - 4) + sens(1)
              Else
                L = 0
              End If
              t = 0
              Worksheets("exp").Cells(2, c + 1) = outpt.Cells(1, c + 1)
              Do
                t = t + 1
                mp(c) = mp(c) + Exp(k * t) * L * dpw * wpl * di(c)
                'outputs Pr[not identifying excursion] to output sheet
                If di(c) > 0 And s = 0 Then
                  Worksheets("exp").Cells(2 + t, c + 1) = Exp(k * t)
                End If
              Loop Until Exp(k * (t + 1)) < alpha Or (t + 1) * L >= p_max
              mar(c) = mw(c) + mp(c)
            Else
              mw(c) = di(c) * vij(r, c)
              mar(c) = mw(c)
            End If

          End If
          If m = 2 And r = mx Then       'outputs mar matrix and total mat'l at risk
            If s = 0 Then
              outpt.Cells(2, c + 1) = mw(c)
              outpt.Cells(3, c + 1) = mp(c)
              outpt.Cells(4, c + 1) = mar(c)
            End If
            mar_tot = mar(c) + mar_tot
          End If
        Next c
      Next r
    Next m
    If s <> 0 Then
      If s < 3 Then
        'outputs sensitivity information
        outpt.Cells(14 + trial, 2 + s * 2) = mar_tot - outpt.Cells(7, 2)
        outpt.Cells(14 + trial, 3 + s * 2) = k_scans + t_scans - _
          Left(outpt.Range("B12"), Len(outpt.Range("B12")) - 12)
      Else
        outpt.Cells(15 + trial - 1, 5 + s) = mar_tot - outpt.Cells(7, 2)
      End If
    Else
      outpt.Cells(7, 2) = mar_tot
      outpt.Cells(10, 2) = k_scans
      outpt.Cells(11, 2) = t_scans
      For clr = 1 To 9
        outpt.Cells(14 + trial, 3 + clr).ClearContents
```

```
    Next clr
  End If

End Sub
```

GAMS input for integer program

```
SETS

    i  source  /1*29/
    j  inspection point  /1*29/
    s  inspection monitor /KLA, TNCR /

  ALIAS (j,jh);
  ALIAS (i,ih);

  PARAMETER

  equiplim(s)  number of possible locations for equipment type s

      / KLA  4
       TNCR  6 /;

  TABLE M(i,j,s)  MAR in dice
.....
  ;


  VARIABLES

  L(i,j,s)  binary equals 1 if inspection for i at j with s
  Y(j,s)   binary equals 1 if inspection at j with s
  Z        total expected defective material ;

  BINARY VARIABLE L,Y;

  EQUATIONS

  TMAR      defines objective function
  SETY(i,j,s)  forces Yjs to be 1 as appropriate
  LIMITY(j)  assures at most one inspection type at each point
  LIMITL(i)  assures exactly one inspection per source i
  CONST(s)  constraint on resource of each inspection type
  insp13    forces inspection at step 13 -- performed by another department
  insp20    forces inspection at step 20
  sort      forces (TNCR) inspection at sort
  boxat(i,j) optional constraint which checks box above i,j;

  TMAR..      Z =e= SUM((i,j,s)$(ORD(j) ge ORD(i)), L(i,j,s)*M(i,j,s));

  SETY(i,j,s)$(ORD(j) ge ORD(i))..  L(i,j,s) =l= Y(j,s);
```

```
LIMITY(j)..  sum(s, Y(j,s)) =l= 1;

LIMITL(i)..  sum((j,s)$(ORD(j) ge ORD(i)), L(i,j,s)) =e= 1;

CONST(s)..   sum(j, Y(j,s)) =l= equiplim(s);

insp13..     Y('13', 'TNCR') =e= 1;
insp20..     Y('20', 'TNCR') =e= 1;

sort..       Y('29', 'TNCR') =e= 1;

boxat(i,j)$(ORD(j) ge ORD(i))..  sum((ih,jh,s)$((ORD(ih) le (ORD(i)-1))
                          $(ORD(jh) ge (ORD(j)+1))), L(ih,jh,s))
                       =l= (ord(i)-1) * (1- (sum(s,L(i,j,s))));

OPTION OPTCR = 1e-09;
OPTION LIMCOL = 0;
OPTION ITERLIM = 100000;
OPTION RESLIM = 100000;
OPTION SOLPRINT = OFF;

MODEL INTEL /ALL/;

SOLVE INTEL using MIP minimizing z;

OPTION L:0:2:1;    DISPLAY L.L;
OPTION Y:0:1:1;    DISPLAY Y.L;
```

```
GENERATION TIME     =      25.017 SECONDS

         S O L V E     S U M M A R Y

MIP Solution  :      3797.750000    (717 iterations, 6 nodes)
Final LP      :      3797.750000    (0 iterations)

Best integer solution possible :      3797.750000
```

	KLA	TNCR
1 .2	1	
2 .2	1	
3 .3		1
4 .4		1
5 .5		1
6 .9	1	
7 .9	1	
8 .9	1	
9 .9	1	
10.13		1
11.13		1
12.13		1
13.13		1
14.16	1	
15.16	1	

16.16	1	
17.20		1
18.20		1
19.20		1
20.20		1
21.22	1	
22.22	1	
23.29		1
24.29		1
25.29		1
26.29		1
27.29		1
28.29		1
29.29		1

---- 140 VARIABLE Y.L binary equals 1 if inspection at j with s

	KLA	TNCR
2	1	
3		1
4		1
5		1
9	1	
13		1
16	1	
20		1
22	1	
29		1