

Comparative Assessment of Yield Learning Tools Using Information Theory

by Charles Weber^a, Vijay Sankaran^b, Kenneth W. Tobin Jr.^c and Gary Scher^d

^aMIT Sloan School of Management, Cambridge, MA 02142

^bSEMATECH, Austin, TX 78741

^cOak Ridge National Laboratories, Oak Ridge, TN 37831

^dSleuthworks, Inc., Fort Collins, CO 80525

ABSTRACT

A model based on information theory, which allows yield managers to choose the optimal strategies for yield management in microelectronic manufacturing, is presented. The data reduction rate per experimentation cycle and data reduction rate per unit time serve as benchmarking metrics for yield learning. These newly defined metrics enable managers to make objective comparisons of apparently unrelated technologies. Four yield analysis tools -- electrical testing, automatic defect classification, spatial signature analysis and wafer position analysis -- are examined in detail to suggest an optimal yield management strategy for both the R&D and volume production environments.

Keywords: semiconductor manufacturing, yield management, automatic defect classification, spatial signature analysis, wafer tracking, parametric test, entropy of information, Shannon's information theorem.

1. INTRODUCTION

Manufacturers of integrated circuits invest billions of dollars in process equipment, and they are interested in obtaining as rapid a return on their investment as possible. Market dynamics allow for healthy profit margins only during the early part of a product's lifetime. The sooner a potentially lucrative circuit yields, the sooner the manufacturer can generate a revenue stream.^{1,2} Rapid yield learning is thus becoming an increasingly important source of competitive advantage. Defects must be detected, analyzed and eliminated within increasingly shorter time periods. The economic benefit to be derived from shortening of data cycles can be better understood in the following example:

If one process tool disperses 100 defects of half a micron in diameter onto product wafers, it can "kill" most of the 200 chips (typically) on wafers with a diameter of 200 millimeters. Let us assume that on average the tool kills half the chips on every wafer, and that these chips are microprocessors that sell for \$100 a piece. The tool will therefore cause \$10,000 of damage in the time that it takes to process one wafer, which may be as short as one minute. If the tool is situated near the beginning of the line, it may damage wafers that will take nearly two months to reach the end of the line, when the problem is discovered. In the interim, the tool continues to reduce the profit of the operation by \$10,000/minute, \$600,000 per hour or more than \$10 million per day. If the problem is discovered after the first damaged wafer that exits the end of the line, more than \$500 million will have been lost.

Clearly, it is in every semiconductor manufacturer's interest to discover process excursions and tool contamination early, even if they do not actually impact the yield. The potential damage is so great that the manufacturer needs to proactively respond to every defect signal. Semiconductor manufacturers have thus invested in expensive in-line inspection tools that detect contamination that may or may not cause electrical faults within a few hours of critical process steps. Defect analysis tools and defect-sourcing methods reduce the data generated by inspection tools and indicate the process steps that are the likely culprits of potential yield loss.

Yield managers have a large but expensive arsenal of yield improvement tools and methods at their disposal. Different tools perform different functions under different conditions and some combinations of tools and methods work better than others. Yield managers need to know which combination of tools works most optimally and cost-effectively to maximize the profitability of their operations. They require metrics that allows them to assess the value of apparently unrelated options. In layman's terms, they need to effectively compare apples with oranges.

This paper uses a model based on information theory in an attempt to create an objective method of comparing technology options for yield analysis. The data reduction rate per experimentation cycle and data reduction rate per unit time serve as benchmarking metrics for yield learning. Four yield analysis technologies -- electrical testing (ET), automatic defect classification (ADC), spatial signature analysis (SSA) and wafer position analysis (WPA) -- are examined in detail to determine an optimal yield management strategy for both the R&D and volume production environments.

2. CONVERTING DATA TO KNOWLEDGE

Yield learning is an iterative experimentation process that is repeated until all sources of yield loss are detected, identified and eliminated, or until the cost of further experimentation exceeds the benefit of the knowledge gained^{3,4}. Yield learning can be accelerated by shortening the experimentation cycle, or by making each experimentation cycle more effective. The former option depends upon the engineering team's ability to reduce the design time of an experiment, fabrication facility's ability to reduce the fabrication cycle time, the test area's ability to accelerate the data generation rate, and the engineering team's ability to increase the data analysis rate. The latter option depends upon how the data are analyzed, and to what extent they are converted to knowledge.

Information theory provides an excellent metric for how effectively data is converted to knowledge, specifically, the entropy of the information source⁵⁻⁸. A source of information reveals an amount of information $I(X_i)$ whenever the source is in state X_i ⁹. $I(X_i)$ is known as the *self-information* and is given by

$$I(X_i) = -\log_2 P(X_i) \quad \text{bits} \quad (\text{Eq. 1})$$

where $P(X_i)$ is the probability of occurrence of state X_i . The unit "bits" is related to the base of the logarithm; it could alternately be "hartleys" for a base of 10 or "nats" for a base of e .

Information entropy is defined as expectation of $I(X_i)$ or the average amount of self-information per state^{10,11}. It is given by the random variable

$$H(X_i) \equiv \langle I(X_i) \rangle = \sum_{i=1}^m P(X_i) \cdot I(X_i) = -\sum_{i=1}^m P(X_i) \cdot \log_2 P(X_i) \quad \text{bits/state} \quad (\text{Eq. 2})$$

Information entropy is at a maximum when all states are equiprobable or $P(X_i) = 1/m$. It approaches zero as the probability of one state approaches unity and the probability of all other states approach zero. In other words, information entropy approaches zero, as knowledge is extracted from the information.

The *relative entropy* of a probability distribution $P(X_i)$ with respect to a second probability distribution $P(Y_i)$ is given by

$$H_{P(X_i)||P(Y_i)} = \sum_{i=1}^m P(X_i) \cdot \log_2 \frac{P(X_i)}{P(Y_i)} \quad \text{bits/state} \quad (\text{Eq. 3})$$

where the sum covers all possible states of the system ⁸, and $P(Y_i)$ plays the role of a reference measure. The relative entropy can thus be used to compare the final state of an experimentation cycle to the initial state or to benchmark the amount of knowledge extraction performed by two different, possibly unrelated processes.

In an economic environment governed by radical experience curves, the actual value of knowledge extraction also depends on the time required to reduce the data, or the length of the experimentation cycle. A tool that identifies the source of an electrical fault with absolute certainty but requires the full wafer fabrication process to be completed may be less valuable than a tool that can identify the source of a fault with significant probability within a few hours. A total systems approach that uses combinations of tools that complement each other may be the most effective strategy ¹², and the global rate of knowledge extraction can serve as the metric that determines the optimal combination.

3. ELECTRICAL TESTING AND WAFER POSITION DATA

Chips on product wafers are electrically tested for functionality shortly after they emerge from the fabrication facility. The wafers also typically contain microelectronic test structures that may reveal characteristics of the fabrication process when subjected to a parametric test, where the effects of varying several electrical parameters on the output signal of the test structures are studied. Functional testing alone identifies defective chips, but both functional and parametric testing of product wafers can localize the source of an electrical fault to within a neighborhood of a few process steps. In doing so they dramatically reduce the entropy of the information source. For example, if the fact of chip failure were the only information available to a yield engineer, the engineer would have a $1/500=0.002$ chance of identifying the culprit process step in a process that consists of 500 such steps. If, however, the engineer had access to the information that the n-type metal oxide semiconductor (NMOS) transistor threshold voltage was out of specification, then the engineer could most likely use his/her expertise to reduce the source of the electrical fault to a neighborhood of about 20 process steps. These include gate oxidation, polysilicon lithography, polysilicon etch, spacer oxidation and a few ion implants. Without access to the history of the process, the engineer would have to assume that each step has a $1/20=0.05$ chance of being the culprit. Assuming the potentially relevant steps are numbered 151 through 170, then substituting the aforementioned odds into Equation 3 yields

$$\begin{aligned}
 H_{POST||PRE-TEST} &= \sum_{i=1}^{150} 0 \cdot \log_2 \frac{0}{0.002} + \sum_{i=151}^{170} 0.05 \cdot \log_2 \frac{0.05}{0.002} + \sum_{i=171}^{500} 0 \cdot \log_2 \frac{0}{0.002} \\
 &= 4.64 \text{ bits/process step}
 \end{aligned}
 \tag{Eq. 4}$$

With prior knowledge of the history of the process, the engineer could assess that some of the 20 candidate steps were more likely to be the culprits. The initial entropy of the 20 candidate steps would then be lower than what is inferred by equiprobability, increasing the relative entropy of the knowledge extraction in Equation 4 and the apparent value of electrical parametric testing.

As mentioned previously, the knowledge extraction rate is the more meaningful assessment of value for yield analysis tools. An engineer who had no access to a parametric tester could, for example, have reduced the possible number of culprit process steps from 500 to 20 by stripping back the wafer to the to the process layers that hosted the problem. However, this would have taken more than 100 times as long as localizing the problem by a parametric tester, which could evaluate a wafer in about an hour. (In addition, a whole wafer, possibly worth a few thousand dollars, would have to be sacrificed for stripback.) A parametric tester can thus be assigned a data reduction rate of about 5 bits/process step/hour, whereas stripback most likely does not reduce data at a rate faster than 0.05 bits/process step/hour. Most engineers therefore value access to a parametric tester very highly.

Nonetheless, by itself parametric testing can rarely be used to pinpoint the source of an electrical fault to the actual process step. Randomizing and recording the wafer order prior to every process step, may yield enough information to identify the culprit step precisely and rapidly. ^{13,14} Data from parametric testing can be correlated to the wafer order at each process step. Any correlation between an electrical parameter and wafer order can potentially infer causality.

Figure 1 shows two examples of how the value of a parameter can correlate to wafer position. Figure 1a shows no apparent correlation for parameter A, whereas figure 1b depicts a distribution in which the parameter B varies in sequence from lowest to highest. The odds against a distribution like the one in figure 1b occurring by chance are one in 25! or about $6.5 \cdot 10^{-26}$. The odds in favor of the process step in figure 1b being the cause of the variation of parameter B are astronomical and it can be concluded that the subsequent entropy would be nearly zero.

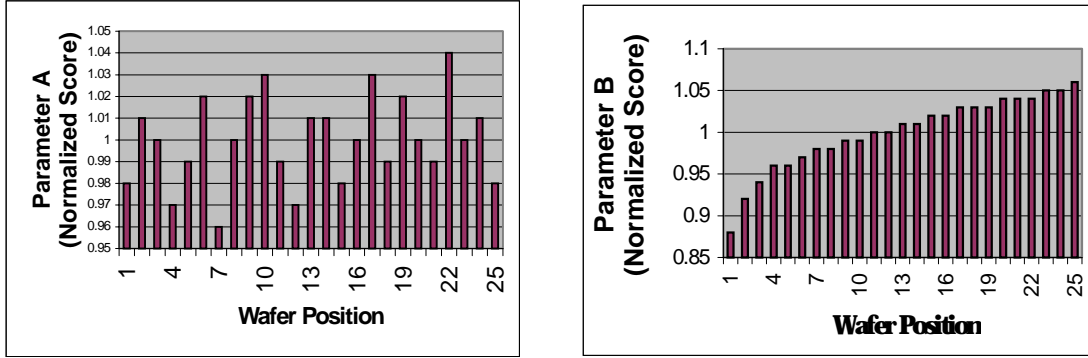


Figure 1 (a). Parametric Data as a Function of Wafer Position (Random Distribution)
(b). Parametric Data as a Function of Wafer Position (Strong Correlation)

Revisiting the case of the out-of-spec threshold voltage of the NMOS transistor with access to wafer position data allows an engineer to reduce the entropy of the data source much further than Equation 4 would suggest. The wafer position data would identify a single process step, say threshold implantation, as the culprit of the electrical fault, effectively reducing the data from equiprobability over 500 steps to virtual certainty. Following the same line of reasoning pursued in deriving equation 4 results in a knowledge extraction of 8.94 bits/process step for the combined approach parametric tester/wafer position data approach, a significant improvement over what a parametric tester can do by itself. Analysis of wafer position data also adds some finite time to the knowledge extraction process, which pegs the data reduction rate of the combined approach to about 8 bits/process step/hour.

To summarize, with no initial information all process steps have the same chance of being the source of the fault. Electrical parametric testing concentrates the probability into 20 steps, increasing the information content and reducing the information entropy in the process. Wafer position data pinpoints the process step and reduces the information entropy to very low levels.

4. SHORTENING THE DATA CYCLE

The data analysis activity described in the previous section is only one in a series of steps in the experimentation cycle³. The other steps have to be included to get a meaningful benchmark of yield management strategies. Thus the sum of the design time, fabrication time and analysis time is the appropriate denominator for knowledge extraction, which slows down the learning rate over a full VLSI process cycle of about 50 days to $(8.94 \text{ bits/process step}) / (50 \text{ days}) = 0.18 \text{ bits/process step/day}$. At that learning rate, hundreds of millions of dollars could be lost by the time the source of a problem has been identified. Semiconductor manufacturers have thus resorted to fabricating fractions of the process in parallel on test wafers^{3, 15-17}. These methods, called short cycles, primarily practiced in R&D and on an ad hoc basis in manufacturing because of test wafer cost, cannot capture faults such as plasma damage where a problem near the end of the process affects a structure fabricated near the beginning of the process.

The value of these methods has to be assessed according to how they extract knowledge from a whole gambit of variables. For this purpose it is useful to assume that n electrical parameters characterize a process. Thus the average knowledge extraction is given by

$$\left\langle H_{P(X_i)||P(Y_i)} \right\rangle_j = \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^m P(X_{ij}) \cdot \log_2 \frac{P(X_{ij})}{P(Y_{ij})} \text{ bits/state} \quad (\text{Eq. 5})$$

Let us now assume that a full VLSI process is broken up into 6 modules with a fabrication cycle of 10 days each, and that these modules are fabricated in parallel. Let us also assume that 100 electrical parameters characterize this process ($n = 100$) and that 10 of them cannot be analyzed on short cycle. These 10 parameters would then remain in a state of equiprobability over 500 process steps: their relative entropy would equal zero. The other ninety parameters would experience the previously calculated data reduction of 8.94 bits/process step. Substituting these values into Equation 5 gives

$$H_{\text{POST}} \parallel \text{PRE-SHORT CYCLE} = \frac{10 \times 0 + 90 \times 8.94}{100} = 8.05 \text{ bits/process step} \quad (\text{Eq. 6})$$

Given an experimentation cycle of 10 days this translates into a learning rate of about 0.8 bits/process step/day, a marked improvement over the 0.18 bits/process step/day for the full VLSI process.

5. AUTOMATIC DEFECT CLASSIFICATION AND SPATIAL SIGNATURE ANALYSIS

The potentially dire consequences of not detecting an electrical fault early during the process have motivated technology managers in the semiconductor industry to introduce an inspection step after about every ten process-steps. During these inspections optical imaging and light scattering tools find defects that could cause faults. Most of these tools have the capability to separate defects that have been added to the wafers from defects that were detected at previous inspections. The inspection tools also transfer the coordinates of the defects to defect review tools, which enable an engineer to classify the defects and identify their source. Defect review was performed manually until the advent of ADC, which proved to be more accurate and orders of magnitude faster than a human inspector. In other words, ADC can reduce information entropy much faster than a human being can.

ADC as applied in the semiconductor industry is the process of automatically categorizing wafer defects into one of multiple classes using data captured by wafer analysis instruments. The type of data that is used by the ADC algorithms varies with the application. It could be optical microscope image data, scanning electron microscope (SEM) image data, material composition information (e.g. from SEM energy dispersive spectroscopy), or confocal microscope image data. ADC compares the defect image to a set of images of known defect types and attempts to classify them into previously established categories. These categories are typically associated with process steps through historical data. ADC therefore has an excellent chance (typically, 70 to 80% of all cases) of identifying the source of a fault-causing defect.¹⁸

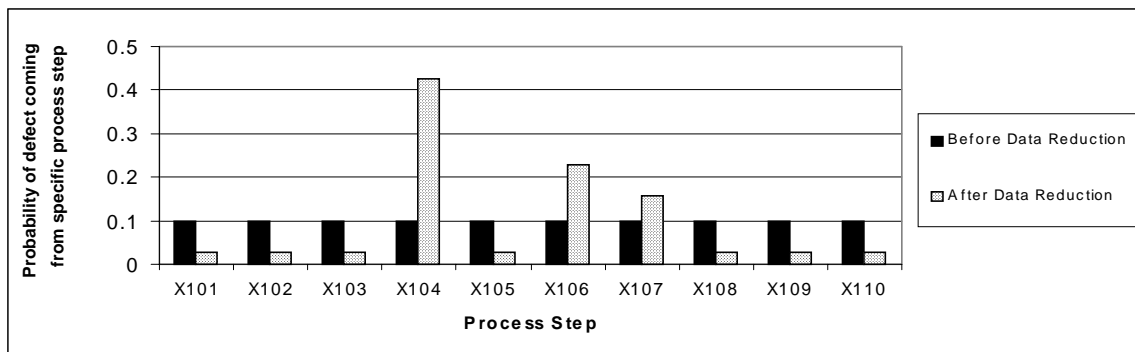


Figure 2. Probability Mass Functions of an Automatic Defect Classification

The phrase "excellent chance" implies finite odds of misclassification or classification into a category called unknown. This complicates the entropy picture by perceptibly adding to the entropy of the source during the data reduction process. Figure 2 illustrates this effect in a hypothetical case where ADC identifies the culprit process step in 73% of all attempts. The other 27% of all attempts either result in misclassification or classification as an unknown. In the absence of a well-documented history of the process, misclassified or unclassified defects have an equal chance of being generated in any of the 10 possible process steps. Let us also assume that for the specific case in Figure 2, ADC points to only 3 defect sources out of the 10 under consideration. The classifier would assign 40% of all defects to step X104, 20% to step X106 and 13% to step X107.

The remaining probabilities of 0.027 per process step come from misclassification or classification as unknown. Therefore, $P(X_{104}) = 0.4+0.027 = 0.427$; $P(X_{106}) = 0.2+0.027 = 0.227$; $P(X_{107}) = 0.13+0.027 = 0.157$; and $P(X_{101}) = P(X_{102}) = P(X_{103}) = P(X_{105}) = P(X_{108}) = P(X_{109}) = P(X_{110}) = 0.027$. Considering that at the beginning there is equal probability of the defects being caused by any of the 10 sources, initially $P(Y_i) = 0.1$ for all ten process steps in question. Substituting these data into equation 3 gives the entropy reduction provided by this ADC classifier.

$$H_{POST||PRE-ADC} = 0.427 \cdot \log_2 \frac{0.427}{0.1} + 0.227 \cdot \log_2 \frac{0.227}{0.1} + 0.157 \cdot \log_2 \frac{0.157}{0.1} + 7 \times 0.027 \cdot \log_2 \frac{0.027}{0.1}$$

$$H_{POST||PRE-ADC} = 0.91 \text{ bits/process step} \tag{Eq. 7}$$

The experimentation cycle for this process sequence, which consists of 10 process steps, equals one day. Therefore the learning rate provided by this ADC system is 0.91 bits per process step per day.

Purposeful randomization of wafer order with prior knowledge of wafer order can narrow the search for root cause whenever some wafer-to-wafer variation is detected at inspection. Thus wafer position data is applicable also for in-line analysis as well as end-of-line correlation to parametric or functional test data. For example, correlating wafer positional data to defect inspection data, short loop monitor results, or visual observations, significantly reduces information entropy and accelerates the defect learning rate.¹⁴

A spatial signature is defined as a distribution of defects originating from a single manufacturing problem. SSA is an artificial intelligence method that has been developed in partnership between SEMATECH, Austin, Texas, and the Oak Ridge National Laboratory, Oak Ridge, Tennessee. The method relies on capturing operator experience through a teaching method to emulate the human response to various manufacturing situations. This has been successfully accomplished through the development and application of an image processing-based, fuzzy classifier system. The technique uses data collected from current in-line inspection tools to interpret and rapidly identify characteristic patterns, or "signatures", that are uniquely associated with the manufacturing process. The SSA system then alerts fabrication engineers to probable yield-limiting conditions that require attention.^{4,19-21}

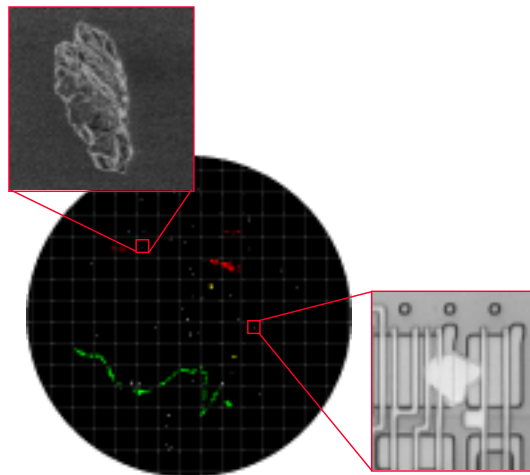


Figure 3. Random defects (inserts) and spatial signature

Figure 3 illustrates the difference between random defects and spatial signatures. The distribution of random defects, such as the ones shown in the inserts of Figure 3, has no clearly discernable pattern. However, spatial signatures, like the scratch on the wafer in Figure 3, are macro phenomena that can kill multiple chips. They generally exhibit distinctive patterns that can be readily discerned by machine vision software. Figure 4 shows an array of such patterns, including one that the SSA classifier has categorized as random due to its apparent lack of a pattern.

We can extend the argument of information entropy reduction to the SSA approach as follows. We assume 100 defects have been detected and that there are 10 possible sources of these defects: Steps X_{201} through X_{210} . Before applying SSA, the probability $P(Y_i)$ of determining the source of any defect is 0.1. Let us also assume that SSA separates 92 of these defects into 3 large clusters such as the ones shown in Figure 4, and only leaves behind 8 isolated defects to classify. Thirty-five of the hundred defects are assigned to step X_{203} ; fifteen to X_{205} ; and forty-two to step X_{208} . The remaining eight unclassified defects could come from any step between and including X_{201} through X_{210} , adding 0.008 to the odds of a defect coming from each of the ten states. Thus $P(X_{201}) = P(X_{202}) = P(X_{204}) = P(X_{206}) = P(X_{207}) = P(X_{209}) = P(X_{210}) = 0.008$; $P(X_{203}) = (0.35 + 0.008) = 0.358$; $P(X_{205}) = (0.15 + 0.008) = 0.158$; and $P(X_{208}) = (0.42 + 0.008) = 0.428$. Then the relative entropy of this knowledge extraction is given by Equation 8, which corresponds to a learning rate of 1.45 bits per process step per day.

$$H_{POST||PRE-SSA} = 0.158 \cdot \log_2 \frac{0.158}{0.1} + 0.358 \cdot \log_2 \frac{0.358}{0.1} + 0.428 \cdot \log_2 \frac{0.428}{0.1} + 7 \times 0.008 \cdot \log_2 \frac{0.008}{0.1}$$

$$H_{POST||PRE-SSA} = 1.45 \text{ bits/process step} \quad (\text{Eq. 8})$$

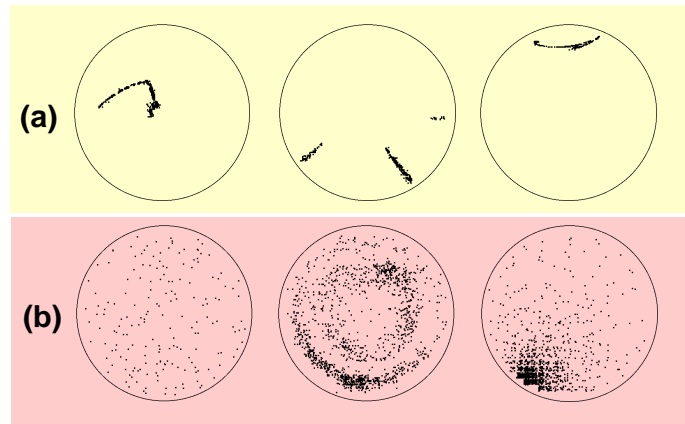


Figure 4. Various types of signatures found by SSA. (a) Mechanically induced signatures caused by scratching, spinning, and handling, and (b) random and systematic defect distributions caused by tools and processes.

6. CONCLUSION

Technology managers in the semiconductor industry need strategies for yield management and fault reduction. Accelerated yield learning gives competitive advantage in R&D environment, which is characterized by radical experience curves. Capital productivity generates competitive advantage in the production environment, where an undetected source of electrical faults can cause enormous losses. The authors have identified “knowledge extraction rate” to be a new success metric for that can serve as a benchmark for the valuation of technologies in both environments. The knowledge extraction has also been quantified by using the Kullback-Leibler formula for relative information entropy, as demonstrated by an estimation of the relative value of four technologies -- electrical testing, wafer position data, automatic defect classification, and spatial signature analysis.

In practice, technology managers in the semiconductor industry are not confronted with an either/or decision between these technologies. Electrical testing on some level is a must, because it is the only technology that identifies sophisticated faults. Inspection-related technologies are also mandatory, because they can identify sources of faults on short notice. Managers would thus like to know the optimal amount of resources to invest in each of the aforementioned technologies. Cost-of-Ownership models, which have been customized for the semiconductor industry over the last 15 years²²⁻³⁰, give managers a very good idea of the true cost of technologies. Managers also typically have ample historical data on the relative frequency of fault types that have plagued their factories. Matching this knowledge to the value metrics established in this paper will allow managers to significantly improve upon the current level of cost/value propositions for yield management and fault reduction.

7. REFERENCES

1. P. Silverman, "Capital Productivity: Major Challenge for the Semiconductor Industry," *Solid State Technology*, Vol. 37, No. 3, p. 104, March 1994.
2. C. Weber, D. Jensen and E. D. Hirleman, "What drives defect detection technology?" *Micro*, June 1998, pp. 51-72.
3. S. H. Thomke, "Managing Experimentation in the Design of New Products," *Management Science*/Vol. 44, No. 6, pp. 743-762, June 1998.
4. V. Sankaran, C. Weber, K. W. Tobin, and F. Lakhani, "Inspection in Semiconductor Manufacturing," to be published in Wiley's Encyclopedia of Electrical and Electronics Engineering (J. G. Webster, ed.) Volume xx (Pattern Analysis and Machine Intelligence), pp. xx-xx, 1999.
5. L. Brillouin, *Science and Information Theory*, (New York: Academic Press, 1956)
6. N. Abramson, *Information Theory and Coding*, (New York: McGraw-Hill, 1963)
7. P. Beckmann, *Probability in Communication Engineering*, (New York: Harcourt, Brace & World, Inc., 1967)
8. S. Haykin, *Neural Networks: A Comprehensive Foundation*, Macmillan College Publishing Company, New York, 1994.
9. R. V. L. Hartley, "Transmission of Information," *Bell System Tech. J.*, 31, pp. 751-763 (1928)
10. C. E. Shannon, "A Mathematical Theory of Communication," *Bell Systems Tech. J.*, 27, pp. 379-423, 623-656 (1948).
11. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, (Urbana: University of Illinois Press, 1949)
12. C. Weber, B. Moslehi and M. Dutta, "An integrated framework for yield management and defect/fault reduction," *IEEE Trans. Semicond. Manufact.*, Vol. 8, no. 2, pp. 110-120, May 1995.
13. Gary Scher, Dennis Eaton, James Sorensen, Barry Fernelius and Jerry Akers, "In-line statistical process control and feedback for VLSI integrated circuit manufacturing," *IEEE Transactions on Components, Hybrids and Manufacturing Technology*, Vol. 13, No. 3, pp. 484-489, September 1990.
14. Gary Scher, "Wafer tracking comes of age," *Semiconductor International*, pp. 126-131, May 1991.
15. U. Kaempf, "Statistical significance of defect density estimates," *IEEE/ICMITS*, 1988, pp 107-113.
16. C. Weber, "Standard defect monitor," *Proc. IEEE/ICMITS*, 1988, pp. 114-119.
17. C. Weber, "A standardized method for CMOS unit process development," *IEEE Trans. Semicond. Manufact.*, Vol. 5, no. 2, pp. 94-100, May 1992.
18. M. H. Bennett, K. W. Tobin, and S. S. Gleason, "Automatic Defect Classification: Status and Industry Trends," *SPIE Metrology, Inspection, and Process Control for Microlithography IX*, Vol. 2439, p. 210, San Jose, CA, March 1995.
19. K. W. Tobin, S. S. Gleason, T. P. Karnowski, M. H. Bennett, "An Image Paradigm for Semiconductor Defect Data Reduction", *SPIE's 1996 International Symposium on Microlithography*, Santa Clara Convention Center, Santa Clara, CA, March 10-15, 1996.
20. S. Cohen, C. Weber, K. Tobin, A. Shapiro, S. Gleason, F. Lakhani, K. Torres and G. Felton-Upton, "Defect detection and analysis at SEMATECH," *Ultra Clean Technology*, vol. 9, no. 5/6, 1997, pp. 308-312.
21. S. S. Gleason, K. W. Tobin, T. P. Karnowski, "An Integrated Spatial Signature Analysis and Automatic Defect Classification System", *191st Meeting of the Electrochemical Society, Inc.*, May 1997.
22. R. Leckie, "A model for analyzing test capacity, cost and productivity," *IEEE International Test Conference*, 1986 pp. 213-218.
23. R. Carnes and M. Su, "Long term Cost of Ownership," *IEEE/SEMI Int'l Manuf. Sci. Symp.* 1991, pp. 39-43.
24. P. DiSessa and S. Stone, "Cost of Ownership for advanced lithography," *Proc. IEEE/SEMI-ASMC* 1991, pp. 54-63.
25. R. Martinez, V. Czitrom, N. pierce and S. Srodes, "A methodology for optimizing Cost of Ownership," *SPIE* Vol. 1803 (1992) pp. 363-387.
26. D. Dance, R. Burghard and R. Markle, "Reducing process equipment cost of ownership through in situ contamination prevention and reduction," *Microcontamination*, May 1992, pp. 21-23 & 64.
27. R. Burghard, D. Dance, R. Markle and T. Silvestri, "Reducing tungsten etch equipment cost of ownership through in situ contamination prevention and reduction," *Microcontamination*, June 1992, pp. 33-36.
28. R. Burghard, D. Dance and R. Markle, "Reducing ion implant equipment cost of ownership through in situ contamination prevention and reduction," *Microcontamination*, Sept. 1992, pp. 27-29 & 63.
29. R. Burghard, D. Dance, R. Markle and T. Silvestri, "Reducing tungsten deposition equipment cost of ownership through in situ contamination prevention and reduction," *Microcontamination*, Oct. 1992, pp. 23-25.
30. C. B. Case and C. J. Case, "Intermetal dielectric cost of ownership," *Semiconductor International*, June 1995, pp. 83-88.