This article was downloaded by: [152.14.136.77] On: 13 August 2015, At: 11:30
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA

**Management Science**

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

# Noise and Learning in Semiconductor Manufacturing

Roger E. Bohn,

**To cite this article:**
Roger E. Bohn, (1995) Noise and Learning in Semiconductor Manufacturing. Management Science 41(1):31-42. http://dx.doi.org/10.1287/mnsc.41.1.31

**Full terms and conditions of use:** **http://pubsonline.informs.org/page/terms-and-conditions**

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

© 1995 INFORMS

**Please scroll down for article—it is on subsequent pages**

INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

# Noise and Learning in Semiconductor Manufacturing

Roger E. Bohn
*School of International Relations and Pacific Studies, University of California, San Diego, La Jolla, California* 92093-0519

Rapid technological learning is critical to commercial success in VLSI semiconductor manufacturing. This learning is done through deliberate activities, especially various types of experimentation. Such experiments are vulnerable to confounding by process noise, caused by process variability. Therefore plants with low noise levels can potentially learn more effectively than high noise plants.

Detailed die yield data from five semiconductor plants were examined to estimate process noise levels. A bootstrap simulation was used to estimate the error rates of identical controlled experiments conducted in each plant. Absolute noise levels were high for all but the best plants, leading to lost learning. For example, the probability of overlooking a three percent yield improvement was above twenty percent in all but one plant. Brute-force statistical methods are either expensive or ineffective for dealing with these high noise levels. Depending on the criterion used, there was a four- to ten-fold difference among the plants.
(*Yield; Learning by Experimentation; Noise*)

## 1. Introduction

Rapid technological learning about manufacturing processes is critical for success in many industries. New process startups require particularly rapid learning. Production volume must be increased rapidly while costs are brought down. In fact the speed and success of the ramp to high volume is determined by the rate at which problems and opportunities on the line are detected, diagnosed, and solved.

This paper presents an empirical examination of the magnitude of process variability and its impact on the rate of process improvement. Examining the case of the semiconductor industry, where issues of rapid learning and process improvement are crucial to success, we develop a simple model of technological learning by controlled experimentation. In this model, process variability creates noise in the experiments, which leads to ambiguous or erroneous results.

It is clear that process variability, by obscuring the true cause and effect relationships in the manufacturing process, makes process improvement and learning more difficult. For example, two plants making the same product but with different process variability will have different functions relating managerial effort to the rate of process improvement, and therefore have learning curves of different slopes.[1] But despite the importance of rapid process improvement in many technology driven industries, process variability and its impacts on learning have received little analytical or empirical analysis. For example, the extensive literature on learning curves is devoid of discussion about process noise as a factor influencing the rate of learning (Dutton et al. 1984).

This lack of discussion implies that process variability is not an important factor in process improvement. This paper presents a preliminary investigation of this issue and concludes that process variability is large enough

[1] See Zangwill and Kantor (1993) for a formalization of the concept of managerial effort versus slope of the learning curve. A more general discussion of learning issues is Jaikumar and Bohn (1992).

0025-1909/94/4101/0031$01.25
Copyright © 1994, The Institute of Management Sciences



Downloaded from informs.org by [152.14.136.77] on 13 August 2015, at 11:30 . For personal use only, all rights reserved.

RIGHTSLINK

Copyright © 2001 All Rights Reserved

Downloaded from informs.org by [152.14.136.77] on 13 August 2015, at 11:30 . For personal use only, all rights reserved.

to have an important impact in the semiconductor industry.

**Prior Work on Yield Variability and Related Topics**
This section reviews literature relating to yield variability in semiconductor manufacturing, and briefly touches on literature on other related fields. Various authors have analyzed the nature of yields in VLSI integrated circuit manufacturing. An important observation in that literature is that the number of defective dice on a wafer does not follow a Poisson distribution, due to spatial clustering of defects. For example, the variance of defects may be ten times the mean, in contrast to the Poisson, which has the variance equal to the mean number of defects (Stapper 1986). In consequence, standard formulas for probabilistic calculations involving yields can be quite erroneous (Stapper 1989). Albin and Friedman propose the use of a Neyman Type-A distribution. They show that it leads to very different acceptance sampling plans (Albin and Friedman 1989) and control charts for detecting out-of-control processes (Friedman and Albin 1991).

Wein and various coauthors investigate the issue of yield variability and its impact on normal plant operations. A plant with constant yield (no matter how low) can be balanced and scheduled with a known ratio of machine capacity at different process stages. In contrast, varying yields can cause shifting bottlenecks and reduce overall plant performance by more than the average yield loss. For example, if a plant is making multiple chip types which are used as a set and sold in fixed proportion, variability in the yields leads to a decrease in the number of good sets produced (Avram and Wein 1992). Sometimes the variability can be turned to advantage. If yield is serially or spatially correlated and if yields are especially low on part of a wafer, it may not even be worth the time to test neighboring wafers or parts of wafers (Longtin et al. 1992, Ou and Wein 1992).

Spanos (1989) analyzes a different type of variability in semiconductor fabrication–measurement error. He shows that ignoring measurement errors can lead to incorrect inferences about process performance. In a sense, this is analogous to the effects of experimental error.

This paper differs from previous work on semiconductor yield variability in two principal respects. First, it is primarily empirical, attempting to establish the magnitude of this problem in a sample of actual plants. Perhaps because of the highly confidential nature of yield data throughout the industry, previous work has been primarily theoretical and has not included empirical measures. Second, it emphasizes the impact of yield variability on learning, rather than on short run operating, cost, or quality issues. It attempts to estimate by how much yield variability makes it more difficult to learn about and improve causes of yield loss. The underlying statistical issues are similar to those for control charts and static performance measures, but the emphasis is on the impact of variability on dynamic performance.

There is also a large literature within the quality control tradition on the causes and effects of process variation. The key point in this literature is that process variation is inherently bad because it leads to out of specification conditions, hurting product quality. Thus quality improvement is in large part a struggle to reduce variability (Deming 1986, Ch. 11). The additional role of variation in creating noise in the learning process is recognized but not emphasized in this literature.

Related to the quality control literature is the statistics literature, which is one of the disciplines underlying process improvement. The statistics literature treats the issue of noise in experiments quite extensively (using different terminology), but pays little attention to the role of process variation in causing that noise. The underlying process variation is taken as given; the role of statistics is to quantify the resulting noise level and to use statistical tools to reduce it (Box et al. 1978, Hogg and Ledolter 1987).

Finally, there is an economics literature on process improvement. Most models of improvement are based on the concept of "the experience curve," which relates declining cost to increases in cumulative production volume. Consistent evidence across many studies and industries shows that the rate of cost improvement (per unit of volume) varies across companies and plants making the same product using the same technology. Dutton and Thomas (1984) survey 200 studies of cost reduction curves over time, and comment that "contrary to widespread assertion, [the slope of the experience curve] depends on firm behavior," i.e., is not determined solely by the technology. Nonetheless, there has been



RIGHTSLINK

Copyright © 2001 All Rights Reserved

Downloaded from informs.org by [152.14.136.77] on 13 August 2015, at 11:30 . For personal use only, all rights reserved.

little effort to study the micro-foundations of the experience curve. Notable exceptions include (Fine 1986, Mody 1989, Muth 1986, Kantor and Zangwill 1991). None of these models includes the effect of process variability.

## 2. Process Description

This section provides background on semiconductor manufacturing, emphasizing issues related to yield variation. VLSI integrated circuit fabrication is one of the most complex of modern manufacturing processes. It involves several hundred process steps conducted on scores of highly automated computer controlled machines. In modern multi-layer fabrication processes there are approximately 1,000 control variables and a comparable number of environmental variables. They interact nonlinearly to determine die yields. Any of these variables could potentially be the target of a process improvement effort.

This study considers five plants, each of which makes multiple products using several processes. In all five plants, the wafers are moved and processed in batches. The standard batch size is 25 or 50 wafers. Batches of different products are interspersed. A large plant may have several thousand batches in process at one time. Plant layouts are functional, and a single expensive piece of equipment is often used at several points in a process and sometimes in different processes. Therefore a machine setup is needed before each batch at most machines.

A key performance indicator in semiconductor fabrication is *yield*. Blank wafers are used as the base for creating circuits. A single wafer may have ten to one thousand identical circuits, called dice. *Line yield* is the fraction of wafers which go all the way through the process without being irreparably damaged by breakage, gross processing errors, or other problems. Line yields are dependent on good plant-wide operating practices, especially in material handling. One hundred percent testing is done at the end of fabrication. The percent of the original dice which function correctly is defined as the *die yield*, also called *probe yield*. Die yields depend on product design, initial process design, manufacturing practice, and cumulative learning/improvement since the process started. Overall yield is the product of die yield and line yield.

The basic economics of fabrication are simple: most costs are fixed, while output is proportional to wafer starts times the final yield.[2] Therefore yields are crucial to profitability. This is magnified early in the life of new products, when the product often sells for a premium price, due to limited production volume and better performance than earlier products. Wafer starts depend mainly on production capacity of bottleneck equipment, while die yields depend on initial process design and manufacturing practice, and cumulative learning/improvement since the process started. Overall die yields are multiplicative, since the dice must survive every one of a number of serial processes.

Yield data is highly proprietary, and little hard information has been published. "One of the most critical, elusive, guarded and controversial aspects to analyze in the semiconductor process is probe yield" (McClean 1985). Freeze et al. (1984) indirectly document a factor of 14 improvement in overall yield over two years for one product, and a factor of 7 improvement for another. These improvements were measured starting from the earliest stages of pre-prototype production, whereas the plants in this study were starting from a much more mature level. Wasserman and Clark (1986) describe a case in which overall yields of a new product were close to zero for several months.

### Yield Improvement

Early in the life of a new process, die yields are usually very low due to a number of problems. They improve as engineers systematically discover these problems, then change the process to reduce or eliminate them. Changes are made to methods (such as changing maintenance procedures on a piece of equipment; changing a fixture) or to the process recipe, (such as dose, temperature, and time for a particular step, material supplier or specifications) or occasionally to the design mask for a layer. Any change can have ramifications for other problems and parts of the process.

Because of the complexity of the process and the potential for unforeseen side effects, most process improvements are made carefully and systematically,

[2] The back end of VLSI semiconductor manufacturing is *encapsulation*, in which individual dice are turned into chips. The economics of encapsulation are somewhat different, and will not be considered in this paper.



RIGHTSLINK
Copyright © 2001 All Rights Reserved

Downloaded from informs.org by [152.14.136.77] on 13 August 2015, at 11:30 . For personal use only, all rights reserved.

under the direction of engineers. The basic test or validation of a proposed change is an "engineering trial." Typically an engineering trial is conducted as *split batch experiment*, which compares two production methods by making some wafers according to each method. A regular production batch is split in half just before the step where the change is to be made. Half the wafers in the batch are processed in the conventional way, and half according to the proposed new recipe. The split halves are recombined and processed normally through the rest of the plant. At the end, the individual wafers are measured and the average measurements for each of the split batches calculated. Differences in the averages are due to the different recipes they went through, plus the effects of noise. Note that this split lot procedure provides blocking for most of the between-batch noise, but does not block any of the within-batch noise.

A single product often runs at 500 wafer starts per week, 25,000 per year, or even higher. Wholesale value of a single wafer is in the range of $1,000 to $10,000, depending on a variety of factors.[3] Thus an experiment which gives a one percentage point yield improvement for one product can be worth approximately 250 wafers per year, or $250,000 to $2.5 million. In fact, most yield improvements after the first year are in the range of 0.5 to three percent, not larger.[4]

Experiments to improve yield can be confounded by experimental noise, which can arise due to inherent

[3] Dehmel and Parker (1987) estimate $10,000 per wafer in the future. Zorich (1991, ch 2.2), presents sample calculations for the value of yield improvement on different products.

[4] Personal interviews with factory engineering managers. This is confirmed by the following simple calculation. A factory which runs only 50 experiments per year must have an average improvement per experiment of 0.004, if the annual rate of yield improvement is 20 percent per year. Experiments with negative estimated improvements will be truncated, so if half of all experiments lead to negative or no results, the average size of the remaining experiments would be 0.008.

An article about a prototype production facility at Hewlett Packard gave a higher estimate:

> It is estimated that an experiment designed to improve a process will most likely result in a yield improvement from −2.5% to 12.5% (*percent, not percentage points*). It can be further estimated that some small, but positive yield improvement (typically 5%) for each completed engineering experiment is the most probable result (Brooksby et al. 1981).

Since average yields are much lower, noise levels much higher, and numbers of experiments much higher in prototype facilities than in production facilities, this estimate is consistent with the lower numbers used for the production factories in this paper. More empirical analysis would be useful.

**Table 1 Summary of Data Sources**

| Plant code | Product Maturity (approx) | Name for Data Set in Tables | Number of Batches | Comments |
|---|---|---|---|---|
| A | 1.5 years | A1.5 | 5 | Same product as Plant G |
| | 3.5 years | A3.5 | 5 | |
| C | 1 year | C1 | 11 | |
| | 1.5 | C1.5 | 9 | |
| | 2 | C2 | 10 | |
| | 2.5 | C2.5 | 10 | |
| | 3 | C3 | 12 | |
| B | Unknown | B | 13 | |
| F | 1 year | FF | 8 | |
| G | Pre-qualify | G | 6 | Same Product as A |

process variability, measurement error, or experimental error (e.g. applying the wrong procedure). This paper analyzes process variability and its impact on experimental noise. Sources of process variability can be classified in various ways. Random point defects caused by both ambient and machine-generated particles are usually a major and highly variable cause of yield loss (Zorich 1991, ch. 3). Other sources of batch to batch and wafer to wafer variation include operator adjustments, machine wear, machine maintenance, chemical contamination, and outright processing errors. Note that while many of these hurt mean yield as well as raise the variance of yield, the two effects are not perfectly correlated. Therefore a noise reduction strategy is not identical to a yield improvement strategy.

## 3. The Magnitude of Process Noise

The magnitude of process noise in actual semiconductor fabrication was investigated empirically. Five plants provided production yield data on one product each. Two of the plants provided data for multiple time periods (Table 1). Each was a high volume, multiproduct MOS fabrication facility. All except G were U.S. plants in a single company; G was a foreign subcontractor.



RIGHTSLINK

Copyright © 2001 All Rights Reserved

Downloaded from informs.org by [152.14.136.77] on 13 August 2015, at 11:30 . For personal use only, all rights reserved.

Plants A and G made the same product using the same process. All of the products were medium to high volume, where high volume is thousands of batches and millions of completed chips per year.

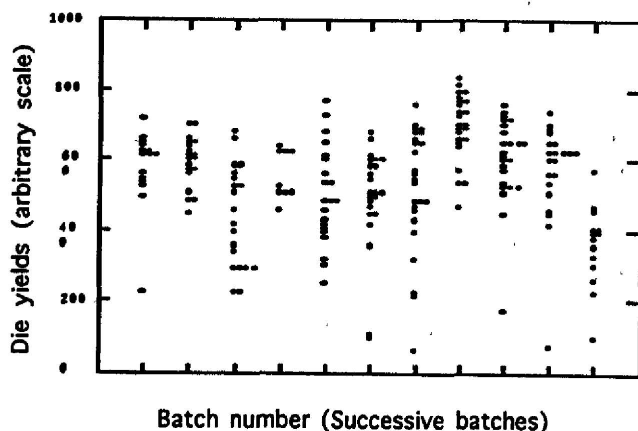
The data are very disaggregated. The data consist of wafer by wafer die yield counts (good dice per wafer), for every wafer in the each batch. Thus the data give a precise measure of die yield and line yield. This is the most detailed yield data normally recorded in the plant.

Data were provided by individual engineers in each plant. In at least one case (plant A), this biased the data since the engineer who selected the data chose from batches with high line yields. Thus the results for this plant will understate the noise levels of experimental batches. In plant FF, a number of anomalous batches, which appeared to be experimental batches, were excluded. One batch was excluded for the same reason in plant C. One entire period of production (4 batches) was excluded for plant A because of a number of anomalies. This was by far the highest noise plant/period in the data set; including it would therefore have strengthened the conclusions further.

Section 3 presents basic descriptions of the noise as revealed by the data. The following section translates the noise into meaningful consequences. All absolute yield data are disguised; only data on noise can be fully presented.

Figure 1 shows dot plots from plant C1 (plant C, one year after the beginning of production for that product). All batches completed the production process consecutively during the same week, and are for the same product in the same plant. Each column represents one batch, while each dot represents the yield of one wafer in that batch. The yields are arbitrarily scaled to protect confidentiality; 1,000 on the graph does not correspond to any particular yield level. Within-batch variability in die yield is the spread of each column. Between-batch variability is the difference among the columns.

**Figure 1 Die Yields of Individual Wafers from Plant C1**




The range of shapes shown in Figure 1 is surprising, considering that all batches were produced under what should have been identical conditions. The mean yields, as well as the variance and skewness of yields vary from batch to batch, suggesting that the underlying production process was not stable. A Bartlett test for homogeneity of group variances gave probability less than 0.005% that all ten batches from C1 had the same variances.

Final die yield is a multiplicative process. A given process change is likely to *multiply* the die yield of each surviving wafer by a constant, rather than *adding* a constant. If the process change is favorable, the yield multiplier will be greater than 1.0. Therefore, it is appropriate to use the natural logarithm of die yield as the measure of experimental outcome. For example, a change of + 0.05 in log[die yield] means a 5.1 percent improvement in average die yield. The standard deviation of log[die yield] measures the percent variability of the yield. It is approximately equal to the coefficient of variation of absolute yields. It will be referred to as the "within-batch noise level."

Because of the way experiments are designed and conducted, the within-batch standard deviation of production die yields will be shown to have the largest impact on experimental noise. Figure 2 summarizes this measure for all five plants. Each column corresponds to one of the plant/time combinations in Table 1. Each entry is the standard deviation of log die yields of a single batch. The simple average across all batches of the within-batch standard deviations is indicated by the x mark.

Based on Figure 2 we can make the following observations.

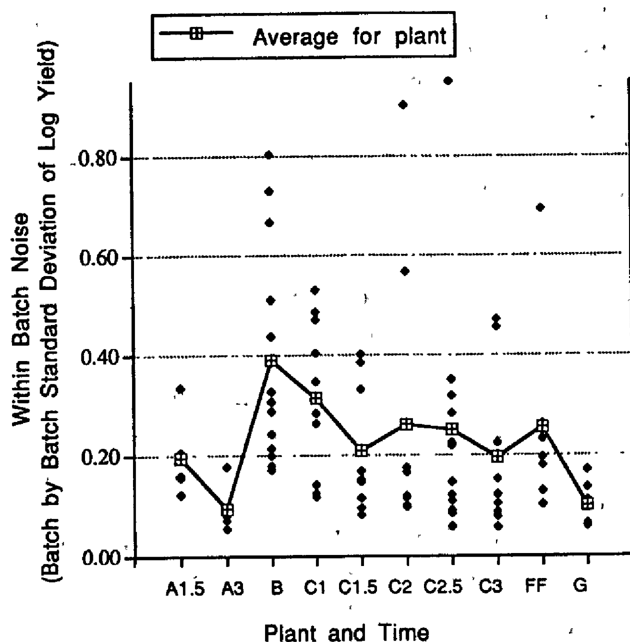
Observation 1. *Most batches within most plants have high levels of within-batch variability, relative to the effects being learned about*. In some batches, the standard deviation of yields is half the mean yield. Considering



Copyright © 2001 All Rights Reserved

Downloaded from informs.org by [152.14.136.77] on 13 August 2015, at 11:30 . For personal use only, all rights reserved.

**Figure 2** **Summary of Noise by Batch and Plant**




that the magnitude of process improvement being sought is less than 0.05, this will be a significant impediment to learning.

OBSERVATION 2. *Noise levels vary considerably across plants and time*. Plants B, C1, and FF have average noise levels more than double those in the best plants. The only processes with moderate noise levels were a mature U.S. plant (A3), and a startup process in an existing non-U.S. plant (G).

OBSERVATION 3. *The noise level varies greatly across batches in each plant*. This is in addition to high batch-to-batch variation in mean yields. This suggests that statistical tests on experiments should use "unknown variance" models, rather than assuming that variances are known from past data.

OBSERVATION 4. *Line yields (surviving wafers per batch) cannot be shown directly but also varied considerably among the batches and between plants*. Average line yields in each plant varied from below 80 percent to above 95 percent.

These observations are consistent with manufacturing processes which are not under good process control. Whatever the causes, it is likely these factors will create high noise levels in experiments.

## 4. Effects of Process Noise

This section examines the effects of process noise on learning, by simulating the conduct of experiments using the data described in the previous section. Note that we are using normal production data to simulate the conduct of experiments. An alternative data collection strategy would be to use actual experiments conducted at each plant to measure noise levels. The use of normal production yield data allows large homogenous data sets which are directly comparable across plants. It allows us to examine a consistent hypothetical experiment in each plant and time period. Actual experiments, in contrast, are highly idiosyncratic. It would be difficult to collect enough similar experiments in a single plant to give a reliable estimate of the noise level at a particular time. In addition cross-plant comparisons would be more suspect, in part because standard experimental procedures may have subtle variations across plants.[5]

The underlying model of manufacturing which we will assume is an additive independent model in the log of yields:

$$Y_{new} = Y_{old} + \Delta Y + \epsilon \quad \text{where} \tag{1}$$

$Y_{new}$ = die yield after the process change.
$Y_{old}$ = original die yield of the process.
$\Delta Y$ = change in average die yield as a result of the experimental treatment (positive or negative). Larger is better.
$\epsilon$ is the noise in the die yield.

All quantities are measured in natural logarithms. This assumes that the process change does not affect the process noise level.[6]

### Methodology

Learning is modeled as occurring through simple split batch experiments as described in §2. Each experiment consists of $2N$ wafers, $N$ of which receive the experimental treatment at the critical process step. The $2N$ wafers are processed as a single batch at all other process

[5] A distortion introduced by use of normal production data may be to reduce measured noise levels. Experimental batches require special handling at a few steps. This is error prone; one informant estimated that up to one-third of special batches were misprocessed.

[6] This is a reasonable approximation for all but the largest changes since total noise is the sum of noise in number of steps, while process changes affect only one or a few steps.



RIGHTSLINK
Copyright © 2001 All Rights Reserved

Downloaded from informs.org by [152.14.136.77] on 13 August 2015, at 11:30 . For personal use only, all rights reserved.

steps. The standard test statistic for such an experiment is the difference in average yield between the two split groups. The larger the difference, the larger the likely process improvement from the new method. The test statistic is

$$\Delta Y_{est} = \left\{\frac{1}{N_1}\sum_{i=1}^{N_1} Y_i\right\} - \left\{\frac{1}{N_2}\sum_{i=1}^{N_2} Y_{N+i}\right\} \qquad (2)$$

which deviates from the true effect of the treatment according to

$$\Delta Y_{est} = \Delta Y_{true} + \epsilon_{experiment} \quad \text{where} \qquad (3)$$

$\Delta Y_{est}$ is the estimated effect of the new production method.

$\Delta Y_{true}$ is the unknown true effect of the new production method.

$Y_i$ is the yield of the $i$th wafer. The first $N$ wafers are the experimental group; the next $N$ are the control group.

$N$ is the initial sample size of each split group in the experiment. $N \leq 12$ since batch size in most plants is 25.

$N_1$ is the number of wafers which survive in the experimental group; $N_1 \leq N$.

$N_2$ is the number of wafers which survive in the control group; $N_2 \leq N$.

$\epsilon_{experiment}$ is the noise of the experiment, which depends on the process noise level and the number of wafers in the experiment.

If die yields were distributed normally or according to another known distribution, and if line yields were 100 percent so that $N = N_1 = N_2$, we could use statistical theory to find the distribution of the experimental noise $\epsilon_{experiment}$. Albin and Freedman argue persuasively that defects and yields will follow a complex compound distribution with clustering behavior. They recommend using the Neyman Type 1 distribution for defects, which is a compound Poisson distribution that exhibits clustering. However, the Neyman distribution does not appear to fit well the actual empirical data described above. Furthermore, using any single distribution to summarize the actual wafer-by-wafer data is risky, since the batch-to-batch comparisons suggest that the manufacturing process parameters were not stable. Finally, the impact of missing wafers caused by line yield losses must be incorporated. This reduces the effective sample size below the nominal sample size $N$.

To evaluate the effectiveness of these experiments without assuming an underlying distribution function for die yields, standard bootstrapping techniques were used to simulate what would have happened if experiments had been conducted on these batches in each plant. (Efron and Gong 1983, Cryer and Miller 1991, Chapter 19). The wafer-by-wafer die yields from a single batch, discussed in §3, were repeatedly sampled with replacement, to construct the two split batches of $N$ wafers each which would result from a single experiment. Wafers which did not survive the line yield were removed from each subsample. The test statistic, $\Delta Y_{est}$ (difference of the average log yields), was then calculated for the case that $\Delta Y_{true} = 0$ (i.e. an experiment on a process change which has no effect). This gives the outcome of a single simulated experiment. This procedure was repeated 600 times for experiments with $N = 12$, and 2,000 times for experiments with $N = 3$. Sampling was conducted equally from each batch of a particular plant/time period. Symmetry was then used to double these sample sizes to 1,200 and 4,000 respectively. These 5,200 simulated experiments per plant/period form the basis for evaluating the error rates of real experiments in the plants.

We will start with the simplest possible test criterion. If $\Delta Y_{est} > 0$, treat the new production method as better; otherwise, stay with the old production method. This decision rule is optimal only if the costs of false positives and false negatives are symmetric, which is unrealistic, but it has other useful properties. More complex decision rules will be discussed later.

From the bootstrap data, we construct the power function $G(\Delta Y)$ of the hypothetical experiment. $G(\Delta Y)$ = Probability of choosing the new production method, if the true value of the change is $\Delta Y$. The power function gives a complete measure of an experiment's information content, and can be used to evaluate the experiment according to any criterion, such as significance regions. A ideal power function would rise steeply through $\Delta Y = 0$, so that for $\Delta Y < 0$ the old method would be chosen most of the time, and conversely for $\Delta Y > 0$.

Using the test statistic of Eq. (2), $G(z)$ is given by $\text{Prob}(\Delta Y_{est} < z / \Delta Y_{true} = 0)$. The impacts of process improvement and process variation are independent ac-



Copyright © 2001 All Rights Reserved

**Table 2 Consequences of Within-Batch Noise**

| Plant name<br>Avg. in-batch noise | A1.5<br>0.196 | A3<br>0.094 | B<br>0.391 | C1.0<br>0.315 | C1.5<br>0.209 | C2.0<br>0.261 | C2.5<br>0.251 | C3.0<br>0.196 | FF<br>0.256 | G<br>0.100 |
|---|---|---|---|---|---|---|---|---|---|---|
| Probability of Missing Process Improvements: | | | | | | | | | | |
| $N = 12$, True $\Delta Y = .01$ | 45.5% | 37.5% | 47.0% | 47.0% | 45.0% | 43.5% | 43.5% | 43.5% | 45.5% | 40.5% |
| $N = 12$, True $\Delta Y = .03$ | 36.0% | 18.5% | 40.5% | 40.0% | 32.5% | 32.5% | 32.0% | 28.5% | 37.0% | 24.5% |
| $N = 12$, True $\Delta Y = .10$ | 11.5% | 2.5% | 22.5% | 19.5% | 11.5% | 12.0% | 12.5% | 9.0% | 16.0% | 4.0% |
| $N = 3$, True $\Delta Y = .10$ | 22.5% | 6.5% | 34.0% | 29.0% | 20.0% | 22.5% | 21.0% | 17.0% | 25.0% | 10.5% |
| Smallest Effect Which Can Be Found with Error Prob $\leq$10%: | | | | | | | | | | |
| for $N = 12$ | 0.109 | 0.044 | 0.216 | 0.188 | 0.111 | 0.114 | 0.117 | 0.092 | 0.129 | 0.057 |

*Note.* Lower numbers are better, for all rows.

cording to Eq. (1). Therefore, we can construct the power function for all possible values of $\Delta Y_{true}$ just by bootstrapping the case $\Delta Y_{true} = 0$, and shifting the $G(z)$ function to the right by $\Delta Y_{true}$. In this way, even though the bootstrap only explicitly evaluates experiments where $\Delta Y_{true} = 0$, we can use the power function to calculate what would have happened for different values.

## 5. Results: Effect of Noise on Experimental Outcomes

Recall from §2 that learning in semiconductor manufacturing proceeds on the basis of multiple small improvements, in the neighborhood of $.01 \leq \Delta Y \leq .03$. This is the size of the signal being sought by the engineer, which is much smaller than the within-batch standard deviations of 0.10 and above, found in the empirical data.

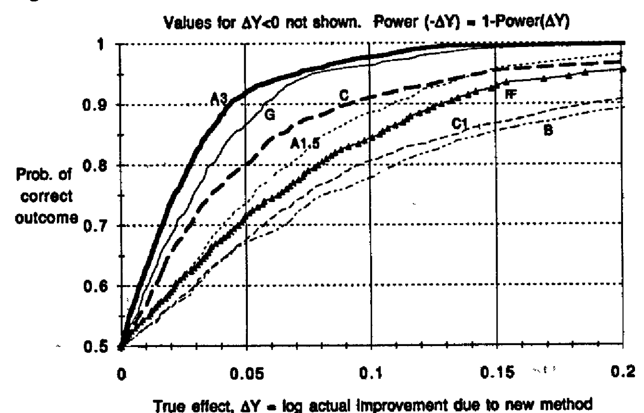
Table 2 shows some of the effects for each plant. Based on information presented in Table 2, we can make the following additional observations:

Observation 5. *The impacts of noise in most plants were so large as to make the chance of overlooking process improvements (Type 2 errors) quite high, except for very large improvements.* To find a process change which has a ten percent effect ($\Delta Y_{true} = 0.10$) is quite rare. But in plant B, even such a large effect would be missed in an experiment more than 20% of the time, even with a sample size of $N = 12$. Only plants A3, G, and C3 have probabilities of error below 10% for a change of 0.10. Experiments on process changes with $\Delta Y = 0.03$ have error rates ranging from 18% to 40%. As this model is formulated, 50% is the highest possible error rate, so plants B and C1 do little better than pure chance. None of the plants does much better than pure chance for $\Delta Y = .01$.

Observation 5b. *All results are considerably worse for experiments conducted with samples of* $N = 3$. In fact the noise levels are so high that all experiments should be run with $N \geq 12$.

Observation 6. *As with the noise levels themselves, the consequences of noise differ considerably across plants and time.* For most measures of learning performance, there is roughly a 4:1 ratio between the best plants and the worst plants. Plants A3 and G were generally the best, while B and C1 were the worst. Correlation analysis confirms that the performance measures are closely

**Figure 3 Power Functions for $N = 12$**

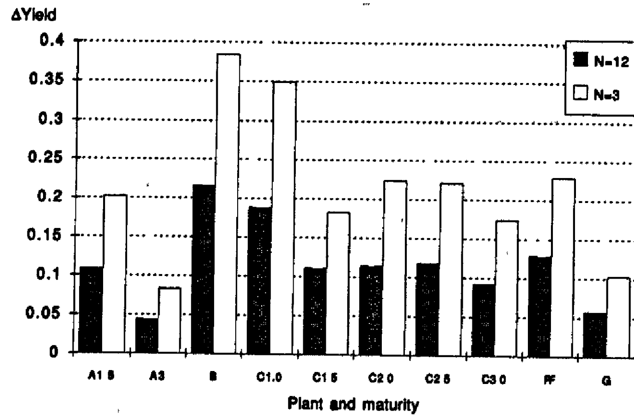





Downloaded from informs.org by [152.14.136.77] on 13 August 2015, at 11:30 . For personal use only, all rights reserved.
Copyright © 2001 All Rights Reserved

Downloaded from informs.org by [152.14.136.77] on 13 August 2015, at 11:30 . For personal use only, all rights reserved.

**Figure 4 Smallest Change (Δ Yield) Detectable with 10% Probability of Error**




related to average within-batch noise levels (standard deviation of log die yields).

The numbers in Table 2 were derived from the power functions of our hypothetical experiments, as calculated from the bootstrap results. The power function for experiments at each plant provides complete information needed to analyze the tradeoffs among signal size $\Delta Y_{true}$ and probability of error. Figure 3 shows the power functions for all plants for selected time periods for full batch experiments of $N = 12$ wafers per sample. Each line shows the probability of accepting the new production method as a function of its true effect on yield $\Delta Y_{true}$, which is unknown to the experimenter. For any true value $\Delta Y$ of log yield improvement, the height of the power function is the probability that the engineer will accept the hypothesis that $\Delta Y_{true} > 0$, i.e. that the new method is better.[7] Probabilities of accepting inferior new methods ($\Delta Y_{true} < 0$) are given by symmetry.

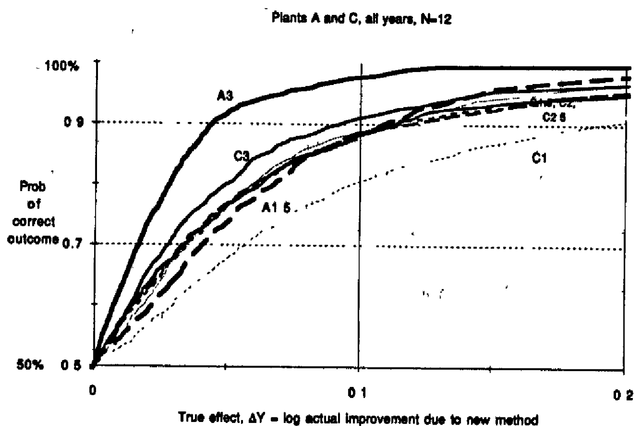
For example, in plant FF, if the true value of a process change is $\Delta Y = .03$ (a 3 percent improvement in die yield), the probability of accepting the new method is 63 percent, and the probability of rejecting it (type 2 error) is about 37 percent. Each power function is symmetric and passes through ($\Delta Y = 0$, prob. = 50%) because we modeled the engineer as using a symmetric test criterion. If the new method were in fact worse, with $\Delta Y_{true} = -.03$, the probability of rejection would be 63 percent and the probability of acceptance (type 1 error) would be 37 percent. If the engineer sets a cutoff of $\Delta Y_{est} \geq \Delta Y_{cutoff} > 0$, in an effort to defeat the effect of noise, this would shift each curve to the right by $\Delta Y_{cutoff}$. These probabilities are conditional on the experiment being successfully completed. There is also a 13 percent chance the experiment would have to be repeated due to too many wafers being lost in the production process (line yield). This is not shown in Figure 3.

OBSERVATION 7. *Even with the largest possible single-batch experiments ($N = 12$), it is difficult for most plants to meet standard statistical criteria for any but the largest process changes. Small sample experiments ($N = 3$) should be avoided in all plants.*

Figure 4 shows one of the consequences of the power functions: the true process change magnitude such that each plant can reliably detect that signal, where "reliably" is defined as less than a 10% chance of error. The importance of setting a nominal sample size of $N = 12$ instead of $N = 3$ is clearly visible. Even with $N = 12$, most plants have detection thresholds higher than 0.10, and one is higher than 0.20.

There are large differences in the effects of noise among plants. These differences are large enough to

**Figure 5 Changes over Time for Plants A and C**




[7] Sampling error: Since Figure 3 and associated tables are derived from bootstrap analysis with 1200 repetitions, they are subject to sampling errors. Since they are order statistics, these errors are largest in the tails. A visual assessment of the error is given by the waviness of the lines in Figure 3; the true power functions are smooth. I also re-ran plant C1, 10 times. For G ($\Delta Y = .10$), the runs had mean and s.d. of .793 ± .010.



Copyright © 2001 All Rights Reserved

Downloaded from informs.org by [152.14.136.77] on 13 August 2015, at 11:30 . For personal use only, all rights reserved.

**Table 3** Number of Batches Required to Overcome Noise (Approximate)
B needed to have 90% chance of detecting true process improvement of size 0.01

| Plant Name | A1.5 | A3 | B | C1.0 | C1.5 | C2.0 | C2.5 | C3.0 | FF | G |
|---|---|---|---|---|---|---|---|---|---|---|
| Number of Batches | 148 | 41 | 630 | 444 | 221 | 392 | 394 | 184 | 365 | 50 |

indicate that different plants should use different strategies and tactics for ameliorating the effects of noise. Even within one plant, conditions change over time. Figure 5 shows the time trends for plants A and C. Although Plant C made great progress between year 1 and year 1.5, years 2.0 through 3.0 were almost stagnant. These results suggest that the progress of noise over time is a complex phenomenon.

### Brute Force Solution to Noise

The orthodox statistical approach to dealing with high noise levels in experimental data is to increase the sample size. We have already seen that even full-batch $N = 12$ experiments are not large enough in many situations. To go beyond $N = 12$ requires multi-batch experiments, which are more expensive, slower (due to variability in batch completion times), and more prone to errors than single batch experiments. Nonetheless they may be economically justified, since the economic value of yield improvement is so large.[8]

We evaluate the following problem. Suppose an engineer wants to run an experiment which has a 90% chance of detecting a process improvement of $\Delta Y_{true} = 0.01$. He or she decides to run a multibatch experiment for this purpose. How large should the experiment be? We can approximate the answer as follows. Let $B$ be the number of batches in the experiment; each batch has $N = 12$. Let $\Delta Y_B$ be the average of the $\Delta Y_{est}$ estimates calculated for each batch according to Eq. (2).[9] Then if $B$ is large, the Central Limit Theorem implies that $\Delta Y_B$ is approximately Normal, with variance proportional to $1/B$. Making the appropriate calculations gives the (approximate) experiment sizes shown in Table 3.

Observation 8. *Increasing the sample size to overcome noise is expensive in all plants*. For example, if the product runs at 500 wafer starts per week (20 batches), in plant $A1.5$ seven consecutive weeks of production would have to be devoted to the experiment, or fourteen percent of a year's production. This is slow and expensive, especially if the experimental treatment has a significant chance of reducing yields or production capacity.

Observation 9. *There is an order of magnitude difference across plants in the sample sizes needed to achieve the same level of statistical certainty, if no other countermeasures are taken.*

Once again, these observations point to the value of noise reduction and other approaches to mitigating noise's impacts on learning.

## 6. Conclusions

The empirical analysis of this study shows that process variability, leading to noise in experiments, is high enough to have a major influence on how efficiently engineers in semiconductor plants can learn by controlled experiments. In most of the plant/product/maturity combinations studied, the noise levels were so high that full length controlled experiments, using an entire production batch, would have error rates above ten percent, even for very large process improvements. A conservative estimate is that most of the plants studied therefore lose more than one quarter of the potential information content of such experiments.

There were large differences among the plants, which indicates that different plants should manage experimentation differently. Comparison of two plants making

[8] The optimal choice of sample size requires an analysis of the incremental value of additional knowledge (reduced probability of error) versus the incremental cost of a larger sample size (lost output in the short run due to the experiment).

[9] Calculating the test statistic in this way provides blocking for the large between-batch noise in the data. An even better test statistic would take advantage of the diversity of within-batch noise levels, by weighting more heavily those batches with smaller estimated within-batch noise levels.



RIGHTSLINK

Copyright © 2001 All Rights Reserved

Downloaded from informs.org by [152.14.136.77] on 13 August 2015, at 11:30 . For personal use only, all rights reserved.

the same product, and the anecdotal evidence about the origins of noise in the process and in experiments, suggests strongly that overall reductions in process noise levels are possible. However, even in the best plants of the sample, the noise levels are high enough to cause a significant number of errors.

The textbook antidote to noise in experimentation is the use of statistical tools. Yet low and erratic use of statistical methods by engineers was observed in visits to plants in several companies. This can be partly explained by the finding that noise levels were so high that standard statistical tools (such as increased sample sizes) would be insufficient or expensive. Non-statistical methods are therefore desirable in conjunction with statistical tools. These include using short-loop experiments, natural experiments (e.g. wafer tracking data (Scher 1991)), and changing to experiments which measure causal factors as the dependent variable instead of measuring yield directly. Such methods require additional knowledge compared with split-batch experiments, so they are not always possible.

Appropriate statistical methods including factorial designs, sequential experimentation, and Bayesian inference, would still be useful for capturing as much information as possible per unit of effort. Their low use may therefore also be due to ignorance. More recent visits to the same plants show growing awareness of standard statistical tools, and some awareness that learning is a process which can itself be managed. Hiatt and Urquhart (1987) describe a directed learning project in Motorola which used multiple statistical and non-statistical methods, and paid careful attention to noise levels.

Among the issues not dealt with in this paper are the root causes of the high noise levels, and the reasons why noise is not addressed more directly and solved by plant management. Noise in final die yields is in principle due to specific machine/process variability at specific process steps. But the large differences in noise among plants using similar equipment and processes suggest deeper levels of causality. Phenomena of this complexity cannot be captured by studying experimentation in isolation from other aspects of plant operation.

In both plants for which time series data are available, the noise levels fall over time, as would be expected if some of the learning is about factors which themselves cause noise. Therefore the management of noise-related issues in experimentation should change over time. The noise level is most critical early in the ramp up, making learning more difficult. Yet this is when the most knowledge needs to be learned. This suggests a time dependent model, in which early process improvement should emphasize noise reduction more than yield enhancement, when the two are in conflict. As the noise level falls, it becomes easier to discover new improvements.[10] On the other hand, the best improvements will tend to be discovered first, so that the $\Delta Y$ s will also fall over time. The signal-to-noise ratio ($\Delta Y$/noise level) would thus initially improve, but ultimately stabilize or worsen as smaller and smaller effects remain to be discovered.[11]

[10] The decline over time in plant C's noise level was not monotonic. As Table 3 shows, it fell by one third from year 1.0 to 1.5. But it worsened from year 1.5 to year 2.0, and was stagnant through year 2.5. In year 3.0 it improved to about its level at year 1.5. Interestingly, the actual die yields themselves were improving throughout the two year period, although slowly. This suggests that more effort was being put into improving average yield than into improving yield variability. The specific die yield numbers cannot be reported due to confidentiality.

[11] Valuable comments on versions of this paper were provided by Christopher Auschnitt, John Bishop, James Cook, Richard Dehmel, M. Thérèse Flaherty, R. Jaikumar, Uday Karmarkar, Andy Urquhart, Larry Wein, Roy Welsch, two referees, and seminar participants at Harvard and Boston Universities. Michael Watkins assisted with data analysis. I remain solely responsible for the paper's omissions and any remaining errors. Author's Internet address: Rbohn@UCSD.edu.

## References

Albin, Susan and David J. Friedman, "The Impact of Clustered Defect Distributions in IC Fabrication," *Management Sci.*, 35, 9 (1989), 1066–1078.

Avram, Florin and Lawrence M. Wein, "A Product Design Problem in Semiconductor Manufacturing." *Oper. Res.* 40, 5 (1992), 986–998.

Box, George E. P., William G. Hunter, and J. Stuart Hunter, *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, John Wiley & Sons, NY, 1978.

Cryer, Jonathan D. and Robert B. Miller, *Statistics for Business: Data Analysis and Modelling*, PWS-Kent Publishing Co., Boston, MA; 1991.

Dehmel, Richard C. and Gerry H. Parker, "Future VLSI Manufacturing Environment," *Solid State Technology*, 1987, 115–121.



RIGHTSLINK

Copyright © 2001 All Rights Reserved

Downloaded from informs.org by [152.14.136.77] on 13 August 2015, at 11:30 . For personal use only, all rights reserved.

Deming, W. Edwards, *Out of the Crisis*, MIT, Center for Advanced Engineering Study, Cambridge, MA, 1986.

Dutton, John and Annie Thomas, "Treating Progress Functions as a Managerial Opportunity," *Academy of Management Review*, 9, 2 (1984), 235–247.

——, ——, and John E. Butler, "The History of Progress Functions as a Managerial Technology," *Business History Review*, Summer 1984, 204–233.

Efron, Bradley and Gail Gong, "A Leisurely Look at the Bootstrap, the Jackknife and Cross-Validation," 37, 1 (1983), 36–48.

Fine, Charles, "Quality Improvement and Learning in Productive Systems," *Management Sci.*, 32, 10 (1986), 1301–1315.

Freeze, Karen J., Neil Wasserman, and Kim B. Clark, *Seeq Technology*, 1984, Harvard Business School, 1984. HBS Case 9-685-081.

Friedman, David J. and Susan Albin, "Clustered Defects in IC Fabrication: Impact on Process Control Charts," *IEEE Transactions on Semiconductor Manufacturing*, 4, 1 (1991), 36–42.

Hiatt, Mark and Andy Urquhart, "Experimental Technique for Resist Process Evaluation," *Semiconductor International*, 1987, 146–151.

Hogg, Robert V. and Johannes Ledolter, *Engineering Statistics*, New York; Macmillan Publishing Co., 1987.

Jaikumar, Ramchandran and Roger E. Bohn, "A Dynamic Approach to Operations Management: an Alternative to Static Optimization," *International J. Production Economics*, 27, 3 (1992), 265–282.

Kantor, Paul B. and Willard I. Zangwill, "Theoretical Foundation for a Learning Rate Budget," *Management Sci.*, 37, 3 (1991), 315–330.

Longtin, Mark D., Lawrence M. Wein, and Roy E. Welsch, *Sequential Screening in Semiconductor Manufacturing II: Exploiting Spatial Dependence*, MIT, Cambridge, MA, 1992.

McClean, William J., *Overview of Yield Projection Techniques*, Integrated Circuit Engineering Corporation, Scottsdale, AZ, 1985, Icecap report.

Mody, Ashoka, "Firm Strategies for Costly Engineering Learning," *Management Sci.*, 35, 4 (1989), 496–512.

Muth, John F., "Search Theory and the Manufacturing Progress Function," *Management Sci.*, 32, 8 (1986) 948–962.

Ou, Jihong and Lawrence M. Wein, *Sequential Screening in Semiconductor Manufacturing, 1: Exploiting Lot-to-Lot Variability*, Sloan School, MIT, Cambridge, MA, 1992.

Scher, Gary M., "Wafer Tracking Comes of Age," *Semiconductor International*, 1991, 126–131.

Spanos, Costas J., "Statistical Significance of Error-Corrupted IC Measurements," *IEEE Transactions on Semiconductor Manufacturing*, 2, 1 (1989), 23–28.

Stapper, C. H., "The Defect-Sensitivity Effect of Memory Chips," *IEEE J. Solid-State Circuits*, SC-21, 1 (1986), 193–198.

——, "Fact and Fiction in Yield Modeling," *Microelectronics J.*, 20-1, 1-2 (1989), 129–151.

Wasserman, Neil H. and Kim B. Clark, *Everest Computer*, Harvard Business School, 1986, case study 685-085.

Zangwill, Willard I. and Paul B. Kantor, *Toward a Theory of Continuous Improvement*, GSB, University of Chicago, Chicago, IL, 1993.

Zorich, Robert, *Handbook of Quality Integrated Circuit Manufacturing, San Diego, Academic Press, 1991.*

*Accepted by Luk Van Wassenhove; received May 1992. This paper has been with the author 2 months for 1 revision.*



RIGHTSLINK

Copyright © 2001 All Rights Reserved