Learning during Ramping:

Policy Choices for Semiconductor Manufacturing Firms

by

Richard H. Benfer

S.B. Materials Science and Engineering, Massachusetts Institute of Technology (1984)
S.M. Materials Engineering, Massachusetts Institute of Technology (1986)

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

Master of Science in Management

in conjunction with the

Leaders for Manufacturing Program

at the Massachusetts Institute of Technology

May 1993

Signature of Author _____
Sloan School of Management

Certified by _____
Charles H. Fine, Thesis Advisor
Associate Professor of Management

Certified by _____
Lawrence M. Wein, Thesis Reader
Associate Professor of Management

Accepted by _____
Jeffrey A. Barks, Associate Dean
Sloan Masters's and Bachelor's Programs

Learning During Ramping: Policy Choices for Semiconductor Firms

by

Richard H. Benfer

Submitted to the Sloan School of Management in partial fulfillment of the requirements for the degree of Master of Science in Management

Abstract

As the microprocessor industry becomes ever more competitive, firms will increase the frequency of introductions of new product and process generations. This strategy requires rapidly ramping production to high volumes. A key determinant of the success of this rapid ramping approach is the ability of manufacturing organizations to improve device yields throughout early production.

Ramping semiconductor manufacturing plants are a challenging place for ongoing learning. Large variations of within-lot wafer yields makes traditional split-lot experiments measuring final device yield less useful for gaining additional knowledge about the process. A number of approaches can be used, however, to generate new learning about the process and to improve yields in this environment.

An "incubation period" for sustained yield improvement was observed for several fab sites. This behavior stems from system wide influences during the early ramp. When resources are directed routine production matters, progress in yield improvement will be delayed. Policies that minimize the diversions stemming from excursions and capacity installations enable the manufacturing organization to devote more attention and resources to long term improvements.

Exposure to excursions is influenced by the state that the process is transferred to manufacturing. Processes that are optimized for low volume development environments will require many changes during production ramping in manufacturing plants. Increasing the fidelity of development facilities with manufacturing and surfacing potential production problems before startup will reduce the need for extensive fire fighting during production ramping.

Thesis Advisor:

Charles H. Fine
Associate Professor
MIT Sloan School of Management

The author gratefully acknowledges the support and resources made available to him through the Leaders for Manufacturing Programs, a partnership between MIT and major U.S. manufacturing companies.

Acknowledgments

I greatly appreciate the time and energy of everyone who helped with the research and preparation of this work. My thanks go to the many Intel employees who gave me their generous support and cooperation during my research. Their insights assisted me in understanding the technical, business and management issues of semiconductor manufacturing.

I'd especially like to thank my company supervisor, Cheri Solomon, for her assitance. I appreciate the guidance and timely feedback that she provided throughout the course of my internship.

I would like to thank Professors Charles Fine and Lawrence Wein for their advice and counsel.

Finally, I would like to thank my wife Meredith for making New Mexico a well-remembered adventure and for her understanding during the long hours that went into this effort.

Biographical Note on the Author

Rich Benfer, a native of Warrenton, Virginia, graduated from MIT in 1984 with an S.B. in Materials Science and Engineering. He served five years as an officer in the US Army Signal Corps, where his primary duties centered on the development of process technology for ceramic materials. He has also worked on semiconductor process development for IBM and MIT Lincoln Laboratories. He is working towards a master's degree in managment and expects to graduate in May 1993.

Table of Contents

List of Figures and Tables

Section 1 • Introduction

Intel Corporation is the world's leading microprocessor manufacturing firm. Much of its recent success is based upon its dominance in providing the microprocessors that run IBM compatible personal computers. IBM selected the Intel x86 architecture as the microprocessor that would be used in its initial foray into the personal computer market. Successive generations of this architecture (286, 386, 486 and Pentium™) have each spurred the rapid deployment of new product lines by the PC manufacturers.

The impressive revenues generated by Intel in the microprocessor market have attracted a number of "clone" manufacturers, most notably AMD (Advanced Micro Devices). These rivals have captured a large share of the 386 market since the introduction of their products. This source of competition has compressed the amount of time during a product's life cycle that Intel can reap high margins. Now Microsoft's new Windows NT operating system may change the market dynamics again. Windows NT will run on RISC chips from MIPS (R4000) and Digital Equipment Corporation (Alpha), as well as potentially other microprocessors from Sun and the IBM-Motorola joint venture. Programs currently designed for Windows could then run on these chips, erasing the software advantage Intel has had over RISC competitors.

The ability of Intel, as well as the other semiconductor manufacturers, to ramp new products and processes has become critical in today's business environment. In fact, competition in many industries has moved beyond quality and performance to reaching the target market in the required time window. Intel Corporation has now committed itself to more rapid introductions of new generations of microprocessors. In order to achieve this goal, design teams for different microprocessor generations (e.g., 686,

786) are now working in parallel on their circuit designs. (see Figure 1.1) This approach, as well as the use of sophisticated software tools, will enable more rapid releases of product designs. The important challenge facing Intel and other semiconductor firms is whether its manufacturing facilities can match the accelerating pace of the designers when they ramp production of new microprocessors.

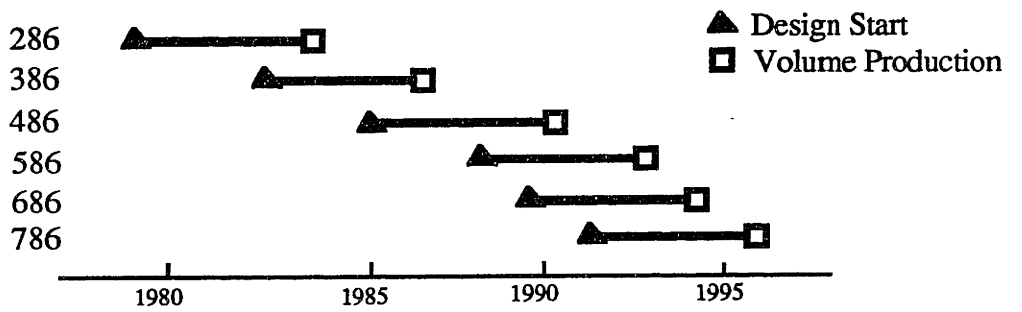


Figure 1.1 Intel Product Development Activity (Hof, 1992)

In figure 1.2, the curve marked "A" shows a schematic product life cycle for a microprocessor product. Much of the revenues generated by product sales occur while production is ramping, since product margins are highest in that time period. It is during this period that the investments in development and physical capital are recouped. Therefore, the success during ramping of a new product/process is critical to the firm's profitability.

A period of relative stable sales exists during the "product maturity" phase. The length of this phase depends on a number of factors, including the introduction of substitute products, price competition, etc. Eventually, sales fall off and the product is discontinued.

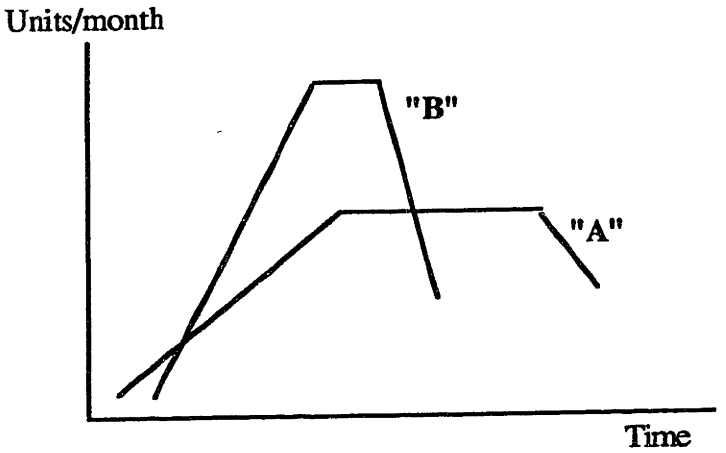


Figure 1.2 Schematic View of Changing Product Life Cycles. These changes present increased challenges to successful production ramping.

The combined effect of Intel's rapid deployment strategy and competitor entrance changes the shape of the product life cycle in three ways. First, products must be ramped much more rapidly to ultimate production. Second, the "plateau" at full production will be shorter due to competition and the introduction of the next generation product. Third, the backside of the life cycle will fall off more dramatically, as "clone" type manufacturers compete as the product migrates to a commodity status. These effects produce the enormous challenges to manufacturing. Curve "B" illustrates schematically the greater intensity of the ramp.

Project Purpose

This project grew from an interest in understanding what are some of the key limiters of output during the early ramp. Output during the ramp can be shown to be a function of installed equipment base, utilization of that equipment and yields (Figure 1.3). Yields

include both line yield (the % of wafers that make it through the entire process) and die yield (the % of good die).

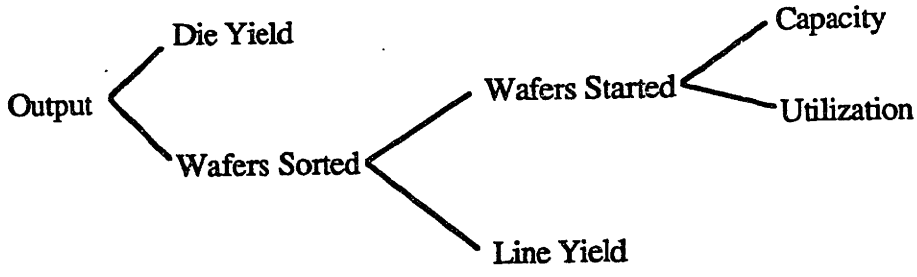


Figure 1.3 Influence Diagram for Manufacturing Output

Output is ramped through the simultaneous installation of capacity and the improvement of yields. Enormous leverage can be gained from die yield improvements in the early phase of ramping, since the yields at the start of the ramp can initially be on the order of a few percent. However, die yield improvement is viewed as more difficult than other actions for increasing output.

The main goal of this project was to understand the observed behavior of process die yield during the ramping of semiconductor fabrication facilities. Of particular interest in this analysis was what the key system drivers to yield improvements were. I looked at how a systems view of manufacturing ramping can help illuminate this behavior.

Major Observations

During the course of this project, I found that die yield improvements for a new process demonstrated an incubation period (as measured by time or cumulative production volume) before sustained progress was observed. Figure 1.4 schematically illustrates

this observation. This behavior was evident in a number of fabrication sites, suggesting that local conditions were not a significant driver to the incubation period. Rather, I saw that the general policies used to govern the ramp had a more systemic influence on performance. Management focus on routine production problems during the early manufacturing ramp delays sustained improvement of die yields.

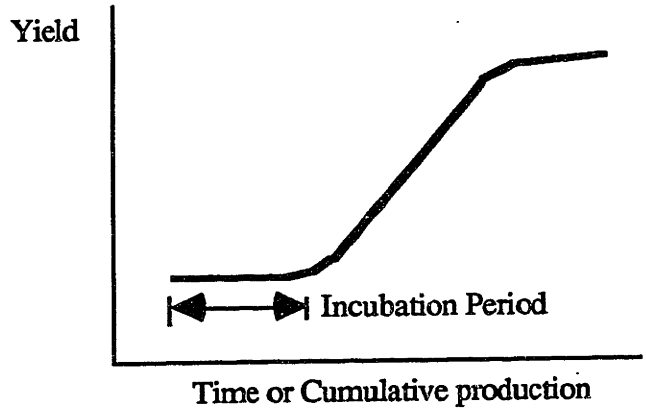


Figure 1.4 Schematic View of Incubation Period

A significant amount of time (4-8 months), between the identification of a yield improvement possibility and the eventual implementation of a verified improvement, is required for yield improvement activities. In addition, many *known* improvement opportunities are not addressed since resources and attention are directed elsewhere. The implication of this finding is that unless yield efforts are identified and focused upon before the production ramp, die yields during the critical early phase of the ramp are likely to remain flat.

The state of the process upon transfer from development determines the level of potential yield excursions. Processes that are optimized for low volume environments

with poor fidelity to eventual manufacturing plants will require many changes during the ramp. These factors serve to increase the delay in sustained yield improvement.

High noise levels in die yields *within* individual lots of wafers make full-loop split lot experiments problematic for identifying and managing improvements. A large amount of potential learning is lost due to this noise. A number of measures, however, can be used to increase learning in this environment.

In this thesis, I propose that a focus on base-line improvement efforts is key to shortening the observed delays in yield improvement. The level of resources diverted to other areas (fire fighting, installations) needs to minimized in order to achieve this goal. Actions within the manufacturing plant can be taken to reduce the impact of these activities.

One proactive measure for increased yield performance during ramping would be to restructure the role of the process development facilities. Surfacing potential production problems while the process is still in the development center is key to reducing excursions in manufacturing. Increasing the fidelity of the development facilities with respect to manufacturing helps achieve this aim.

Outline of Thesis

Section 2 reviews the research methods used in this project. Section 3 provides an analysis of the effect of noise on yield learning. Section 4 presents an overview of the yield improvement process. Section 5 discusses the importance of maintaining a focus on sustained improvements. Section 6 suggests some approaches that can be taken in the factory to improve focus. Section 7 considers the total manufacturing system and

reevaluates the role of technology development. Section 8 provides a summary of obserations and conclusions of this thesis.

Section 2 • Research Methods

This project is based upon a 5 month internship conducted at a semiconductor manufacturing site. This site consists of 3 distinct fabrication areas which produce a variety of semiconductor products. Insights drawn from this location provided the bulk of this analysis, though I was able to draw upon information and experiences from a number of other facilities.

Figure 2.1 outlines the general research approach used in this study. Data was collected from both clinical and statistical sources. This data was used to construct systems models of relevant areas. These models help in analyzing the tradeoffs among various policy decisions.

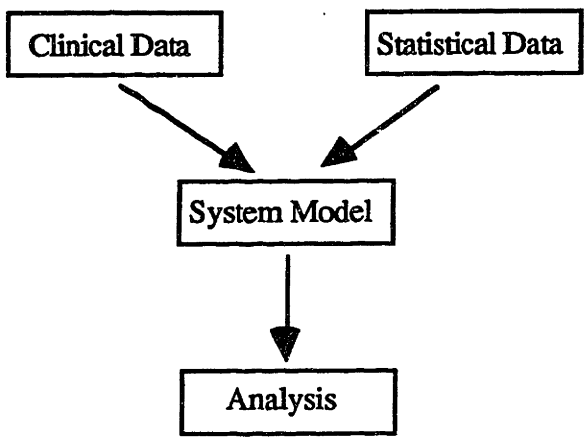


Figure 2.1 General Research Approach

Data Collection

Both statistical and clinical data were collected and used in this study. A brief outline of the ways in which this data was collected follows.

Clinical Data

The clinical data from this project was used to explain the patterns of yield behavior and to provide insights for building systems models. This information proved critical in developing a realistic view of the operating policies used in manufacturing. On-site discussion and observation proved helpful in assessing the impact of those policies.

A number of approaches were used: face-to-face interviews (both structured and unstructured), remote conversations (via telephone and email), attendance of meetings, written reviews and plant tours. The following lists some of the attributes of each format.

Interviews: Conducted on a one-on-one basis. Interviewed personnel from operations, finance, engineering and senior factory staff. Principals subjects addressed included the yield improvement process, the timing of capacity decisions, background on previous factory ramps and factors they felt influenced the incubation time.

Conversations: Often used email as a chance to ask follow-up questions. Telephone conversations were used to solicit answers from personnel from remote manufacturing and development sites.

Meetings: Attending the numerous meetings (from plant wide to specific functional area reviews) enabled me to understand the priorities assigned to particular actions. These also provided me with an opportunity to make contacts with key personnel and to arrange interviews with them.

Reviews: Written reviews included regularly scheduled weekly/monthly reports, as well as ad hoc post mortems on particular matters. These reports helped identify significant events and obstacles that occur during each factory ramping.

Plant tours: The main objective of the plant tours was to familiarize myself with the process technology used in microprocessor fabrication. This understanding

was necessary for the interviews described above and for interpreting the observed yield behavior.

Statistical Data

Data was collected from historical records as well as using current data generated during the internship period. Existing data from the plant's regular reporting systems was used in order to reduce the work required in generating the data. This data was archived by a variety of groups, including finance, operating and engineering. Restricting the project to existing data imposed certain constraints on the research; data on potentially useful dimensions was simply not available.

In addition, on-line databases concerning current products and processes provided detailed factory information. These databases are often used by factory staff to identify and solve yield issues. I found this data critical when examining the impact of noise on experimental results. Unfortunately, this on-line data was stored for only finite periods of time. To obtain information on earlier time periods it was necessary to search through written files or use the aggregated charts from the various functional groups.

Developing a Grounded Theory

In order to produce a useful and robust thesis, I have attempted to form a theory that creates a better understanding of the observed behavior. The theory must be clear and sufficiently general in order to be applicable to potential users. However, I agree with Strauss when he asserts that, "without grounding in data, this theory will be speculative, hence ineffective." (Strauss, 1987) I therefore attempted to apply some of the concepts of grounded theory research. In addition, I have also used some of the tools from systems thinking (most notably causal loop diagrams) to articulate my hypotheses. Burchill advocates this combination of various aspects of grounded theory and system dynamics, which he names "Inductive Systems Diagrams" (Burchill, 1992). This combined approach proved useful in developing and explaining the hypotheses in this work.

Grounded Theory

Three processes occur concurrently throughout a research project: induction, deduction and verification (Strauss, 1987). Induction concerns the processes that are used to develop hypotheses. These hypotheses are conditional and can be readily modified. Deduction refers to drawing conclusions from the hypothesis (if X, then Y) in order to verify the theory. Verification consists of exploring whether the data supports or negates the theory in whole or in part. Figure 2.2 illustrates this cyclical process. It should be noted that these cycles can be of the order of minutes (e.g., quick conversation) to years (e.g., long term projects). The objective behind this process is to develop a theory, grounded in the data, which creates a better understanding of the targeted behavior.

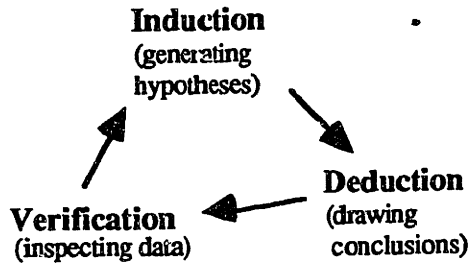


Figure 2.2 Grounded Theory Cycle

Developing a grounded theory generally requires the use of several work processes. Key to this approach is (1) data collection, (2) extensive data coding and comparison and (3) continual writing of analytical memos. These activities include:

1. Examining the data by early and continual coding of each incident and comparing incidents with other incidents and eventually with an accumulation of incidents.
2. Developing provisional linkages among discovered concepts. These linkages are verified and strengthened by each phase of coding and comparison.
3. Identifying the core categories and linkages. Linkages are integrated into more unified theory, often with the aid of theoretical memos. The theory becomes tighter as less salient features are dropped and detailed categories are abstracted into higher level concepts. (Strauss, 1987. Burchill, 1992)

Most critical to this process is the task of coding the data. It is helpful to use the following coding paradigm (Strauss, 1987):

- Focus questions and answers about the available data on aspects pertaining to (1) conditions, (2) interactions, (3) tactics and (4) consequences.
- Maintain a"brainstorming mindset" in order to enable the coding process to snowball and result in many unexpected concepts.
- Winnow down these concepts only in the later stages of the inquiry.

Systems Thinking

Causal loop diagrams can be used to describe complex systems (Kim 1992). Appendix B gives a review of the basic components and structure of causal loop diagrams. These diagrams capture much of the richness of a complex system, but in an intuitively easy to understand manner. A number of systems archetypes have been developed; these general structures aid in quickly developing models.

Burchill and Kim have developed a useful process to build causal loop models (Kim and Burchill, 1992). This process, modified for use by Burchill for inductive system approaches, is outlined below:

1. Selecting the Variable. Identify the core categories and their symptoms. This variable should have central importance on the events being studied.

2. Identify Related Variables. Identify the causes and consequences of the selected variable.

3. Diagram Interactions. Illustrate the interrelationships among the selected features using causal loop diagrams (see Appendix B).

4. Check Diagram for Consistency. Compare diagrams to insure they are grounded in available facts. Check that the diagram is easy to understand and has direct links between causes and effects.

5. Integrate Diagrams into a Systems Diagram. Combine individual causal loop diagrams into a *system diagram* that captures the integrate whole. Develop a central theme and prune less influential loops to simplify the diagram.

6. Clarify the Message. Identify any archetype(s) that is present in the systems diagram. Reformat the diagrams to make them more clear. Prepare a written summary that explains the insights generated by the diagramming process.

Case Study

A brief example of the use of Inductive Systems Diagrams in the development of a grounded theory for tactical factory decisions is shown below. Specifically, this work illustrates the structure and implications of a policy for responding to shortfalls in planned output (Benfer, 1992).

The key variable was selected: shortfall in output. An initial observations was made that start rates were increased when output fell below planned levels. Coding for variables from this observation included: start rate and output gap. Figure 2.3a captures this relationship:

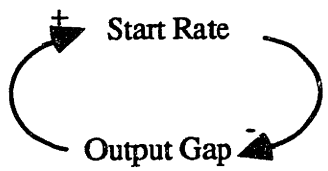


Figure 2.3a Initial Coding: Start Rate and Output Gap

The second observation noted that increasing start rates increased the overall time a wafer spent in the fabrication environment. This empirical observation of the system is in agreement with queuing theory, which, assuming a constant equipment set, shows that queues will lengthen as arrival rates increase. Both published and experiential data suggests that longer throughput times in the fab area result in lower yields (Anastasio, 1991; Wein, 1992). A third observation reported was that an increase in the start rate

can put significant pressure on both equipment and personnel in the factory. The increased pressure can lead to confusion (misprocessing) and delay of scheduled preventive maintenance, both of which have a negative influence on final yields from the process. From this set of observations, we can code for throughput time, production pressure and yields. Figure 2.3b maps the critical relationships:

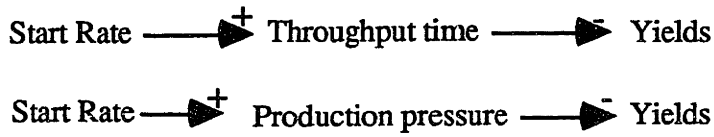


Figure 2.3b Second Coding: Start rate, TPT, production pressure and yields.

In the long run, it was observed that output goals could be addressed by raising yields. Raising yields requires *dedicated* and *sustained* yield improvement activity (see Section 3). However, it was noted that increasing production starts limits the capacity to process engineering wafers through the line. In addition, personnel (especially management) divert their attention from yield issues to the immediate focus of increasing output. These observations are coded as output gap, yield improvement activity, resources, and yields. Figure 2.3c shows the important relationships:

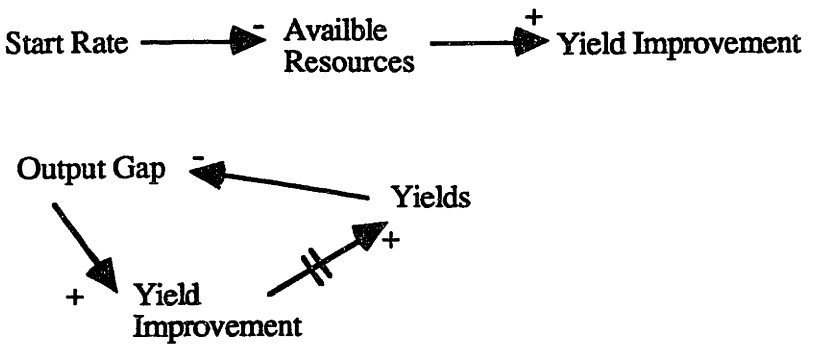


Figure 2.3c Third coding: start rate, resources, yield improvement activity, yields.

The diagrams shown above were grounded in the available data. It can now be integrated into an overall systems diagram. Figure 2.4 fully captures the structure of the policy. All significant variables have been diagrammed and a central theme has been developed:

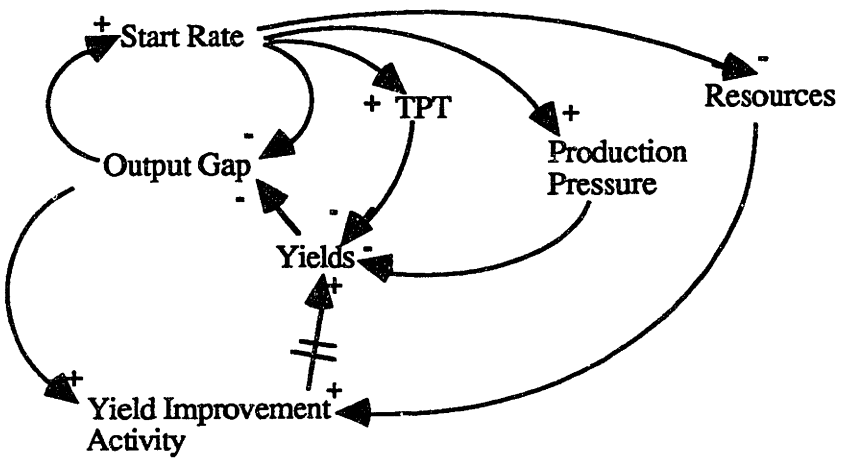


Figure 2.4 Final Causal loop diagram mapping output policy. (Benfer, 1992)

The final diagram is recognized as a familiar system archetype: "shifting the burden". This archetype illustrates how a short term "solution" is habitually used to address the problem symptom, while the longer term solution is left unused or even weakened. This can lead to even greater reliance on the short term fix.

In this case, the short term response is to increase the start rate. A more fundamental solution of yield improvement, which requires longer lead times to generate higher output, is used less. This behavior can "trap" an organization into short-term reactive practices.

Section 3: Noise and Learning by Experiment

Overview

This following analysis uses Bohn's methodology for analyzing noise levels in semiconductor plants (Bohn, 1991). Process improvement experiments are particularly susceptible to process noise. The level of noise, therefore, is a critical factor influencing the rate of learning from full-loop experiments.

The purpose of this section is to analyze the effect of noise on learning by experiments. The key question concerns the level of noise: Is the magnitude of the noise significant enough to seriously degrade the information produced by experiments? And if so, to what extent is learning lost?

The goal of process development is to alter the process (by changing an operating parameter, equipment set, maintenance schedule, etc.) so that die yield Y improves. Experiments are used to guide this process development work. The engineer makes a process change (ΔX) in order to see its effect on yield. The result of an experiment can be stated as:

$$Y_{new} = Y_{old} + \Delta Y + \varepsilon$$

where: Y_{old} is the original die yield

ΔY is the change in average die yield per wafer as a result of the process change

ε is the experimental noise

The experimenter can use the observed results to decide whether to adopt the change. The decision rule can be a simple one: if Ynew > Yold, then implement the change. Or it can be made considerably more complex by taking into account information from other experiments, risk, costs, experiential factors and signal to noise ratio.

The signal-to-noise ratio of the experiment can be stated as the ratio of $\Delta Y/\mathcal{E}$. Low signal to noise ratios can render experimental results valueless. Process variation is a fundamental contributor to experimental noise. Other sources of experimental noise include misprocessing of experiments, measurement errors and design errors. Process variation induces process noise (unexplained variation in yields). Process noise has two components: between batch noise and within batch noise (Figure 3.1).

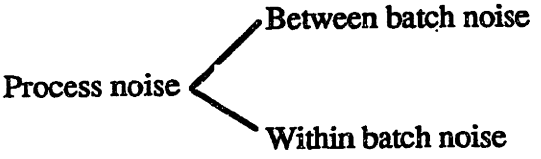


Figure 3.1 Components of Process Noise

The yield of a particular wafer in an experimental batch can be described as:

$$Y_{\text{wafer j in batch k}} = Y_{Ave} + \mathcal{E}_{\text{batch k}} + \mathcal{E}_{j/k}$$

where: $\mathcal{E}_{\text{batch k}}$ = effect of between batch noise and

$\mathcal{E}_{j/k}$ = effect of within batch noise.

This section shows that both between batch and within batch noise levels can be very high. Between batch noise is particularly large, so experiments are usually run on split batchs. That is, a particular batch is divided into two parts: active and control splits.

Both splits are processed together until they reach the process change. The active split is then processed using the change, while the control split is processed according to the normal recipe. The splits are then recombined and processed together through the rest of production. In this way, between batch experimental noise is effectively blocked.

The Level of Process Noise

The fabrication facility which yielded the data for this analysis was a high volume factory (factory "A") producing a variety of integrated circuit products on a single process. Data from one product over a three year period was studied. These data groups are labeled A1, A2 and A3; which correspond to approximately 1, 2 and 3 years of process maturity, respectively. Wafer level data indicating the number of good die per wafer in each batch was provided. All absolute yield data has been disguised.

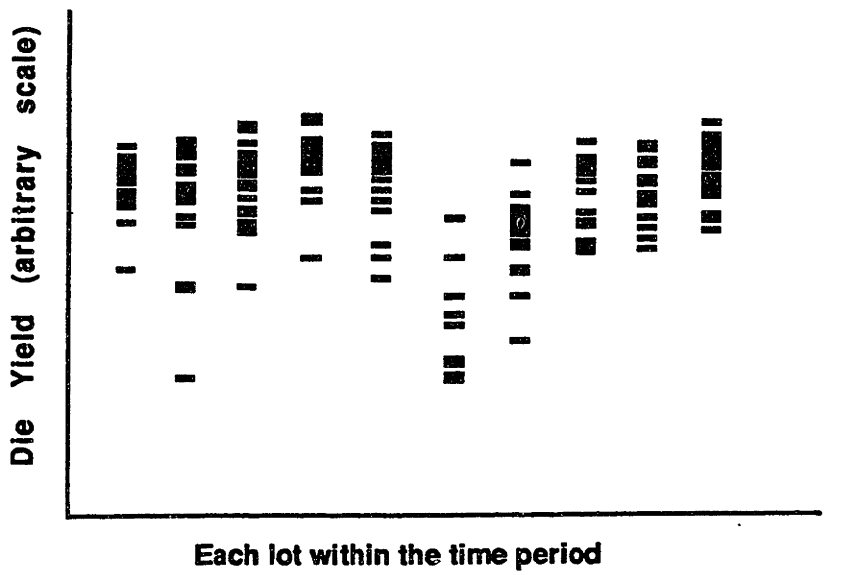


Figure 3.2 Die yields from Factory A1.

Figure 3.2 shows data from the factory approximately one year after the beginning of production (A1). The yields have been rescaled in order to protect confidentiality. The data is for ten batchs of the same product that were sorted in the same week. Each column represents one batch, with each data point representing a single wafer. Within batch variability of the die yield is the spread in each column. Between batch variability is the difference among columns. Especially evident are the differing shapes of the graphed data for each batch. These differences indicate that the batches were not produced under identical conditions, an observation which suggests a manufacturing process not under robust process control.

Die yields can be assumed to arise from a multiplicative process, such that

$$\text{final die yield} = Y_1 \cdot Y_2 \cdot Y_3 \ldots \cdot Y_{n-1} \cdot Y_n$$

where Y_n is the die yield for a particular process step n
n is the total number of process steps.

In this way the yield of each process step has the effect of multiplying the die yield of the rest of the process by some constant less than one. Correspondingly, a positive process change should have a multiplier effect greater than one. Therefore, the natural logarithm of die yield, *ln*(die yield), can be used as a measure of experimental results. For example, a change of +0.03 in *ln* (die yield) is equivalent to a 3 percent improvement in average die yield. The standard deviation of *ln*(die yield) is then a measure of the percent variability of the yield and is approximately equal to the coefficient of variation of absolute yields (Bohn, 1991).

Table 3.1 summarizes the data from A1. An arbitrary constant has been added to the batch mean in order to disguise the true die per wafer yield. There is a considerable range in both the mean and standard deviation of yields among the batches.

Table 3.1 Summary Statistics of Factory Data A1

Batch	ave *ln* Y	sdev *ln* Y
a	5.15	0.18
b	4.97	0.35
c	5.28	0.10
d	5.05	0.51
e	4.95	0.21
f	5.12	0.32
g	5.26	0.14
h	5.28	0.12
i	5.28	0.18
j	5.19	0.26
min	4.95	0.10
max	5.28	0.51
ave	5.15	0.24
stdev	0.13	0.13

* An arbitrary constant has been added to the mean in order to disguise true yield data

Figure 3.3 illustrates the within batch noise levels over three years (A1, A2 and A3) in Factory A. The standard deviation of *ln*(die yield) for each batch is recorded as a single point. A number of observations can be made. First, there are high levels of variability within each batch. Second, this variability is not consistent among batches within each time period. Third, the average variability decreases with factory life.

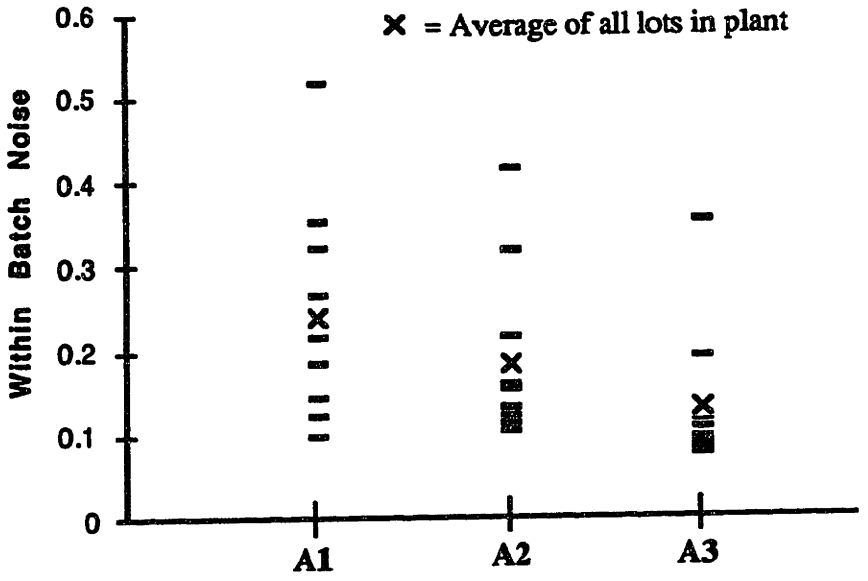


Figure 3.3 Noise levels by Batch for Three Time Periods

Effect of Noise on Experimental Confidence Levels

A test statistic for a split lot experiment is shown below. The statistic is the difference between the average yield in the active and control groups.

$$\Delta Y_{exp} = \{(1/N_a)\Sigma\, Y_i\} - \{(1/N_c)\Sigma\, Y_j\}$$

where Y_i is the yield of the i'th wafer in the active group

Y_j is the yield of the j'th wafer in the control group

N_a and N_c are the number of wafers in the active and control groups, respectively.

ΔY_{exp} of the experiment relates the observed difference in mean yields between the active and control groups. The larger the difference, the greater the likelihood that the process change will have a positive effect. The level of certainty that ΔY_{exp} accurately

reflects the true effect on yield (ΔY_{true}) depends upon a number of factors, including the within batch yield variation and the number of wafers per batch. Using the test for the difference between two means, we can determine the probability that a particular change does have a positive effect.

The following analysis uses a two-sample test (with σ_a and σ_c known). We assume a normal distribution of yields, resulting in a z variable in the following form:

z = (point estimator - hypothesized mean)/ standard error of point estimator.

The point estimator is the difference in the observed means. We will test H_o: $u_a - u_c \geq 0$ against H_a: $u_a - u_c = 0$. The standard error, σ_{xa-xc}, is in the form of

$$\sigma_{xa-xc} = (\sigma_a^2/n_a + \sigma_c^2 n_c)^{1/2}$$

We assume $\sigma_a = \sigma_c = \sigma$ and $n_a = n_c = n$ for a split lot experiment. With the null hypothesis of $\mu_a = \mu_c$, the z-test statistic can be written as:

$$z = (x_a - x_c) / (2\sigma^2/n)^{1/2} = \Delta Y_{exp}/(2\sigma^2/n)^{1/2}$$

The *average* standard deviation in *ln*(yield) for all of the batches within a time period was used as a measure of σ. Table 3.2 states the minimum value of (xa -xc) required to state with 80, 90, 95,97.5 and 99 percent probability that the process change would in fact have a positive influence ($\mu a > \mu c$).

Table 3.2 Minimum Required Die Yield Changes

α =	80%	90%	95%	97.5%	99%
A1	0.062	0.093	0.117	0.138	0.162
A2	0.051	0.077	0.097	0.114	0.135
A3	0.038	0.057	0.072	0.085	0.101

Values are given in Δ (natural log of observed effect).

It is observed that even at an 80% confidence interval, rather large improvements (on the order of 4-6 % are required. In order to reduce the chances of Type II error, engineers may require a high cutoff for accepting the results of experimentation. This cutoff criteria, however, increases the chances of rejecting a positive true improvement (type I error).

Without tight controls, adopting process changes based upon the results of a few split lot experiments can be a risky undertaking. Conversations with engineers suggest that such an approach has been counterproductive. In one review of process changes made in one fab, it was found that only a fraction of the changes produced positive effects after implementation; indeed, an equally large fraction led to *negative* results.

Effect of Noise on Learning by Experimentation

ΔY_{exp} deviates from the true effect due to experimental noise. The relation between the two is shown as

$$\Delta Y_{exp} = \Delta Y_{true} + \varepsilon_{experiment}$$

where ε is the experimental noise.

A bootstrapping technique (see Appendix C) was used to generate power functions G(ΔY) for the plant during time period A1, A2 and A3. A simple test criterion was then assumed: if ΔYexp > 0, then accept the process change; otherwise, reject it. Each power function shows the probability of accepting a process change if the true value of the change ΔY.

The power functions were used to generate Table 3.3, which summarizes the consequences of within batch noise for the plant at the different time periods. The table lists the probability of missing process improvements in the range of $0.01 \leq ln(\Delta Y) \leq 0.20$ with n = 12. The effect of noise in the factory is dramatic. Only very large yield improvements will be identified. For $ln(\Delta Y) = 0.01$, the probability of missing an 1% yield improvement ranges from 44.3% to 40.8%, indicating that nearly half are rejected.

Table 3.3 Power functions for factory A

	A1		A2		A3
ΔY			Prob of miss		
0.01	44.30%		44.10%		40.80%
0.03	34.10%		30.90%		24.30%
0.1	13.00%		8.10%		3.60%
0.15	6.20%		3.40%		1.50%
0.2	2.60%		1.50%		0.90%
Prob miss			ΔY		
10%	0.1162		0.0886		0.0605
20%	0.0690		0.0526		0.0378

Figure 3.4 illustrates the power function for the factory for the time periods A1, A2 and A3. Each curve shows the probability of accepting the new production method, as a function of ΔYtrue (the true effect on yield), which is unknown to the engineer. The height of the power function curve shows the probability that the engineer will accept the hypothesis that ΔY>0 (i.e., the new method is better).

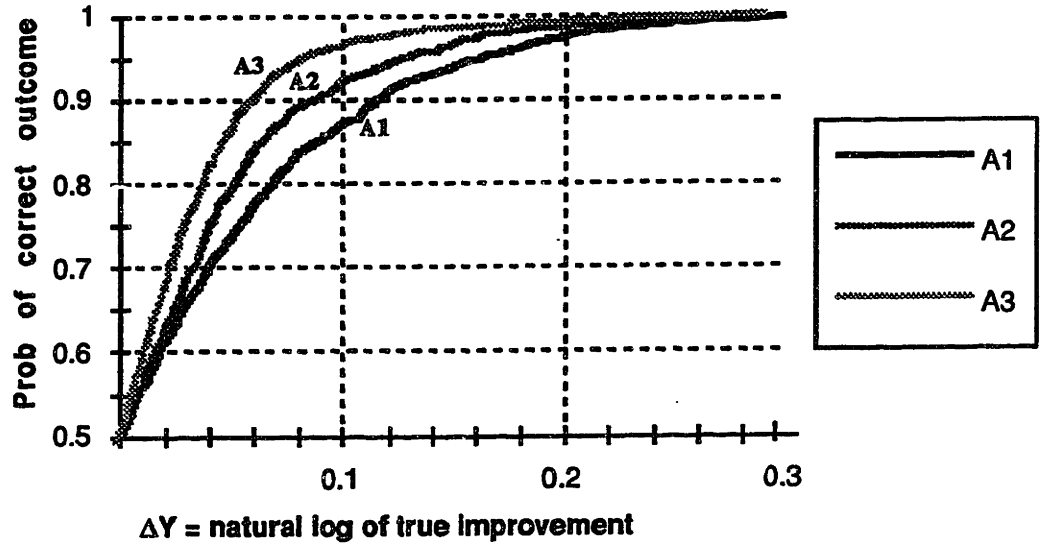


Figure 3.4 Power functions graphed

Figure 3.4 shows that the power function changed over time. Plant A improved its ability to recognize yield improvement over the three year period. This improvement was realized through decreased variation in within batch yields.

Decision rule ΔYest>ΔYcutoff

Because of the symmetry in this model, there are equivalent probabilities for accepting true negative changes (Type 1 error). For example, a change with a ΔY of -0.03 during period A1 has a 34.1% chance of being mistaken for a positive change. The analysis in the previous sub-section showed various confidence levels for factory time periods A1,

A2 and A3. If an engineer applied a decision rule of $\Delta Yest > \Delta Ycutoff$, where $\Delta Ycutoff$ was determined by a 80% confidence level, then the power function curves would shift to the right. Figure 3.5 shows the consequence of this shift for each time period. It illustrates the minimum yield increase that the factory can detect in each time period with less than a 10% chance of rejecting the improvement.

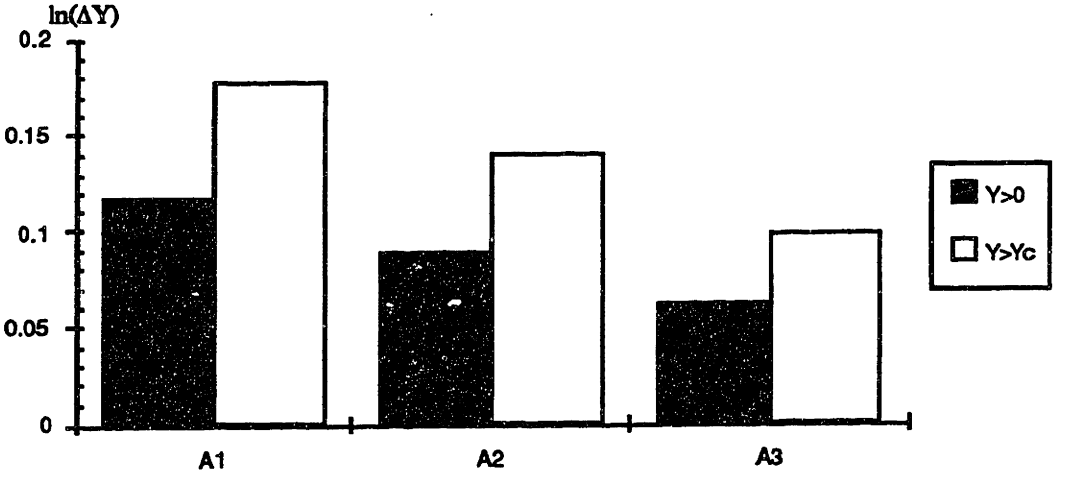


Figure 3.5 Smallest detectable change with 90% accuracy.

We see that the effect of the shift in the power function curves is to raise the threshold for the minimum detectable change.

A summary of observations concerning noise levels in semiconductor manufacturing includes:

- Noise levels are high but reduce over time
- A large amount of learning is overlooked because of noise
- Significant experimental improvements are necessary to insure reasonable confidence levels.

Noise levels seriously degrade the information obtainable from full loop experiments. Therefore, full loop experiments can not be the primary tool for yield learning. Their utility is in a large degree restricted to verification that proposed changes (based upon the results of other types of experimentation and analysis) do not crash yields.

Tactics for Learning in Noisy Environments

Policies that emphasize systematic reductions in process variation are critical for reduced experimental noise levels. The thousands of process variables in a modern semiconductor process interact non-linearly and in ways that are not easily recognized. Reducing the impact of these non-linearities by limiting variation at all steps is a broad brush method for both improving the signal to noise ratio and average yields.

Improvements can be accomplished by reducing the sensitivity of both the process and the product. Robust process design strategies can limit the effects of disturbances on the process. Design for manufacturing (DFM) approaches can aid in improving the process window for specific products. This type of analysis would examine design tradeoffs for feature size and layout geometries. The later method requires a strong feedback loop from manufacturing yield engineers to product designers. It is usually acknowledged that this interaction is not particularly strong.

A variety of experimental methods can be used to reduce the effects of noise on learning. Most have serious drawbacks, though, especially when one considers information cycle time and the consumption of production resources. Multi-batch and sequential experiments are examples of procedures where these drawbacks are pronounced. Increasing the sample size of split lot experiments would increase signal to

noise ratios, but is not generally feasible since most factories move batches in 25-wafer lots.

Several leading Japanese manufacturers have improved process control using computer integrated manufacturing (CIM) systems (Hodges, 1990). CIM systems that capture large amounts of operational and yield data enable the use of "natural experiments" to enhance yield learning. They improve signal to noise ratio largely through increases in effective sample sizes. Correlations between operating variables, physical measurements (defect density, critical dimensions) and performance data can be made through the regression approaches. In addition, common factors (equipment, personnel, run time, throughput time) for low yielding lots can be identified and examined. This approach is particularly helpful during early ramps when line yields suffer due to misprocessing. The main drawback of this approach are the capital and time requirements necessary for (1) developing the software and hardware and (2) training the employees to use it.

Perhaps the most often used approach are short-loop experiments. These experiments are conducted through a limited number of process steps. Instead of directly measuring die yield, the engineers evaluate other parameters which have a high correlation with final yield and device performance. Such metrics include particle generation, uniformity, critical planar dimensions, thickness, etc.

The benefits of this approach are fast information cycle time (often on the order of minutes to hours), lower experimental noise and reduced costs (those associated with tying up production equipment). However, in order to achieve these benefits and increase learning, the process must be modular enough so that yield models can associate the defect structures with final yield and performance.

Unfortunately, semiconductor fabrication is a highly integrated process. The interdependency between process step 45 and step 125 may be significant but go unrecognized. Therefore, a process change that appears to improve parameters associated with step 45 may have unknown and drastic negative consequences at step 125. In other words, the state of knowledge about the process may not be high enough to support making changes based upon short-loop data alone. Instead, these potential changes are required to be first verified through full-loop runs.

The high level of process variation and its associated impact on experimental noise affects the choice of approaches used by the engineering staff to improve yields. The next section reviews a general roadmap for the yield improvement process.

Section 4 • Yield Improvement Process

The yield improvement process relies on many resources. These include the different functional groups (yield engineering, process engineering, industrial engineering, test engineering, operations), automation systems, materials analysis, dedicated time on process equipment, and software tools. This section outlines a general model of yield improvement activity. This model illustrates the various activities that occur and the relative time each activity requires.

Improvement Activity Flowchart

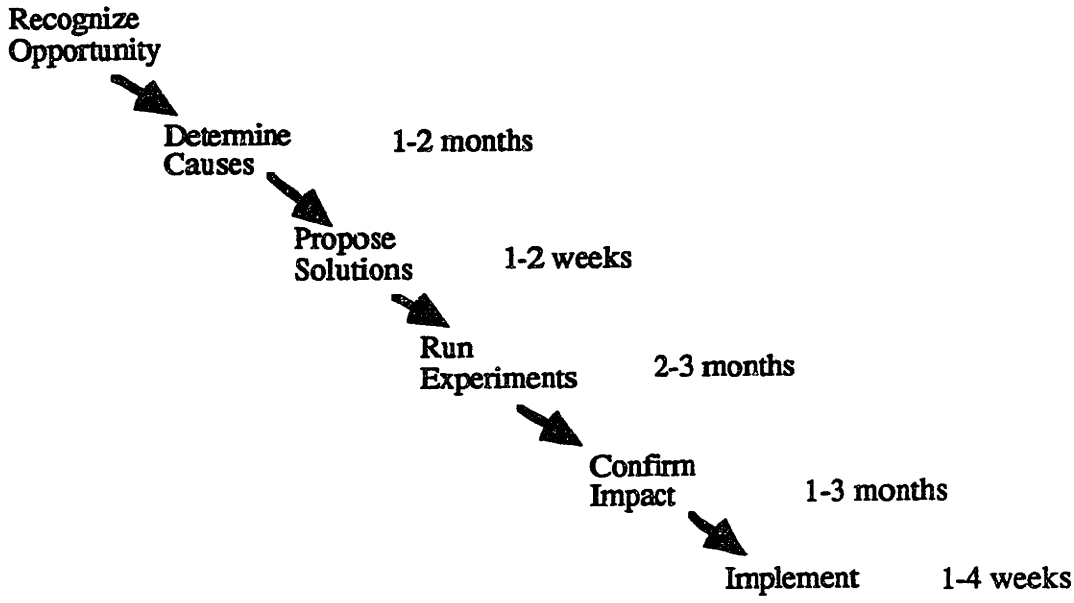


Figure 4.1 Yield Improvement Activity

Figure 3.1 diagrams a generalized flowchart for yield improvement activity. The approximate times required for each step are indicated beside it. What is first evident is the total amount of time that is required to complete a project once it is started: 4-8 months in most cases. This results in a significant delay for yield learning.

Engineers manage this sequence of events during a specific process improvement project. The activities can be viewed as part of a larger PDCA cycle for continuous improvement:

Plan	Recognize Opportunity Determine Causes Propose Solutions
Do	Run Experiments
Check	Confirm Impact
Act	Implement Changes

These activities are described in more detail below.

Recognizing Opportunities

The confusion levels in a manufacturing environment often make recognizing a yield opportunity (or problem, depending upon your mindset) a difficult activity. Some of this confusion stems from the high level of noise that exists in yield data (see Section Three). One sure way in which a problem can be highlighted is when a yield excursion occurs. This event has the immediate effect of focusing attention on a problem area. Unfortunately, responding to excursions is rather reactive by definition. In addition, the solution is usually that of a quick fix (e.g., replace defective part, re calibrate equipment) rather than baseline process improvement.

The use of historic data (particularly CIM data) has been advanced by Bohn (Bohn, 1988). Careful examination of yield data (especially on low yielding lots) can ID opportunities for improvement in a more proactive manner. Yield engineers can look for high leverage opportunities by examining pareto charts of losses. This technique focuses further work on the largest contributors to yield loss.

Other opportunities are developed via the experience and knowledge of engineers and operators. Engineers develop experience on both the current and previous processes; this experience provides valuable lessons for identifying opportunities. Knowledge gained from external agents (literature, short courses, colleagues and vendors) can suggest new improvement approaches. Examples of these types of opportunities are changes in die layout, adapting new components to existing processing equipment, changes in maintenance performance, etc.

Before a problem/opportunity can be addressed, a decision to allocate finite resources (personnel, time, materials, equipment usage) must be made. Opportunities can be ranked in order of their "expected" effect on output. For yield improvements, this translates into the largest expected increases in yield. The relative risk of the project, and the associated costs (personnel, materials, disruption), are also considered. I found that a number of successful projects had been identified well in advance of actually being launched; for various reasons (some of which are described above) these projects were delayed until a later date.

Causal Analysis

After a problem has been identified and resources allocated to a project, an effort is made to determine the contributing factors to the problem. A number of methods, with varying "cycle times', are used. An automated database containing processing records

of individual batches can be used to search for common attributes. That is, if a several lots show the same type of problem, their processing histories can be cross-checked, and common occurrences, equipment sets, operating personnel can be identified. This helps in narrowing the range of potential causes.

Performance test data is also used to target causes. For instance, certain electrical failures are associated (through experience) with particular problem sources. Information from these tests can be used to spot the occurrence of these types of problems. Another type of electrical test aids in identifying the location of an affected area (defect). Certain products have regions that can be scanned electronically. Signals that are abnormal can be pinpointed to certain areas (and sometimes levels) of the device. These areas are then targeted for further evaluation.

A more time consuming method is to exploit materials analysis tools. An SEM can be used to visually inspect the device for any defective area. X-ray spectroscopy will give an elemental breakdown of suspected defects. Various etch chemicals can then be used to strip one layer for another round of analysis. This process of removal and inspection can be repeated until bare silicon is reached. As it can be imagined, this approach is very time intensive.

Equipment related causes can often be identified through the use of monitor wafers. These wafers can be evaluated for the uniformity and accuracy of the process step. The monitor wafers can also relate the number of defects (particles) generated by the equipment. This information is helpful in ascribing problems to equipment operation. However, since monitor wafers are not equivalent to production wafers, care must be taken not to assume that the operating conditions that exist during normal production exist during the processing of monitor wafers. For instance, in-situ measurements

suggest that particle levels are higher when product wafers are run that when monitor wafers are processed (Bordon, 1989).

Developing Solutions

Engineers usually have a number of solutions in mind after the correct cause and effect relationships have been attributed. Often, these solutions stem from experience gained in solving identical or analogous problems. Others sources of ideas include colleagues, technical conferences and vendors.

The solution approach can either attack root causes or treat the problem symptoms. For instance, take the example of a process step that generates large numbers of particles. One approach is to add an inspection/removal procedure after the process step. A more fundamental approach is to reduce or eliminate the generation of particles. The tradeoffs between these approaches include experimental costs, project risks and the ultimate effect on production operations.

Experiments

The previous section detailed the impact of noise on full loop experiments. The impact of noise and the high costs of full loop experiments lead engineers to rely on other methods to specify and test proposed solutions. Chief among these are the use of short-loop experiments.

Short loop experiments are taken through only a few steps of the process. This reduces both the noise and the costs associated with this method. Moreover, the information turnaround time is tremendously improved. Short loop experiments usually focus on a

key attribute linked to final yields: particle density, critical dimension, uniformity, thickness, etc.

Good models for correlating these attributes to final yield aid in prioritizing and addressing yield improvement opportunities. The low state of knowledge about many parts of the process (especially during a ramp of a new process) hinders the development of models in two significant ways. First, the key parameters for a particular process are not well known. Second, the interdependencies of one part of the process on another have not been established. These factors make it more likely that short loop experiments will emphasize local optimums at the expense of more systemic improvements.

The effectiveness of short loop experiments is bounded by the experimental methods used by the engineers. The greater use of Design of Experiments (DOE) and statistical tools has added greater rigor to the improvement process. These approaches enable more and better quality information to be extracted from experimental results. Of equal importance is the correct following of experimental procedures. By their nature, short-loop experiments deviate from the normal production process. This increases the likelihood of misprocessing. In fact, Bohn suggests that the misprocessing of experimental batches can be as high as one third (Bohn 1991).

Confirming the Impact

The fabrication process can looked at as a web of interrelationships. As stated earlier in this section, these relationships are often not documented or even well understood. Because of this situation, process changes are not made on the basis of positive results

from short loop experiments alone. They must also be confirmed by full loop experiments.

After the initial rounds of short loop experiments identify a potentially favorable process change, full loop experiments are performed in order to gauge the effect on final yields. Their use is not to quantify the exact yield improvement, since the high noise levels associated with full loop experiments would necessitate a prohibitively costly number of experiments. Rather, it is to verify that the change does not cause yields to crash. Deciding the exact number of full loop experiments requires a tradeoff between the costs of such experiments and the risks of not identifying the exact effect on yields.

Full loop experiments require that lot be processed through the entire production line to electrical testing. Therefore, throughput time will restrict how quickly the lots are processed. Actual throughput times for experiments can differ from the normal production throughput time. If lots are put on engineering hold for extended periods for additional inspection, then the throughput time can be significantly longer. This increases the information cycle time for experimental learning.

This cycle time can be reduced in two ways. First, efforts can be made to minimize the time that engineering lots are kept on hold. Second, particular lots can be expedited. This approach, however, disrupts the line and can unduly influence results via the careful handling and attention these lots receive as they are shepherded through the line.

Implement the Improvement

A policy of tweaking the process after experiments verify improvements can still be very counterproductive. This is particularly true if there is more than one factory running a single process. The processes at each factory would diverge over time and any synergy that might have been gained by joint effort will be diminished. Therefore, the change process needs to be tightly controlled. One way to accomplish this is to institute a multifunctional (and multisite) group to filter and authorize suggested changes. In this way the change process can be controlled and the specifics disseminated to other organizations within the firm.

The decision over whether to first pilot the change or to fully implement it is based upon a judgment of its risk. Changes that do not effect critical device dimensions (and thus less likely to degrade performance) can usually be implemented without a full scale pilot. More significant changes can first be run on a partial equipment set and then gradually implemented on all equipment.

Caveat

This treatment of yield improvement suggests a linear sequence with respect to time. In fact, there can be a great deal of iteration within each step (causal analysis, experimentation, confirmation) and between these steps. This iteration can lead to even longer times required for a yield improvement effort. To the extent that this iteration is not planned, a wide variance in project completion dates can be expected.

Improvement Cycle Time and Rate of Learning

The finite time it takes for process improvement projects to reach completion limits the rate at which learning about the process can occur. The rate of learning, in turn, gates the effectiveness of the yield improvement process. These relationships can be described as follows:

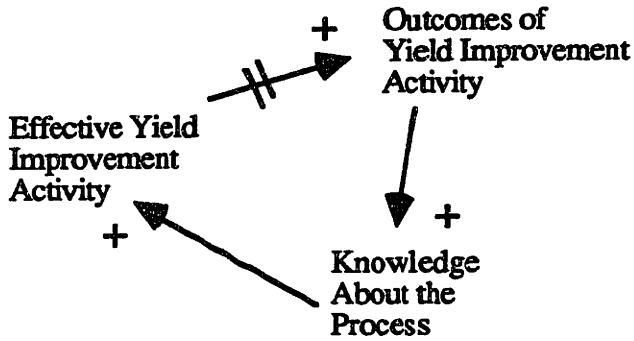


Figure 4.2 The Yield Learning Cycle

Recognizing and reducing the delay between identifying improvement opportunities and implementation are key to developing more effective yield learning. Understanding that there is a definite time lag for this process should induce managers to focus on base line yield improvement earlier in the ramp. If this activity is put off until after the initial phase of the ramp, there will be further delays in actual progress (results) in yield improvement.

Improving the Process

There are a number of approaches to improving the effectiveness of yield improvement work. Two broad categories include (1) specific tools and (2) management practices.

Use of Tools

Yield models

Yield models relate final yields with defect density (see Appendix D). Dance outlines the benefits of using yield models in conjunction with short-loop monitors to improve yield improvement efforts (Dance, 1992). Fine-grained yield models use more than a single defect density estimate in their calculation; they use the contribution of specific process zones or defects to estimate defect limited yield. This allows engineers to understand their specific area's effect on final die yield. Coupled with short loop experiments, yield models accelerate learning.

In-situ monitors

In-situ measurement of operating conditions in processing equipment provides benefits over using short loop monitors. First, these measurements look at the actual production that a wafer experiences. There is greater fidelity to the actual manufacturing process. Second, this information can be captured by the CIM systems in place and used to analyze contributing factors to yield losses. The results from in-situ monitors can be used to guide the development of equipment modifications.

Management Practices

Manufacturing synergy

Process can be transferred from development centers to more than one manufacturing site during the life of the process. There is therefore an opportunity for synergistic learning to take place between these organizations. The keys to managing this activity among geographically dispersed factories are control, communication and synergy of effort. Strong control of process changes is necessary to prevent migration from a process standard at the various sites. In such a case, opportunities for synergistic learning are diminished, since each location is running a specialized version of the process. Control also begins at the time of the process transfer into manufacturing. The process should be robust enough to be transferred intact to the manufacturing site; alterations should be absolutely minimized.

Instituting a directing group among the factories which run the same process helps provide the control function (especially over authorized changes in the process). It is also a vehicle for communication, where the work in one factory can be disseminated to the other sites. Other means of communication include direct peer-to-peer meetings and reports (both formal and informal) and manufacturing symposia. Electronic transfer of information (email, fax,file transfer) also allows for the rapid disclosure of successful projects.

Finally, coordination of work is necessary to reduce duplication of effort. In certain circumstances it may in fact be wise to have multiple efforts directed on a single issue; however, this should be a planned activity rather than a haphazard occurrence.

Goal Setting and Tracking

Thorough goal setting on key performance measures is critical to motivating improvement activity. Benchmarking should begin early in the process life cycle in order to identify best-in-class performance. Benchmarking both internal and external organizations provides a standard from which to set the organizations objectives. A number of issues should be researched, including management policies, operational methods and performance levels. Comparisons with an internal self-review allows the plant to identify areas on which to focus attention: these areas are generally the ones with the greatest difference with best-in-class performance.

The benchmarking results allow the organization to set realistic, yet challenging, goals. For yield improvement, these goals would address issues such as particle levels, uniformity, etc. A valuable tool for setting these goals is the half-life method of analysis (Appendix E). This analysis can be used to set time period for reducing targeted variables in half. Relatively dramatic results can arise in just a few time periods if targeted metrics are closely matched in the plant.

Key to *sustained* improvement is the careful tracking of results over time. Without a visible method of tracking performance, attention to long term goals suffers. The half-life charts provide a very easily understood rendering of performance towards identified goals. These graphs allow for rapid comparison of performance among various plants. This ease of comparison allows for continuous "internal benchmarking" highlighting both low and high performing factories. This motivates lower performing fabs to use approaches adopted by the higher performing factories.

Prioritizing efforts

At any one time, there are a variety of improvement opportunities to which resources can be directed. Choosing projects on which to focus those resources is an important task. This focus is easier to obtain when one zeros in on the goal of the firm: to make money. Under the paradigm proposed by Goldratt, there are three plant level measures of progress towards this goal: inventory, operating expenses and throughput (Goldratt, 198x). In the early ramp of microprocessors, output is of critical concern. Demand is high and production is constrained. Therefore, the criteria upon which any project must be judged is its estimated effect on output. In order to standardize this calculation across different products, an *equivalent* output measure can be used.

A key task of fab management, therefore, is to get the engineering team to identify and solve important problems (Flaherty, 1990). This requires managers who are well-versed in *both* technical and business matters. The manager must comprehend the process in order to push engineers past quick fixes to fundamental improvements. The manager must also grasp critical business issues and understand how the plant can address them.

The importance of directing project efforts was made clear to me through conversations with process and yield engineers. Yield losses "are driven by surprisingly few issues". These observations were supported by an analysis of yield improvement curves for two of the fabrication areas. The yield improvements (as well as excursions) could be reasonably well attributed to specific actions (events). The implication is that improvement can be accelerated if a finite number of key issues are focused on early in the process. The next section delves more deeply into this issue.

Section 5 • Management Focus

This section outlines the importance of focus in managing yield improvements during production ramping. The first part describes the factory environment during ramping. The second part describes the observed behavior and outlines possible contributing factors. The third part presents a system diagram which helps explain the observed behavior.

Climate in Ramping Semiconductor Factories

The operating environment in ramping factories influences the ability of an organization to focus attention and resources. Hayes and Clark define two terms that help explain the somewhat chaotic nature evident in production ramps: complexity and confusion (Hayes, 1988). The factory system generally becomes more complex as it expands and increases output. There are more orders, personnel, and process equipment to manage. Confusion stems from policies that disrupt the factory stability. Such actions as equipment installation and modification, changes in production schedules, altering the process all contribute to confusion.

The following list details some of characteristics of ramping facilities:

Immediate output goals predominate. Reaching the output goals negotiated for the plant is a crucial measure of success for the factory and its management. I was told that a factory manager could fail to meet other measures (such as throughput time, yields, or cost targets) yet still be deemed successful if she attained her planned die out. This

importance of meeting market requirements is fully understandable (see Section 1). However, if this sense of urgency is not well managed, reliance on short term fixes will prevail. Consequences of the reliance include fixing the problems more than one time and weakening the ability of the organization to focus on more fundamental improvements.

Equipment is installed to add capacity. Managing factory ramping while adding new capacity a fact of life in most American semiconductor fabs. The timing and increment of the installation of new processing equipment influences the ability of the factory to absorb and adapt the new capacity while still focusing other key issues. Some of the effects of adding capacity in small increments over time include continued disruption of the line (as equipment is installed) and added process variation (due to slight differences among sophisticated equipment sets acquired in different time periods).

New workforce is training. Ramping of a new facility will often require the hiring of new managers, engineers and technicians/operators. Dedicated training is necessary to develop the skills needed by the workforce. This training can be accomplished in parallel with the ramp (and therefore relies on a lot of experiential learning) or beforehand, while the site is being prepared for ramping. Tradeoffs between the two approaches include costs associated with early hiring and training, the impact of each role on overall factory performance and the existing skill sets of the workforce.

New process is often immature. Processes are often transferred from the development labs in an "immature" state and require further development. The need for ongoing change means that the process is still evolving as it is being ramped at the manufacturing site. The process changes will cause disruption to the production line.

They can also cause frustration if these changes hurt the ability of the factory to meet its immediate output targets.

New equipment is utilized. New microelectronics processes often require the use of new and more capable production equipment. Older equipment with well known characteristics will be replaced by newer models that have unknown idiosyncrasies. The new equipment renders obsolescent a great deal of the experiential knowledge possessed by the firm's workforce. The loss of this knowledge limits the degree to which existing engineering intuition can be used to devise creative improvement strategies.

Die Yield Behavior in Ramping Fabrication Facilities

Yield improvement behavior can be viewed as a learning process. Traditional approaches relate learning rates to cumulative experience or the time required to absorb and understand new insights. While these factors are important, my view is that learning is a function of conscious and directed effort.

Yield curves from a number of fabrication facilities are graphed schematically on Figure 5. Actual yields and production volume are not indicated in order to protect confidentiality. What is interesting to note in the behavior is the production delay required for sustained yield improvement to take place. This delay can be called the "incubation period". It should be noted that these factories have significantly different local environments, yet all display nearly the same behavior. This suggests that policy structures that cut across the factories of the firm drive this behavior.

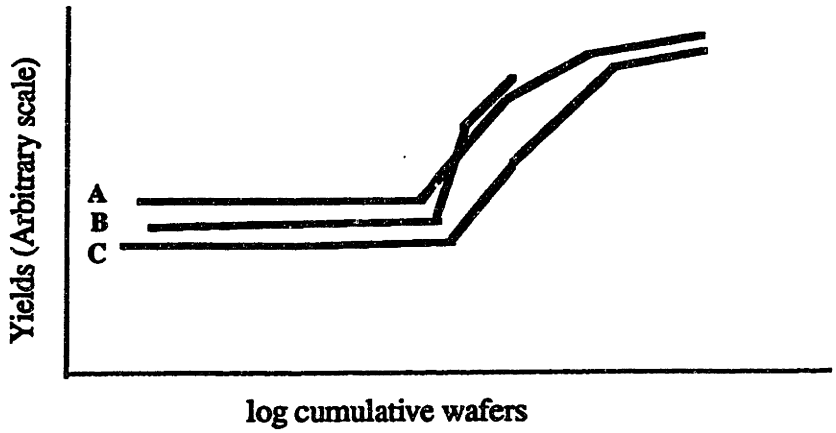


Figure 5.1 Schematic sketch of "incubation period" among three manufacturing facilities.

This observed behavior has a significant influence on performance during the early ramp. Since yield improvements are flat, output can only be enhanced by increases in capacity or utilization. Since capacity is costly in the equipment-intensive microelectronics industry and yields are typically low early in production, there are enormous economic advantages to be had in shortening the observed "incubation period".

There are many different interpretations of this behavior. One is that a certain amount of exposure to repeated problems is required before improvement opportunities can be identified. However, my research has shown that many improvement opportunities are already identified at the development site or lead manufacturing site. Another interpretation is that improvements are simply a function of learning by doing; i.e., that task performance is improved as personnel learn by experience. This view may have a bearing on line yield (where misprocessing is a contributing factor) but would seem to

be less applicable to die yield (where systematic elimination of defects is a result of directed effort).

An alternative hypothesis is that learning is a function of directed effort: yield improvements are the result of applying the knowledge gained through continual learning about the process. In this view, the observed behavior is a function of two factors: the degree of focus placed on yield improvement and the time required to implement yield improvement changes (outlined in the previous section) once an effort is initiated. When resources are focused on routine production issues (responding to excursions, capacity installation) work on base-line yield improvement will lag. In addition, a finite time is still required to complete a project once focussed effort is directed on it. The combined effect delays sustained yield improvement in manufacturing.

The Importance of Focus

Flaherty outlines the importance of a proactive and focused approach on the ability of a manufacturing plant to make rapid improvements (Flaherty, 1992). Table 5.1 outlines her estimate of the resources directed to various activities in "slow" and "fast" factories. The slower performing fabs spends an inordinate amount of time on routine production problems while faster fabs devote considerably more resources to fundamental improvements.

Table 5.1 Differences between "fast" and "slow" factories.

	Routine Production Problems	Introduction of New Technology	Improving Existing Technology
"Fast"	25%	50%	25%
"Slow"	>90%	5%	<5%

The allocation of engineering resources can have a large effect on yield improvement in the fab, since it is the engineers who develop and implement changes to the process and equipment. Interviews with fab personnel suggests that attention and resources are primarily focused on two types of routine production issues: (1) responses to excursions and (2) capacity installation during the observed "incubation period". The statement that "yield was not focussed on until well into the ramp when it became evident that there was a lot of leverage to be gained" is a telling comment on the priorities during the early production ramp.

Flaherty's observations show that slow progress during that period is hindered by policies that direct resources to normal production problems. The implication of her distinction between fast and slow factories is that management can shorten the "incubation period" by applying a greater focus on yield improvements.

That separate fab facilities show the same dynamic performance suggests that the structure of management systems significantly influences behavior. The simple systems diagram in Figure 5.2 illustrates a view of the key contributors of the yield improvement behavior ("incubation period") during the early ramp. The diagram is

based upon the analysis of the previous sections. It attempts to explain how the effects of focus and yield improvement cycle time work together to drive the observed behavior.

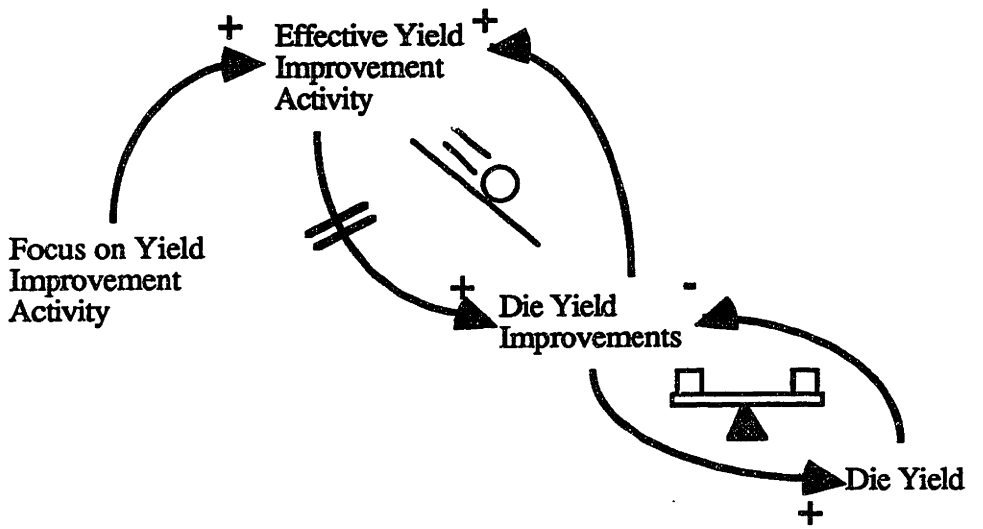


Figure 5.2 System Diagram of Die Yield Improvement

The components of this diagram are easily understood. The key behavior of interest is the actual base-line yield (die yield). Yields are increased through successful improvement activities. Higher yields become more difficult to achieve as they approach their theoretical maximum. Die yield improvements were the direct result of effective die yield activity. A delay is noted in this relationship and corresponds to the time required to accomplish the tasks outlined in Section 4. Effective die yield activity increases as knowledge gained from results and focus on this activity increase.

A key limiting factor shown in the diagram is the level of resources dedicated to yield improvement activity. The level of resources is a direct function of the focus

management places on this activity. The next section proposes strategies for improving the focus on and resources devoted to yield improvement during early production.

Section 6 • Improving the Manufacturing Focus

According to the model forwarded in Section 5, two strategies for reducing the incubation time include increasing the focus on yield improvement and decreasing the cycle time of this activity. A number of approaches for enhancing yield improvement activities were introduced in Section 4. This section looks at intensifying the focus on yield improvement activity by reducing the diverting effects of other activities. In particular, I address capacity additions and fire fighting (excursions).

Capacity Decisions

A number of different policies govern production ramping. I examine policies that determine the incremental quantities in which capacity is added. There are two distinct schools of thought on how to outfit a fab that will ramp to high production. At one extreme, there are factories in which the total planned equipment set is installed before production begins. This strategy requires a large initial investment that would seem especially risky to most American manufacturers. However, in spite of this "risk", some Japanese semiconductor companies have used this approach to quickly ramp their manufacturing plants.

At the other extreme, a firm could install a very limited equipment set and gradually add capacity. Capacity will be added in single increments when a particular type of machine (e.g., etcher, diffusion tube) becomes the limiting constraint. This "spider web" approach to capacity addition is illustrated in Figure 6.1.

Finally, there is an intermediate approach where somewhat less than the ultimate equipment set is installed and capacity is added in larger blocks of equipment. Some of the benefits of this approach are presented later in this section.

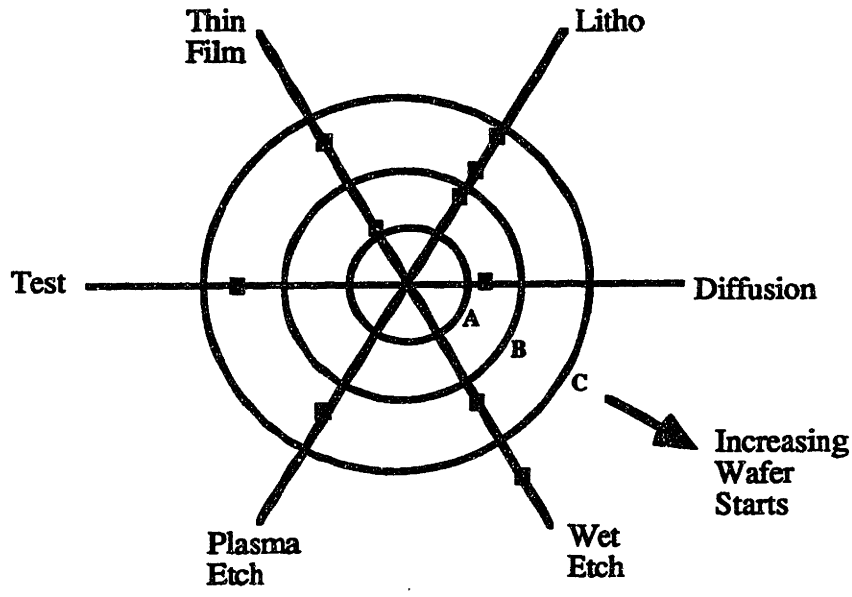


Figure 6.1 "Spider Web" Capacity Addition Model. Individual equipment are incrementally added as production ramps (as indicated by the square symbols on the spokes). For example, when ramping from B wafer starts per month to C wafer starts per month, the plant will first add lithography capacity, then etch capacity, then tester capacity and so on.

Side effects of capacity installation

Adding capacity requires an expenditure of resources (space, money and time) and creates many effects other than the planned one of increasing output. An obvious example is interference; moving and installing equipment into the clean room area disrupts the fab line operations. This interference can affects many areas, including work flow and team structures. In addition, installation can result in an increased generation of ambient particles in the clean room.

The introduction of equipment that differs slightly from existing machines can lead to increased process variation. For example, lens systems for lithographic systems are produced in batches and therefore differ slightly from batch to batch. Purchasing lithography equipment over time can result in lens systems that slightly deviate from on another. This divergence increases process variation and thus has a negative effect on yields.

Figure 6.2 illustrates the effect of incremental capacity additions on output. Capacity is added as the ramp in planned output increases. The decision to add capacity, after delays due to installation and qualification, ultimately enables greater output for the system. However, the increase in variation and interference serve to reduce output (through misprocessing and lower yields). These negative effects can result in lower than expected output (at least initially) for the system.

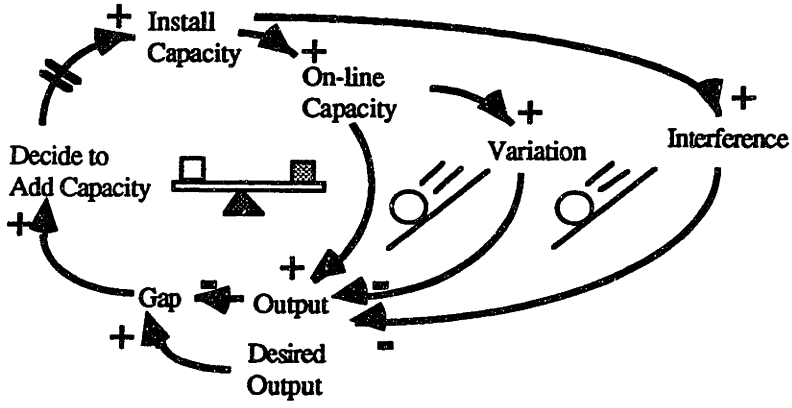


Figure 6.2 Effects of an Incremental Capacity Addition Policy

The implications of this analysis of an incremental capacity policy include:

- Capacity installations continually consumes the attention of fab engineers and managers.
- Each installation creates disruptions on the line (and increases the level of "confusion")
- Process variation is increased as new equipment is installed.

The next subsections detail approaches which address these concerns.

Use of External Personnel

Estimates on the amount of resources devoted to capacity addition during the ramp ranged up to 70% in some fabrication facilities. Reducing the allocation of personnel resources to routine capacity installation can be accomplished through increasing the involvement of outside employees, including:

- Greater involvement of vendors in site preparation and installation.
- Use of contractors to execute routine equipment qualifications.

The utilization of outside personnel keeps the firm's engineers and technicians from being pulled away from long-term projects for routine capacity additions. This enables focused yield improvement to proceed with less disruption. The utility of this approach is shown through the experience of one fab area, where the percentage of engineering time spent on installs declined even as installation increased. This trend is shown schematically in Figure 6.3 and is due to the use of vendor/contractor personnel.

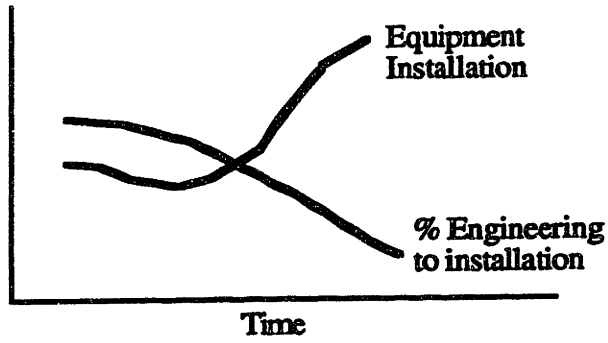


Figure 6.3 Effects of Using Contractors for Equipment Installation.

The use of external resources helps alleviate some of the personnel diversion that occurs during each installation. However, it does not address the disruptive effects. In addition, it does not eliminate the need for "management overhead" to plan and supervise the installation of each piece of equipment.

Leveraging equipment installations

Adding capacity in larger increments would reduce the total management and engineering resources allotted to this activity. Each equipment addition, no matter how small, requires a finite use of critical resources. Adding the capacity in larger, pre-planned blocks would reduce the frequency and total amount of "engineering overhead" that is devoted to equipment installation over the course of the ramp. In addition, the installations can be less disruptive to manufacturing operations if they are bundled together rather than released incrementally to the production floor. Figure 6.4 illustrates the distinction between incremental and clustered capacity addition.

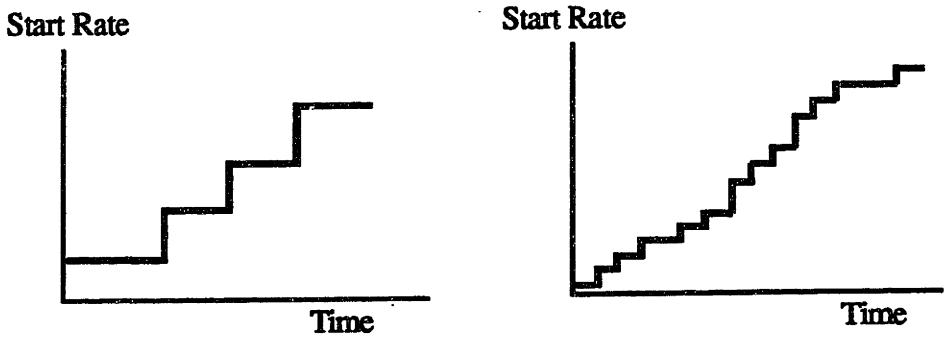


Figure 6.4 Capacity Addition Policies. Adding capacity in larger quantities (left) results in fewer installation periods as compared to incrementally adding capacity (right). It can also provide for distinct periods of time for adapting to the additional equipment.

Combining equipment installation into concentrated time periods can also be used to create a window of opportunity for technical and managerial change. Window of opportunity refers to the short time period, after the introduction of a new technology or resource, that organizations can use to modify systems (both technical and operational) before routine application creates inertial barriers to further change (Tyre, 1992). It was noted earlier in Section 4 that yield improvements are derived from discrete projects, rather than from continuous fine-grained improvement. In a similar vein, manufacuring organizations can revise and improve their operations (actions which are often impeded during output ramping because of production pressures) by capitalizing on the short periods of intense capacity installation to make necessary adaptations.

How can organizations leverage capacity installations to creating periods of adaptation? Management must first use the interim periods of relative stability to study the effects of previous solutions and to collect data for the upcoming round of managed adaptation. They can then take advantage of the relatively short periods of opportunity by using them to focus attention and resources on solving problems which have been

accommodated by the day-to-day operations. Capacity installation cycles can therefore be effectively coupled with the 4-phase PDCA cycle and help foster continuous improvement throughout the production ramp.

This process of cyclic adaptation has a number of advantages. First, tackling groups of problems/opportunities at a time can force consideration of the interdependencies that exist in the system. Instead of identifying local optimums for individual problems (and thus creating sub optimal routines for the system), the relationships between issues can be examined. This consideration of a set of solutions is especially relevant in when dealing semiconductor manufacturing processes, which inherently have a high degree of interdependence.

A second advantage is the ability to test the effectiveness of modifications during the period of relative stability between episodes of adaptation. Section three presented an analysis of the "noisy" environment characteristic of wafer processing facilities. One consequence of this noise is the difficulty of collecting and analyzing data in order to make valid assessments. This task is made more challenging (if not impossible) in an environment of constant change. Therefore, the periods of relative calm create a valuable opportunity for learning and reflection.

Considerations

A manufacturing firm must analyze a range of issues when devising a capacity addition strategy. These issues include equipment costs, process variability, interference from installations, etc. The firm must consider tradeoffs among these variables when developing its capacity expansion policy. It also needs to insure that its chosen policy fits the overall business goals.

A complicating factor in making these tradeoffs is the mixture of quantifiable and intangible variables. For some issues, it is relatively straightforward to develop hard numbers. For instance, the cost of equipment and the volume discount savings are readily available. The savings (associated with the time value of money) from incrementally purchasing and installing equipment is also easily calculated.

A strictly quantitative analysis cannot be made for many other issues. For example, buying equipment in larger quantities can insure that they were manufactured in the same period. These machines are more likely to be well matched than machines purchased from different manufacturing periods. One example provided to me concerned stepper lens systems. These lens systems are typically fabricated in batch processes; buying lithographic equipment in the same time period helps insure that the lenses are from the same batch.

Interference from equipment installation is another issue that is hard to quantify. This interference (including interference in the fab area and the diversion of management focus) will be a constant source of disruption under an incremental capacity addition model. Adding capacity in larger increments would restrict this source of confusion in the fab area to specific time periods.

Finally, adding capacity in large increments helps create opportunities for change. These opportunities allow the organization to exploit learning gained from stable periods and implement necessary changes. The benefits of gaining this ability to make adaptations are again difficult to quantify. They are valuable only in so far as aid the overall factory strategy for continuous improvement.

Excursion Policies

Since output considerations are dominant in the early ramp, yield excursions demand and receive immediate attention. Having to continually respond to excursions, just to regain normal yields, is a serious hindrance to long term yield improvements. This sub-section addresses effective policies for solving excursion issues.

Planning excursion responses

The key to effective excursion response is a well thought out recovery plan. A generic flowchart outlining steps for recovery can be constructed and used as a template for specific events. In addition, the roles and responsibilities at all levels (operator, technician, engineer, manager) and functional areas (e.g., process engineering, operations) should be clearly defined *before* an excursion occurs. In this way, the organization is prepared to correct the situation and not hobbled by confusion over which individuals and groups are responsible for various tasks.

Documenting excursions

Thorough and complete documentation of all excursions (events, cause, solution, methods used, etc.) helps build "road maps" that all the factories of the firm can use. This synergy among various fabs avoids expending resources to "reinvent to wheel"; i.e., solving problems that have already been solved. Excursions can be complex phenomena where corrective actions can lead to unintended side effects. Learning from the documented experience of other excursion events (which occurred either at the same site or at another factory) aids solving excursions using confirmed methods.

A more fundamental rationale for documentation is to capture the knowledge that is locked inside the heads of individual engineers. This information is rather "sticky" or hard to transfer from one individual to another. Organizations can develop pockets of knowledge, where tacit and systemic knowledge about the process are isolated in key individuals. Well structured documentation of excursion events helps "unstick" this knowledge and allow for it to be readily accessed by the rest of the organization. In the event that a more intuitive understanding of the issues are required, the local site can identify and involve the previous problem solver ("the expert") with the current problem.

Response Methodology

The engineers responding to an excursion face a choice between two general types of solutions. They can address the symptom (quick fix) or attack the root cause (fundamental solution). The benefits of a quick fix approach are initial savings in time and effort. The drawbacks can include generating unanticipated side effects (since the root causes and interdependencies have not been fully examined) and having to respond to the same problem again (repeatedly fixing the same problem).

The more powerful approach is to correct the root cause of the problem. However, the likelihood that an organization would adopt this approach is dependent on the existing management policies of the factory. People will not generally follow this course of action unless they (1) have the skills to accomplish the task and (2) have motivation to do so. In the cause of addressing root causes of excursions, plant personnel must first be trained in those methods. The measures on which employees are judged must then be designed so as to encourage the desired behavior.

As an example, consider the case of a machine operator whose machine malfunctions. One possible response would be merely to replace the broken part. Perhaps the machine breaks down again, perhaps not. Another more proactive response would be to find *why* the part broke and how to *prevent* it from happening again. This thorough approach is likely to take more time (at least initially). If the operator is not trained in root cause analysis or is measured solely on output, the "band-aid" solution will predominate.

A pattern of rewards for firefighting rather than long term problem solving may help instill a penchant for short term solutions. Figure 6.5 illustrates one view of this dynamic.

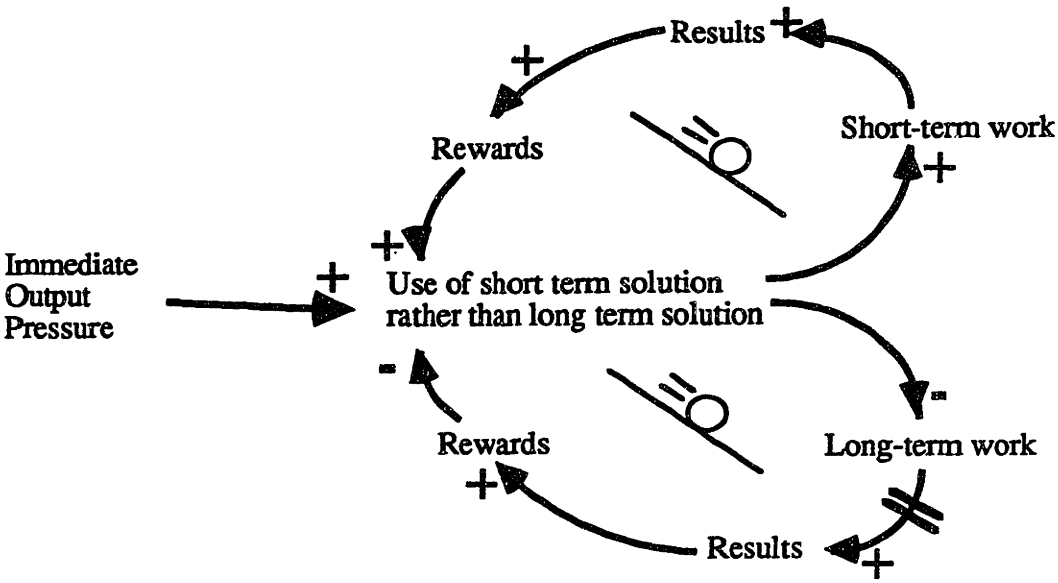


Figure 6.5 Predominance of Short Term Solutions. Pressure for immediate output and factory rewards systems encourage the use of "band-aid" solutions. Fundamental solutions are not investigated and key learning is not advanced. The short term loop is strengthened while the long term loop is weakened, thus reducing opportunities for true improvement.

If the emphasis on current output is high, then the preferred use of short term solutions will prevail. This situation is exacerbated if short term fixes (successful firefighting) are better rewarded than long term fixes (which, ironically, may not even be recognized since they *remove* potentially dire -- and noticeable -- problems). Over time, the organization gets "addicted" to firefighting, but its ability to pursue long term solutions atrophies since that approach is not exercised. Because both individuals and the organization lack fundamental problem solving skills it becomes very difficult to reverse this situation.

Proactive Measures

Most of the above discussion focused on enhancing the *response* to excursions. Excursions can also be *prevented* by proactive activities within manufacturing (e.g., preventive maintenance, raw materials quality assurance, reduction of misprocessing). In addition, the strict controls on engineering process changes that were mentioned in Section 5 reduces the likelihood of unintended excursions.

However, a large percentage of excursion events cannot be adequately prevented by actions taken within the factory system. The more preparatory problem solving that is accomplished before the process is transferred to full-scale manufacturing the less likely hidden problems will arise and the more the manufacturing organization will be able to focus on long term improvements. This problem solving is optimally addressed only when the larger development-manufacturing system is considered. The following section concerns this larger system.

Section 7 • Manufacturing/Development Interface

In order for manufacturing to focus on long-term yield improvement, frequent exposure to diverting yield crashes (excursions) must be minimized. This section briefly explains how the way development organizations bring a new process to the factory affects the manufacturability of the process. The first part considers the role of the development center. The second part relates an example of process development and ramping from the steel industry in Japan. The third part discusses a variety of implications for semiconductor manufacturing.

Role of Development Facilities

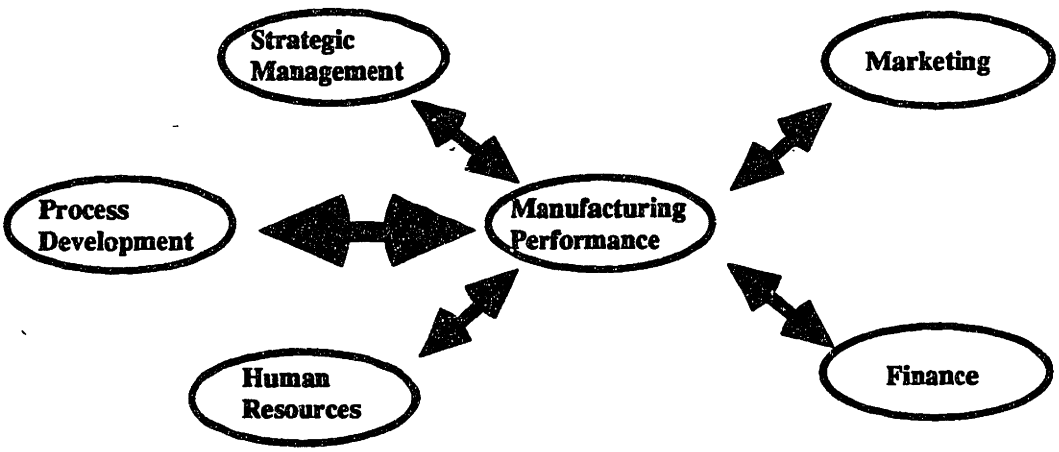


Figure 7.1 Observing the Wider System. The total factory system determines ultimate factory performance. Particularly tightly coupled are development and manufacturing.

Fab performance is determined by the total company system, including process development, marketing, finance and strategic management (Flaherty, 1990). I

consider both the development and manufacturing organizations, therefore, because of the need to analyze a larger system than just an individual manufacturing plant. Process development and manufacturing performance are especially tightly coupled: an uncharacterized and immature semiconductor process creates great obstacles to meeting manufacturing objectives. Figure 7.1 illustrates this view of manufacturing performance.

The traditional role of process development has been to deliver a new process that yields at "manufacturable levels". Or put another way, it is to bring the new process to a point where the die per wafer yields are high enough to make manufacturing the product economically feasible. The emphasis is on developing process technology at in a stable environment (e.g., at low volumes). Under this development paradigm, planning and designing of operational methods such as maintenance, automation, and training are issues that remain for manufacturing to develop.

This overriding goal of meeting yield targets directly competes with another activity: surfacing potential problems. Surfacing potential manufacturing obstacles requires stressing a system that is closely aligned with the factory environment. This instability works against getting the process up and running smoothly *in the development center.* However, exposing the problems early in development helps avoid stabilization of a sub-optimal process, much in the way that reducing inventories "exposes the rocks" in manufacturing operations.

Congruence between Manufacturing and Development

Fidelity is the degree of congruence that exists between actual high volume manufacturing facilities and the environment where the process is developed (Bohn,

1988). Process development in facilities with a low fidelity to manufacturing is likely to lead to a sub-optimal process (i.e., the process is optimized for the low volume development environment). In order to expose and solve potential production volumes, development operations should have a high degree of fidelity to manufacturing operations for some period of time.

Why? Some problems only arise in a full scale manufacturing environment. Full scale manufacturing differ from development in many areas, including personnel policies, equipment utilization and operational procedures. Replicating these conditions in the development plant as the process matures increases fidelity. However, this approach can also slow the rate of learning. Appropriate tradeoffs need to be made in order to determine the level and timing of increased fidelity in development facilities.

The important issue is the state at which the process is transferred. Process transfers that are "thrown over the wall" from development to manufacturing are likely to result in processes that face significant problems in the manufacturing environment. Instead of a distinct development to manufacturing transfer stage, there should be a seamless evolution (Anastasio, 1991). This approach reduces problems stemming from lack of fidelity.

Figure 7.2 illustrates how seamless evolution from development to manufacturing can be accomplished by increasing the fidelity of the technology development facility throughout the course of process development. Traditional transfer of process technology occurs when the process shifts from a low fidelity environment (development facility) to a high fidelity environment (manufacturing plant). The lack of congruence increased the chances that significant modifications to the process would be required once it was running in the manufacturing plant.

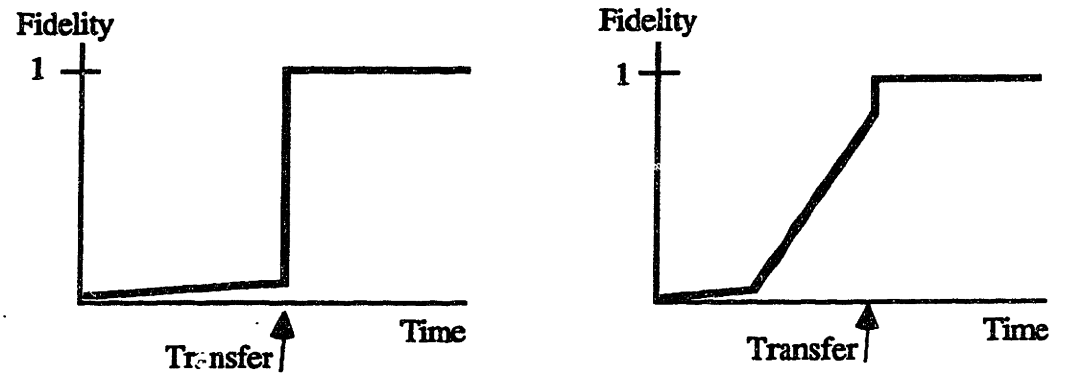


Figure 7.2 Development to Manufacturing Transition. Traditional transfers occur between very dissimilar environments. An alternative strategy is to increase the fidelity of the development environment to manufacturing and thus avoid devising a sub-optimal process.

A more successful process transfer can be accomplished by increasing the fidelity of the development facility. This increased fidelity is realized, in part, by increasing the rate and volume of wafer starts, developing production-oriented operational methods (such as preventive maintenance, scheduling, metrics) and standardizing equipment.

This thesis suggests the need for a greater focus on manufacturability in technology development organizations. The charter of the development center becomes two-fold: to develop both the technology *and* the operations for the next generation process. A main objective is to *align* both of these components during development. This goal is accomplished through a mutual adaptation of technology and organization (operational methods, user environment, etc.) Figure 7.3 illustrates this approach. The adaptive cycles help correct poor fits in technology, manufacturing systems and performance metrics.

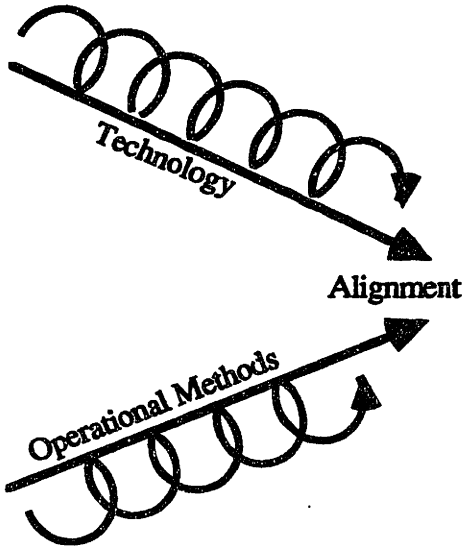


Figure 7.3 Mutual adaptation of technology and operational methods. Technology and operational methods evolve (through adaptive cycles) so that misalignments are reduced.(adapted from Leonard-Barton, 1988)

A number of factors support this expanded role for technology development centers. Three important considerations are:

The transfer of a process better optimized for manufacturing

By better aligning the process to the manufacturing environment, the process can ramp with fewer "unexpected" problems. This reduction in firefighting situations allows manufacturing resources to focus on long term improvements.

A window of opportunity exists before production ramping

As was shown in Figure 7.3, development centers have the opportunity to use repeated cycles of adaptation to align technology and user environment (operations). Volume manufacturing organizations have less flexibility to make changes once they are committed to ramping output. The window close once production begins due to:

- Pressures to produce at high utilization
- Strong institutional barriers to change (e.g., engineering change board)
- Long process throughput times in the early ramp (and thus long information cycle time for experiments on proposed changes)

The "stickiness" of knowledge in process development

Much of the knowledge of the process is difficult to detail and transfer across organizations (or even, at times, from engineer to engineer). One effective strategy of solving the problem of "sticky" data is to have ready access to the "experts" (Ogawa, 1990). The technical personnel who developed the process are located at the development site and can be easily tapped when problems arise.

In summary, additional responsibilities are proposed for development centers so that full-scale manufacturing can rapidly ramp and focus on fundamental improvements. These objectives include:

- Developing operational standards such as
  - PM schedules
  - Automation
  - Performance metrics
- Uncovering "hidden" problems driven by
  - Intensity factors
  - Volume factors
- Increasing the fidelity over time with the manufacturing organization

The proposed approach better aligns the performance metrics of both development and manufacturing. This congruence encourages managers from both organizations to

improve overall system performance, rather than seeking locally optimized performance. It also allows for more successful "templating" of the process from development to manufacturing, since the process is adjusted for the manufacturing environment. While this by itself does not enable a fully "turnkey" transfer, it does improve the probability of success for a "direct copy" approach to process transfer.

These new tasks present significant challenges to existing development organizations. The complexity for development personnel is increased by (1) having more tasks and (2) executing them simultaneously. This requires that development centers acquire new skills and resources. Significant risks can be associated with the ability of these organizations to successfully manage this complex role.

Example: Japanese Steel Industry

Policies that enable development organizations to gain better understanding of how a given process will succeed in a manufacturing environment are important for limiting exposure to problems during production ramping. Ogawa (1991) presents an approach used in a Japanese steel company.

This example concerns a continuous annealing line (CAL) at a Kawasaki steel plant. Engineers noted that large problems would often arise a number of months into production ramping when the output reached high levels. Resolving the problem at that stage was difficult since (1) engineers who developed the process had dispersed and (2) there was little time available to run experiments on the production line. The engineers therefore took longer than expected to correct the problem.

A strategy named "rapid max" was developed to address this situation. Under the rapid max strategy, the process system is stressed to maximum speed *before* production ramping. The tests are carried out during a short period of time or on just one part of the line. This approach is a very efficient way of detecting potential problems.

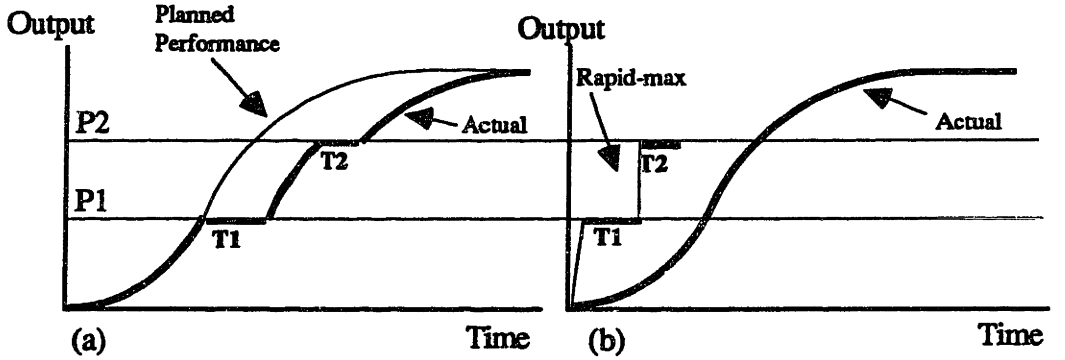


Figure 7.4. **Effect of Rapid-Max Strategy. In (a) problems surface at output volumes (or rates) P1 and P2. The problems take times T1 and T2, respectively, to solve. In (b) problems are surfaced by rapid-max approach before the ramp reaches that level. Delays are avoided.**

The rational for the rapid max strategy is illustrated in Figure 7.4. Assume that fundamental but unknown problems will arise at certain load levels. Under normal process development and start-up, these problems will not surface until production is ramped to high levels. Uncovering the problems during the production ramp causes delays in planned output and the resulting firefighting detracts from long-term effectiveness of any ongoing improvement activity.

A rapid max strategy detects problems *before* production ramping. These problems can then be solved before they affect production performance and thereby avoiding the delays mentioned above. The time required for production ramping lessens.

Ogawa draws a clear distinction between *early start up* and *early stabilization*. Early startup is achieved when the production line runs at capacity in a reliable fashion. Stabilization is achieved when the process is operated without difficulties at output lower than ultimate production. This is often linked to slow startup, since the process may be stabilized for operating conditions which do not match the ultimate manufacturing environment.

Implications for Semiconductor Manufacturing

A number of observations from this example drawn from the steel industry can be applied to semiconductor manufacturing. Among these are:

The rapid-max approach helps raise killer issues earlier.
By exposing and solving these issues before they impede production ramping, this approach reduces the probability of exposure to excursions that demand immediate attention (firefighting) and draw resources from effective long term improvement opportunities. Solving these problems takes less time before full-scale production since equipment and "experts" are more available.

Reducing potential problems makes direct copy transfer more practical.
The process and operations are modified to solve the problems that appeared during rapid-max projects. The overall robustness of the process is thus increased and will help insure that a process developed in an environment with less than perfect fidelity to manufacturing will still have a reasonable chance to succeed without major hang-ups.

Rapid-max in development aids in problem solving using "sticky data"
Systemic and tacit knowledge of the process is difficult to transfer from development. This is knowledge that is not easily captured in written specifications and process flows. Therefore, the development location becomes the preferred sight for rapid-max in order to tap into the knowledge of resident "experts". The presence of this information helps accelerate the problem solving process.

Opportunity for rapid-max strategies also exists in manufacturing.
The initial test runs before fab startup are a good period for rapid-max experiments. Though the process and methods are more fixed than they were earlier in development, flexibility still exists to make adjustments for problems uncovered using this approach. Additional windows for opportunity can be created if capacity is added in large "blocks"; the period before the additional capacity is fully committed to production output is time which can be devoted to rapid-max experiments. Both these time periods give factory management a chance to identify and resolve potential problems before they become serious stumbling blocks to meeting output commits.

Section 8 • Summary and Conclusions

The main focus of this research was to develop an understanding some of the key limiters of die yield improvement during the early production ramp. Die yields have a major impact on production output in semiconductor manufacturing. Since products may yield at only a few percent during initial production, enormous leverage can be gained from die yield improvements.

Process die yields demonstrate an incubation period (as measured by time or cumulative production volume) before sustained rates of improvement are achieved. This observation was made for a number of fabrication plants, suggesting that local conditions were not driving the behavior. Rather, these observations indicate that more systemic causes stemming from the general ramping policies influence performance and give rise to the incubation period.

This research combined grounded theory and systems thinking approaches to analyze the available data. Data included information from both clinical and statistical sources. The clinical data from this project was used to explain patterns of behavior and to provide insights for building systems models. Statistical data was collected from automated databases and standard reports.

High noise levels in die yields within individual lots of wafers make full-loop split lot experiments problematic for identifying and managing improvements. There are high levels of variability within each batch (the standard deviation can be on the order of 25% during the initial ramping phases). This variability fluctuates from batch to batch.

Therefore, rather large improvements (on the order of 4-6 %) from split lot experiments are required to meet even an 80% confidence interval.

Full loop experiments can not be the primary tool for yield learning; a large amount of potential learning from split lot experiments is overlooked because of process noise. Power functions for a single fabrication plant at three time intervals were generated via a bootstrapping technique. This procedure was applied using a simple test criteria: if $\Delta Y_{exp} > 0$, then accept the process change. This analysis showed that only very large yield improvements would be correctly identified. The effect of noise in the factory is dramatic; the probability of missing a 1% improvement ranged from 40 to 44% or not much better than pure chance. Adopting a decision rule that meets the 80% confidence interval further reduces the probability of detecting a true process improvement through the use of split lot experiments.

An enhanced learning environment can be achieved by reducing the sensitivity of the process and the product to variation. Robust process design can limit the effects of disturbances on the process. Design for manufacturing approaches can improve the fit of specific products into the process window. Both methods require strong feedback from manufacturing yield engineers to product designers, an interaction which has traditionally not been very strong.

A variety of tactics can be used to enhance learning in noisy manufacturing environments. Many approaches, such as multibatch and sequential experiments, have serious drawbacks with respect to information cycle time and the use of production resources. Two alternative tactics are (1) the use of computer integrated manufacturing data in "natural experiments" and (2) short-loop experiments. CIM data reduces the effect of noise by increasing the sample size and allows for common factors for low

yields to be identified. The benefits of short-loop experiments include fast information cycle time, lower experimental noise and reduced costs of experimentation (as compared to full-loop experiments).

A components of a simple systems diagram can explain one view of the key contributors of the observed yield improvement behavior. The system diagram illustrates the interrelationships between various factors. The key limiting factors shown in the diagram are (1) delays during improvement activity and (2) focus on yield improvement activity.

A significant amount of time (4-8 months), between the identification of a yield improvement opportunity and the eventual implementation of a verified improvement, is required for the completion of yield improvement projects. This results in a significant delay for yield learning. In addition, many known improvement opportunities are not pursued for long periods of time because resources and attention are directed elsewhere. The implication of these observations is that unless yield efforts are identified and focused upon before the production ramp, die yields during the critical early phase of the ramp are likely to remain flat.

Reducing the delay between identifying improvement opportunities and implementation of changes is key to developing more effective yield learning. The use of tools such as yield models and in-situ monitors can accelerate yield learning. The following management practices also have a tremendous effect on yield improvement work. Capitalizing on manufacturing synergy between fabrication facilities running similar processes can enhance yield improvement rates via information sharing and the reduction of duplicated efforts. Thorough goal setting and tracking of key performance

measures motivates improvement activity. Prioritizing efforts based upon likely impact on output helps focus resources on important tasks.

Slower improving fabs spend an inordinate amount of time on routine production problems while faster fabs devote considerably more resources to fundamental improvements. The allocation of engineering resources has a large effect on the yield improvement rates in the fab, since these engineers develop and implement changes to the process and equipment. This research suggests that attention and resources during the observed "incubation period" are primarily focused on two types of routine production problems: (1) capacity installation and (2) responses to excursions. Base-line improvements in yield are not focused on until well into the ramp.

Adding capacity requires an expenditure of resources (space, money and time) and creates many side effects. Among these are increased process variation caused by the slightly varying operating characteristics of equipment and interference to the manufacturing line during equipment installation. Policies that call for adding capacity in small increments (such as the "spider web" model) result in equipment installations that continuously consume the attention of fab engineers and managers. Estimates of the amount of resources devoted to capacity addition under these circumstances ranged up to 70%.

Reducing the allocation of engineering resources to routine capacity installation can be accomplished through increasing the involvement of outside employees. Examples of this approach include greater involvement of vendors during installation and the use of contractors during equipment qualification. The utilization of outside personnel keeps the firm's engineers and technicians from being pulled away from long-term projects

for routine capacity additions. This enables focused yield improvement to proceed with less disruption.

The use of external resources alone will not eliminate the need for "management overhead" to plan and supervise the installation of each piece of equipment. Adding capacity in larger increments, however, would reduce the required management and engineering overhead for capacity addtion. Installations are also less disruptive to manufacturing operations if they are bundled together rather than continually performed one at at time on the production floor.

Combining equipment installation into concentrated time periods can also be leveraged to create a "window of opportunity" for technical and managerial change. Windows of opportunity refers to the short time periods, after the introduction of a new technology or resource, that organizations can use to modify systems (both technical and operational) before routine application creates inertial barriers to further change. Manufacturing organizations can revise and improve their operations (actions which are often impeded during output ramping because of production pressures) by capitalizing on these periods of intense capacity installation to make necessary adaptations. Management can then use the interim periods of relative stability to study the effects of previous solutions and to collect data for the upcoming round of managed adaptation. These cycles of capacity addition become part of a 4-phase PDCA cycle and thus foster continuous improvement during the production ramp.

This process of cyclic adaptation has a number of advantages. First, tackling groups of problems/opportunities at a time can force consideration of the interdependencies that exist in the system. Instead of identifying local optimums for individual problems (and thus creating sub optimal routines for the system), the relationships between issues can

be examined. A second advantage is the ability to test the effectiveness of modifications during the period of relative stability between episodes of adaptation. The periods of relative calm create a valuable opportunity for learning and reflection.

Continually responding to excursions in order to regain normal yields is a serious hindrance to long term yield improvements. However, since output considerations are dominant in the early ramp, yield excursions demand and receive immediate attention. The key to effective excursion response is a well thought out recovery plan. Roles and responsibilities, as well as the response sequence, should be clearly defined before an excursion occurs. This preplanning reduces the confusion over which individuals and groups are responsible for various tasks and allows for a more focused effort.

The engineers responding to an excursion face a choice between two general types of solutions. They can address the symptom (quick fix) or attack the root cause (fundamental solution). The probability that an organization will adopt the latter approach is dependent upon the existing management policies of the factory. Plant personnel must first be trained in those methods. The measures on which employees are judged must then be designed so as to encourage the desired behavior.

A pattern of rewards for firefighting rather than long term problem solving will help instill a inclination for short term solutions. This pattern is particularly likely to occur if the emphasis on immediate (today's or this week's) output is high. Long term solutions may not even be recognized since they remove potential problems before they are noticed. Over time, the organization can become "addicted" to firefighting as its ability to pursue long term solutions atrophies from underuse.

Thorough and complete documentation of all excursions (events, cause, solution, methods used, etc.) helps build "road maps" that all the factories of the firm can use. Learning from previously documented excursion events (which occurred either at the same site or at another factory) aids problem solving by using confirmed methods and avoids expending resources to "reinvent to wheel" (i.e., solving problems that have already been solved). It also helps to capture the knowledge that is locked inside the heads of individual engineers. Organizations often develop "pockets of knowledge", where tacit and systemic knowledge about the process are isolated in key individuals. This leaves the organization vulnerable if the individual leaves. Well structured documentation of excursion events allows their knowledge to be captured and then more readily accessed by the rest of the organization.

In order for manufacturing to focus on long-term yield improvement, frequent exposure to diverting yield crashes (excursions) must be minimized. A large percentage of excursion events, however, cannot be effectively prevented by actions (such as preventative maintenance, training of raw materials suppliers, reducing misprocessing errors) that take place solely within the factory system. These problems are optimally addressed only when the larger technology development-manufacturing system is considered.

An important issue concerns the fidelity (the degree of congruence) that exists between actual high volume manufacturing facilities and the environment where the process is developed Process transfers that are "thrown over the wall" from development to manufacturing are likely to result in processes that face significant problems in the manufacturing environment. A more successful process transfer (via a seamless evolution from development to manufacturing) can be accomplished by increasing the

fidelity of the technology development facility throughout the course of process development.

This thesis suggests the need for a greater focus on manufacturability in technology development organizations. Process development and manufacturing performance are especially tightly coupled: an uncharacterized and immature semiconductor process creates great obstacles to meeting manufacturing objectives. The charter of the development center becomes two-fold: to develop *both* the technology and the operational methods for the next generation process. These two components can be aligned through the mutual adaptation of technology and organization (operational methods, user environment, etc.) The adaptive cycles help resolve poor fits among technology, manufacturing systems and performance metrics.

The traditional role of process development organizations has been to advance a new process to the point that yields are at "manufacturable levels". This goal of meeting yield targets can directly compete for resources and attention with yet another important activity: surfacing potential problems. The instability created when surfacing manufacturing problems is at odds with getting the process running smoothly and at desired yields while in the development center. However, exposing the problems early in development by stressing a system that is closely aligned with the factory environment helps avoid stabilization of a sub-optimal process (that is, one optimized for a low volume development environment).

A "rapid-max" approach used in the Japanese steel industry addresses this situation. Under rapid max, the process system is stressed to maximum speed before production ramping. This strategy detects problems before volume production. These problems can then be solved before they affect production performance (that is, become

excursions). Under traditional process development and start-up, these problems often do not surface until production is ramped to high levels, resulting in firefighting that draws resources from long-term improvement activity.

Identifying and solving problems before full-scale production takes less time since both equipment and the "experts" are more readily available. Reducing potential manufacturing problems also makes a "direct copy" process transfer more effective. Therefore, under these considerations, the development location becomes the preferred sight for the rapid-max strategy. However, the initial test runs in the fab before startup are also a good period for rapid-max experiments. Additional windows for opportunity can be created if capacity is added in large "blocks"; the period before the additional capacity is fully committed to production output is time which can be devoted to rapid-max experiments. These time periods give factory management a chance to identify and resolve potential problems before they become serious stumbling blocks to meeting output commits.

In summary, the development organization receives a broader range of responsibilities in order to enable full-scale manufacturing to rapidly ramp and focus on fundamental improvements. Responsibilities include increasing its fidelity with manufacturing over time, developing operational methods and uncovering "hidden" problems driven by intensity and volume related factors. The proposed approach better aligns the goals of technology development and manufacturing. This congruence encourages managers from both organizations to improve overall system performance, rather than seeking locally optimized performance.

Bibliography

Adler, Paul, "Shared Learning", *Management Science*, Vol. 36, No. 8, August 1990.

Adler, Paul and Kim Clark, "Behind the Learning Curve: A Sketch of the Learning Process", *Management Science*, Vol 37, No. 3, March 1991.

Anastasio, Frank and David Miller, "Technology Development: Focus on Manufacturability", *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, Vol. 14, No. 3, September 1991.

Benfer, Richard and Chip Burczak, 1993,"A systems Thinking Approach to Manufacturing" Proceedings of Intel Manufacturing for Excellence Conference, San Diego, CA .

Bohn, Roger, "Noise and Learning in Semiconductor Manufacturing", Working Paper, March 1991.

Bohn, Roger, "Learning by Experimentation in Manufacturing," Harvard Business School working paper 88-001, June 1987.

Borden, Peter and Lawrence Larson, "Benefits of Real-Time, In-Situ Particle Monitoring", Proceedings of 1989 IEEE International Semiconductor Manufacturing Symposium, Burlingame, CA .

Bradbee, George et al, "Early Manufacturing Involvement in New Process Technology", Proceedings of 1989 IEEE International Semiconductor Manufacturing Science Symposium, Burlingame, CA .

Burchill, Gary, "Inductive System Diagrams", Working Draft, October 1992.

Chew, W. Bruce, "Productivity and Change: The Short Term Effects of Investment on Factory Level Productivity", Unpublished Ph.D. Thesis, Harvard University, 1986.

Dance, Daren and Karen Gildersleeve, "Estimating Semiconductor Yield from Equipment Particle Measurements" Proceedings of 1992 IEEE International Semiconductor Manufacturing Science Symposium, San Francisco, CA.

Dance, Daren and Richard Jarvis, "Using Yield Models to Accelerate Learning Curve Progress", *IEEE Transactions on Semiconductor Manufacturing*, Vol. 5, No. 1, February 1992.

Flaherty, M. Therese "Managing Manufacturing and Engineering in VLSI Fabs", Proceedings of 1990 IEEE International Semiconductor Manufacturing Science Symposium, Burlingame, CA .

Flaherty, M. Therese "The Management Factor", 34th Joint National Meeting, ORSA/TIMS, San Francisco, CA 1-4 November 1992.

_____, Harvard Business School, "An Informal Note on Knowledge and How to Manage It", HBS Case 9-686-132, 1988.

Hayes, Robert, Steven Wheelwright and Kim Clark, *Dynamic Manufacturing: Creating the Learning Organization*, The Free Press, 1988.

Hodges, David "VLSI Management in Japan and the United States" Proceedings of 1990 IEE Advanced Semiconductor Manufacturing Conference, Danvers, MA.

Hof, Robert, "Inside Intel", *Business Week*, 1 June 1992, p. 86.

Kim, Daniel and Gary Burchill, "Systems Archetypes as a Diagnostic Tool: A Field-based Study of TQM Implementation", Working Paper, 1992.

Kim, Daniel, "Guidelines for Drawing Causal Loop Diagrams", The Systems Thinker, Vol.3, No. 1, February 1992.

Leonard-Barton, Dorothy "Implementation as Mutual Adaptation of Technology and Organization", *Research Policy*, Vol. 17, 1988.

Milling, Peter, "Time- A Key Factor in Corporate Strategy", Proceedings of the 1990 International system Dynamics Conference, *System Dynamics '90*, edited by Anderson, D., Richardson, G., and Sterman, J.

Morecroft, John D.W., "System Dynamics and Microworlds for Policymakers", European Journal of operational research, Vol. 35, 1988.

Reagan, Mark, "Turning R&D into Productive Manufacturing Knowledge - A Tour Through Pilot Lines in Japan", *Dataquest Newsletter* 0007746, 1990.

Reagan, Mark, "A Comparison of Japanese and US Fabs", *Dataquest Newsletter* 0005328, 1989.

Senge, Peter, "The Leader's New Work: Building Learning Organizations", *Sloan Management Review*, Vol. 32, No. 1, Fall 1990.

Senge, Peter, *The Fifth Discipline: The Art & Practice of the Learning Organization*, 1990 Doubleday.

Stata, Ray, "Organizational Learning -- The Key to Management Innovation", *Sloan Management Review*, Vol. 30, No. 3, Spring 1989.

Strauss, Anselm, (1987) *Qualitative Analysis for Social Scientists*, Cambridge University Press.

Tyre, Marcie, "Managing Innovation on the Factory Floor", Technology Review, October 1991.

Tyre, Marcie and Eric von Hippel, "Trial and Error Learning with New Technologies: Where it Occurs, and Why that Matters", Draft, March 1993.

von Hippel, Eric and Tyre, Marcie, "How 'Learning by Doing' is Done: Problem Identification in Novel Process Equipment", MIT Sloan School of Management Working Paper #BPS 3521-93, January 1993.

von Hippel, Eric "The Impact of 'Sticky' Information on Innovation and Problem Solving", Working Paper 3147-90-BPS, MIT, April 1990.

Wein, Lawrence M., "On the Relationship Between Yield and Cycle Time in Semiconductor Wafer Fabrication", IEEE Transactions on Semiconductor Manufacturing, vol. 5, no. 2, May 1992.

Appendix A: Definitions

Die Yield - Percent good die per wafer. Usually treated as a function of defect density.

Excursion - A significant drop in yield in a lot or a number of lots. Often attributed to equipment failure or misprocessing.

Line Yield - Percent of wafers which are processed through the whole line. Usually treated as a function of wafer handling, equipment failure, etc.

Wafer starts - Number of wafers started into production in a defined time period.

Wafers sorted - Number of wafers which are selected to proceed to die level testing.

Appendix B: Guide to Causal Loop Diagrams

[This section is extracted from "Systems Thinking Approaches to Manufacturing" (Benfer, 1993)]

People are often prisoners of their own thinking. The mental models that they hold of their organization and environment limit their abilities to identify new opportunities for high leverage change. System thinking helps to overcome this limitation by providing tools that allow organizations to develop a more holistic outlook. Causal loop diagrams are one such tool.

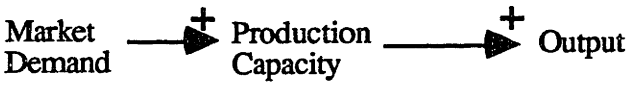


Figure B1 Linear mapping of production capacity

Linear thinking pervades most organizations. Figure B1 illustrates a linear view of the relationships between market demand, production capacity and output. A mental model such as this one encourages a reactive response to capacity planning. A more thoughful model would incorporate important feedback and delays. The mental model mapped in Figure B2 fosters a more proactive policy towards capacity planning, since it explicitly recognizes the delay to install capacity and the important influence of output availability on market demand.

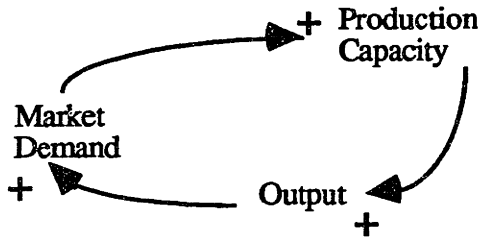


Figure B2 Feedback mapping of production capacity.

Causal loop diagrams help illustrate key relationships in an easily understood graphical format. The diagrams capture the system structure -- delays and interdependencies -- that drive system behavior. Many of the reoccurring structures have been identified and catalogued as *system archetypes*. These generic templates can be readily modified for specific systems under study and thus increase awareness of dynamic behavior of the system.

The arrows in causal loop diagram depict interactions between two variables. In Figure B1, the interaction between product capacity and output was signified by an arrow connector. The sign by the head of the arrow denotes the polarity of the relationship. The positive sign reflects the fact that as production capacity rises, output rise (all other factors remaining constant). See Table B1 for further explanation.

The basic building blocks of causal loops are readily grasped. These are reinforcing loops (positive feedback), balancing (negative feedback) and delays. Figure B3a illustrates both types of loops. Reinforcing loops amplify dynamic behavior: the familiar snowball effect. The effect can be either positive or negative, as shown in the Figure B3b. Reinforcing loops reach their limit at some point; eventually the system reaches a limiting condition as part of a large balancing loop.

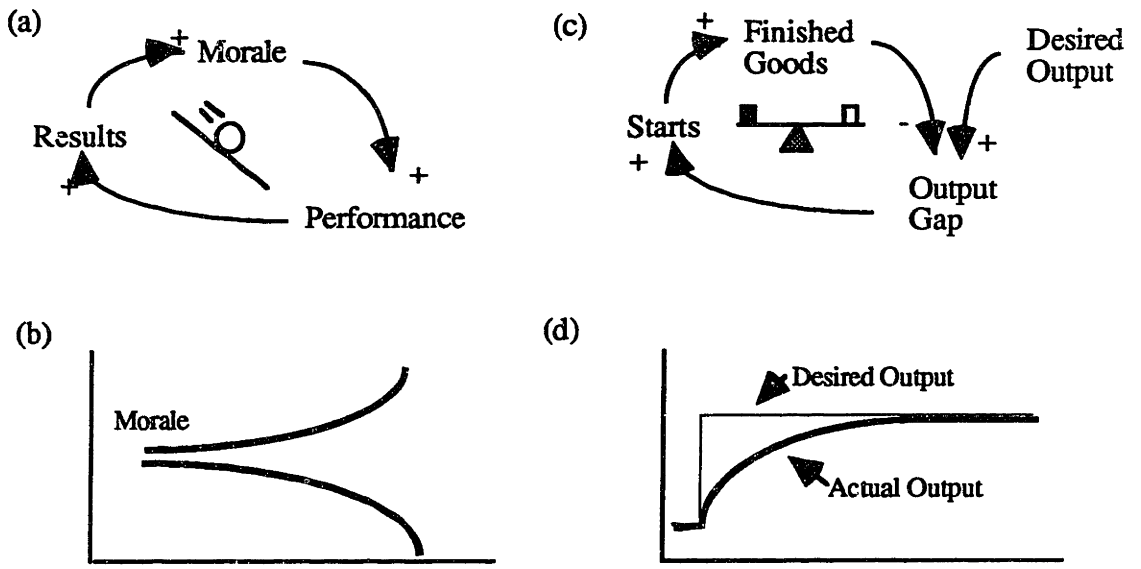


Figure F3 Basic Feedback Structures (a) Reinforcing feedback loop. (b) Dynamic behavior in either direction. (c) Balancing feedback loop. (d) Dynamic behavior.

Balancing loops are goal seeking, much like a thermostat. Figure B3c illustrates an example of a balancing loop from a manufacturing environment. The dynamic behavior of this system to a disturbance (in this case, a step increase in desired output) is shown in Figure B3d. Organizations that are unaware of balancing loops in their systems will often push the system harder to gain increased performance. Unfortunately, the harder the organizations pushes, the harder the system will push back. Real leverage can be found in removing or enhancing the limiting condition. Awareness of balancing loops is often obscured in practice, since these structures do not exhibit the dramatic dynamics seen in the reinforcing system. Without long term trend information, one may not recognize significant forces acting upon the system.

Recognizing and reducing delays in the system can also be a high leverage strategy to greatly enhance system performance. Unknown delays can often lead to instability

(oscillating behavior), especially when the delays are long and thus the relationships between cause and effect are obscured. Figure B4 illustrates this behavior. The observed overshoots often arise due to the correcting action. Ironically, the more forceful the correcting action, the longer it will take for the system to stabilize. Assertive action can, therefore, lead to higher levels of instability.

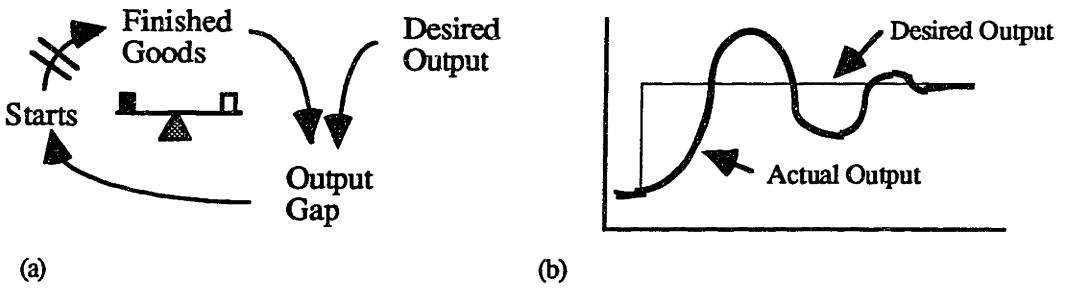
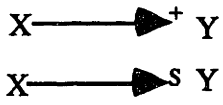
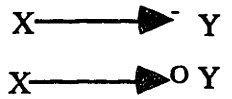
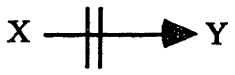
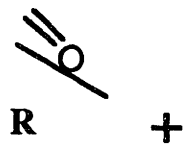


Figure B4 Effects of system delays: (a) Diagram including a delay between starts and finished goods. (b) System performance shows oscillating dynamic behavior.

Table B1 Description of Symbols used in Causal Loop Diagrams.

Symbol	Interpretation
	X and Y tend to move in the same direction, when all else is equal.
	X and Y tend to move in opposite directions, when all else is equal.
	Time delay between X and Y.
	Reinforcing loop (positive feedback). The number of negative relationships counted around the loop is even or zero.
	Balancing loop (negative feedback). The number of negative relationships counted around the loop is an odd number.

Appendix C: Statistical Analysis of Noise

Bootstrapping Methodology

In order to evaluate the consequences of noise on learning by experimentation without first assuming a known distribution of die yields, I used the bootstrapping method outlined by Bohn (1991). Bootstrapping uses Monte Carlo sampling to construct statistical power functions. This methodology simulates experimental results for hypothetical split lot experiments on the batches from Factory A. The wafer level die yields from a single lot were sampled with replacement to construct two "split batches" of N wafers, which represent the results of a single experiment. No correction was made for line yield losses.

A test statistic ΔY_{exp} (the difference in the observed average *ln*[yield])was then calculated assuming that $\Delta Y_{true} = 0$ (in other words, the hypothetical experiment had no actual effect). The sampling process was repeated 500 times for each time period, with sampling conducted equally from each batch of a particular time period. This resulted in 500 values of ΔYest; this was doubled to 1000 through symmetry. The 1000 simulated experiments were then used to estimate the error rates of real experiments from each time period (A1, A2 and A3).

Generating Power Functions

These results are used to construct a power function G(ΔY) for simulated experiments. The power function gives the probability that change which produces a ΔYtrue will be accepted. We will model a simple test criterion: if $\Delta Y_{exp} > 0$, then accept the change; if

not, reject the change. An ideal power function would resemble Figure C.1, with the curve rising steeply through $\Delta Y=0$ so that for $\Delta Y>0$ the change will be accepted and for $\Delta Y<0$ the change will be rejected.

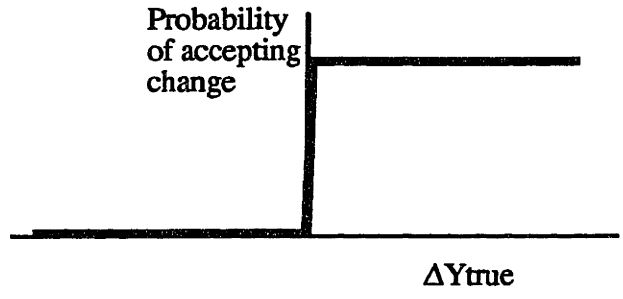


Figure C.1 Ideal power function G(ΔY) for $\Delta Y_{est} > 0$ decision rule.

However, experimental noise insures that the curve is not sharp and instead resembles Curve A shown in Figure C.2. If the decision rule is made more stringent, such as $\Delta Y_{exp} > \Delta Y_{cutoff}$, the power function curve shifts to the right by ΔY_{cutoff}. This is shown in Figure C.2 as Curve B.

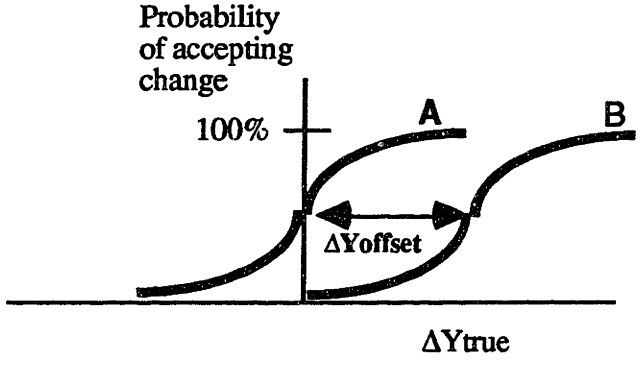


Figure C.2 Example of actual power functions.

Appendix D: Yield Models

Most yield models relate integrated circuit yield using device area and defect density estimates. Since defect density is a random phenomenon, yields (which are a function of defect density) will also be random variable. There are multiple sources of variation, including:

- Defects vary in size.
- Defects are distributed randomly on the wafer.
- Sensitivity of different mask layers to defect size and location varies.
- Sensitivity of circuits to defect size and location varies. (Dance, 1992)

Several yield models are discussed in the literature. The negative binomial is one such example:

$$Y = 1/(1 + AD_oP_f/C)^C$$

where:

Y = Percentage defect limited yield
A = Device area
P_f = Fault Probability
D_o= Defect Density
C = Degree of clustering

However, if C equals ∞, then the formula becomes the Poisson formula:

$$Y = \exp(-AD_o)$$

If C equals one then formula reduces to Seeds formula:

$$Y = 1/(1+AD_O)$$

Appendix E: Half-life Method of Analysis

The half-life method of analysis can be a useful approach to managing quality improvement efforts. Quantities that are targeted for reduction, such as defect levels or failure rates are tracked. They are then plotted against time or cumulative production on a semi-log scale. Figure E1 illustrates an example:

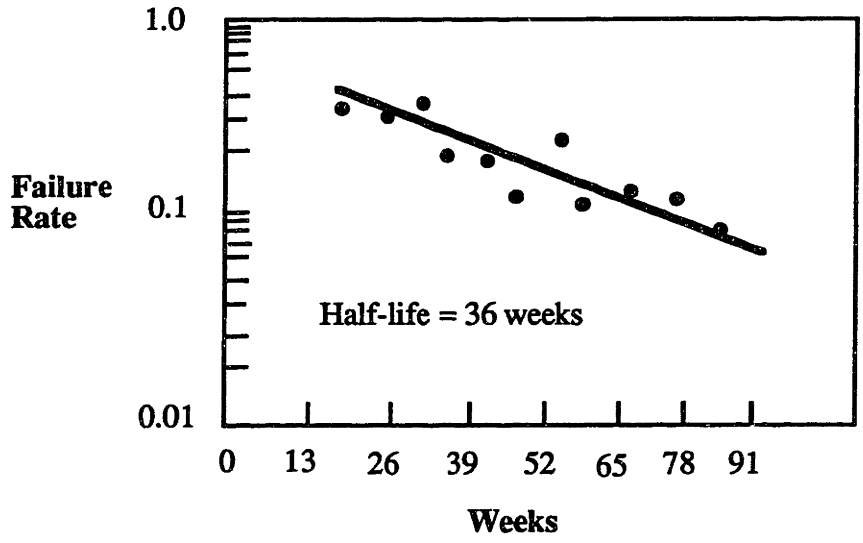


Figure E1 Example of half-life learning curve analysis.

A linear regression determines the best straight-line fit. From this equation a calculation of the *half-life*, or the time it take for 50% reduction in failure rate, can be made.

The half-life for any improvement process is influenced by the rate of organizational learning. The slope of the learning curve is determined by the time it takes to identify and prioritize the causes of a problem and to eliminate those causes. Using pareto

analysis, the top few causes are addressed; the remaining causes are attacked in the next cycle. Influences on the cycle time of this process include amount of dedicated resources, skills of the people and the *complexity and bureaucracy of the organization* (Stata, 1989).

Utilizing the Half Life Concept

Organizations can use the half-life technique to help their quality improvement in many ways:

Setting goals

This analysis enables rational goal setting since the process recognizes the increasing difficulty to obtain improvements as defects are systematically eliminated. Six to twelve month half-lives translate into substantial progress when extended over a reasonable time horizon.

Communicating results

Part of the effectiveness of this approach lies with graphical format. Trends in performance are easily recognized and can be displayed in a compact summary. Employees can readily see their movement towards long term goals.

Motivating Learning and Knowledge Transfer

Readily exposes areas of slow improvement so that they can be targeted for focused attention. Allows for quick comparisons of performance across various facilities with the firm. Outliers (both fast and slow learners) and opportunities for knowledge transfer are identified. Figure E2 illustrates how differences among various sites can be highlighted using this method.

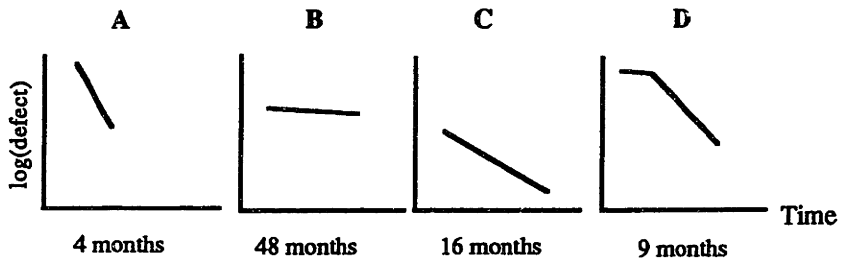


Figure E2. Comparison of defect reduction at four factories. Improvement rates are easily compared. Significant differences are highlighted and can be targeted for further investigation.