Sub-Micron Gate Lithography in an MMIC Production Environment

by

PETER WILLIAM MORAN

B.S., Mechanical Engineering
M.S., Mechanical Engineering
Stanford University, 1985

Submitted to the Sloan School of Management
and to the Department of Electrical Engineering
in Partial Fulfillment of the Requirements of the Degrees of

Master of Science in Management
and
Master of Science in Electrical Engineering

at the
Massachusetts Institute of Technology
June, 1990

Signature of Author ______________________
MIT Sloan School of Management and Department of Electrical Engineering
May 29, 1990

Certified by ______________________
Rebecca M. Henderson, Assistant Professor of Strategic Management

Certified by ______________________
Jesús A. del Alamo, Assistant Professor of Electrical Engineering

Accepted by ______________________
Jeffrey ... Associate Dean, Master's and Bachelor's Programs

Accepted by ______________________
Arthur C. Smith, Chair, Department Committee on Graduate Studies

Sub-Micron Gate Lithography
in an MMIC Production Environment

by

Peter William Moran

Submitted to the MIT Sloan School of Management
and to the MIT Department of Electrical Engineering
on June 15, 1990, in partial fulfillment
of the requirements of the Degrees of
Master of Science in Management
and Master of Science in Electrical Engineering

ABSTRACT

Well controlled processes have been shown to be economically and strategically beneficial to the long term competitive abilities of a firm, yet many firms have consistently under invested in activities to improve process control or failed to apply those investments in an optimal manner. This thesis examines the process for allocating and applying resources to activities that improve process control in one specific industrial setting - a captive supplier of leading edge, solid state electronic devices and circuits. I show that this organization encounters many of the difficulties suggested in the literature on efficient resource allocation. In particular, following a recent shock to the business, the organization does not measure certain critical information and incorporate it into the decision process. A key piece of such information for this yield limited, 0.42 µm MMIC fabrication process is shown to be the location and magnitude of variability in the process. This variability can be measured through a detailed study of the device and process physics and incorporated in the resource allocation process. Variation in one key circuit performance parameter, gain slope, is shown to be caused, primarily, by variation in date length which in turn is affected by the method used for targeting develop time, by developer exhaustion due to loading effects, by variation in developer bath temperature, by the presence of particles on the wafer during contact gate lithography, by variation in mask CD accuracy and by poor repeatability in intra-mask CD.

Thesis Advisors:
Rebecca Henderson, Assistant Professor of Strategic Management
Jesús del Alamo, Assistant Professor of Electrical Engineering

Acknowledgements

This work was conducted under the auspices and funding of the Leaders For Manufacturing Program at the Massachusetts Institute of Technology. This program, begun in April of 1988, is a Partnership between the School of Engineering, the Sloan School of Management and eleven major US manufacturing firms, among them Hewlett-Packard. The partnership nature of the program seeks to foster joint academia/industry research such that the empirical basis of the research will foster more rapid utilization in industry. Additionally, the combinations of Management and Technical disciplines seeks to address those problems in Manufacturing that require a joint vantage point for solution.

I am indebted to the people of MWTD for openly sharing their corporate and personal lives with me. In particular, without the many hours invested by Scott Elliott and Neil Wylie, this learning experience could not have been so rich. Their experiences were richly complemented with the frameworks and analytical methods contributed by my faculty advisors at MIT, Jesús del Alamo and Rebecca Henderson.

The ideas in this thesis and much of the enjoyable time spent at HP in Santa Rosa and at MIT would not have been possible without the challenges and rewards of working with my friends can colleagues in Manufacturing, Jim Lawton and Dean Vlasak. Finally, without the unflagging support of my wife, Anne....I hate to think.

Table of Contents

List of Figures

ABSTRACT

Well controlled processes have been shown to be economically and strategically beneficial to the long term competitive abilities of a firm, yet many firms have consistently under invested in activities to improve process control or failed to apply those investments in an optimal manner. This thesis examines the process for allocating and applying resources to activities that improve process control in one specific industrial setting - a captive supplier of leading edge, solid state electronic devices and circuits. I show that this organization encounters many of the difficulties suggested in the literature on efficient resource allocation. In particular, following a recent shock to the business, the organization does not measure certain critical information and incorporate it into the decision process. A key piece of such information for this yield limited, 0.42 μm MMIC fabrication process is shown to be the location and magnitude of variability in the process. This variability can be measured through a detailed study of the device and process physics and incorporated in the resource allocation process. Variation in one key circuit performance parameter, gain slope, is shown to be caused, primarily, by variation in date length which in turn is affected by the method used for targeting develop time, by developer exhaustion due to loading effects, by variation in developer bath temperature, by the presence of particles on the wafer during contact gate lithography and by poor repeatability in intra-mask CD.

Chapter 1: An Introduction to Resource Allocation in the Manufacturing Environment

Well controlled processes have been shown to be economically and strategically beneficial to the long term competitive abilities of a firm, yet many firms - particularly American ones - have consistently under invested in activities to improve process control or failed to apply those investments in an optimal manner. *Made In America* has suggested that this lack of investment has been a fundamental contributor to the downfall of several U. S. industries, among them: machine tools, VCRs and other consumer electronics, and, most recently, DRAMs and other commodity semiconductors.[1]

In *Dynamic Manufacturing*, Hayes, Wheelwright and Clark argue that this pattern of under investment can be traced directly to the system for allocating scarce resources.[2] It follows that if the resource allocation system could be improved, the observed pattern of investments would more nearly match the "true" needs of the business. In many US industries, this would change the investment bias towards process improvements thus helping to restore competitiveness.

Clearly, the efficient allocation of resources is one of the central challenges of management. It is here that managers exercise one of their biggest available levers for improving shareholder return, yet this lever is a difficult one to master. An efficient allocation system goes well beyond the simple concept of net present value taught as a prime decision criteria in modern Finance courses, for several reasons. These reasons can be thought of as falling into two categories: organizational problems and technical problems.

Organizational

1) Corporations and individuals are slow to respond to change. In particular, after a major environmental shock, organizations often ignore variables critical to efficient resource allocation in the new environment.
2) Strategy formulation, budgeting and allocation processes occur within a hierarchically and geographically diffuse organization.
3) The means for generating resource allocation options may ignore several very important options.

Technical

4) Multiple criteria used to select between projects can not be easily summarized in a single measure such as profit.
5) All allocation options inherently involve uncertainty in terms of technical, economic, and/or commercial success.
6) Time dispersed arrival of options requires a continuous decision process.
7) Existing decision models do not deal well with the ambiguity created by the problems above.

Because of these problems, the literature on resource allocation suggests that it is doubtful that any large organization is truly efficient in its allocation of resources; nonetheless, improving the current level of efficiency remains a goal well worth striving for.

While much research has focused on resource allocation in terms of investigating from a broadly based viewpoint the difficulty of allocating resources efficiently, comparatively little empirical work has been done to validate or investigate, in detail, the truth of the supposed causes of this difficulty and the situational and structural barriers that inhibit overcoming the difficulty. I begin by looking in depth at the allocation of resources, to improve process control, in the manufacturing function of one division of a major U.S. company, namely the Microwave Technology Division (MWTD) of the Hewlett-Packard Company (HP), to examine whether the methods used and difficulties encountered result in a sub-optimum pattern of decisions. Indeed, MWTD appears to suffer from many of the same problems outlined above. In particular, having recently undergone a substantive change in their business, there is evidence that they are not paying attention to variables that are critical in the new business environment of their division. I show only that <u>some</u> critical variables are ignored. Many, if not most, important variables appear to be actively observed and incorporated into management decisions at MWTD. Thus, the emphasis is on seeking to improve what are already generally efficient management methods.

One such group of critical variables that appears to not be receiving sufficient attention surrounds investment in process control. The current pattern of allocation towards process control seems to be one factor that leads to two distinct risks in MWTD's ability to compete:

- Batch to batch yield is not predictable. Therefore, MWTD cannot quickly respond to changes in demand, nor reliably meet customer ship dates without carrying considerable finished goods inventory.

- Efforts to increase yield are unfocused. Managers cannot predictably increase yield through the assignment of resources. This jeopardizes achievement of the fourfold increase in throughput that is a key goal in this division's strategic plan.

MWTD could also realize substantial benefits from increased investment in process improvement: cost per part would decrease by spreading high fixed costs over more "good" circuits, engineering line support and management cost to respond to unexpected problems would decrease, and the cost of improving the process - either within manufacturing or for development of new processes - would decrease.

One of the underlying causes for this situation is that accurate cost/benefit information is not available to the decision maker in the right form, at the right time and with a sufficient degree of certainty, and then incorporated by him or her into the decision process. While this problem has been recognized repeatedly in the past, little empirical work has been done to evaluate precisely what information is being ignored in the decision; and, how and at what cost the information could be made available.

Towards beginning to fill this gap, I continue by showing that an analytical decision could be aided by the availability of three pieces of information:

- the additional profit from a change in yield,
- the yield improvement per process improvement project, and
- the cost of implementation for each such project.

One of these three, the yield improvement per process improvement project, requires identifying where and how variability occurs in each part of the manufacturing process. I present a method for identifying this variability.

each part of the manufacturing process. I present a method for identifying this variability.

Because the sources of variability are spread throughout a multi-step process and because the propagation and effects of this variability are ingrained in the physics of the manufacturing process and the circuit being produced, identifying the variability requires a detailed, technical look at the manufacturing process and an understanding of the process and device physics involved. Implicitly, I suggest that the critical variables of concern in the resource allocation decision are heavily dependent on the technology involved. Thus, understanding what and why information is ignored cannot be separated from understanding the technology.

The MMIC-A process studied here produces a variety of high frequency, integrated circuits. By physics and empirical data correlation, I show that the key physical determinant of the electrical performance (in particular, gain slope) of a finished 2-26.5 GHz traveling wave amplifier made on this process is the transistor gate length; hence, control of this feature will, to a large degree, provide control of circuit performance. The aforementioned variability study shows that variability in six key process inputs are the dominate causes of variability in gate length:

- presence of particles on the wafer during contact gate lithography,
- repeatability of the intra-mask critical dimensions (CDs),
- accuracy of mean mask CD,
- developer exhaustion due to loading effects,
- developer bath temperature, and

- method for targeting develop time.

Identifying the variability and understanding the yield and cost impact of this variability, allows for a conversion to yield loss per process input and estimation of the dollar impact to MWTD for improving this yield.

This explicit estimation of costs and causes of process variability provides one of the critical pieces of information that is currently being ignored by MWTD. However, even by incorporating this and other critical variables, such as the costs of implementing these process improvement projects, it is not clear that the pattern of allocations at MWTD would change. There are other structural impediments in the resource allocation system, such as internal distribution of manpower, that make incorporating critical variables a necessary, but not sufficient condition for improving resource allocation.

Finally, the narrow industrial base and type of allocation option chosen may limit the broad applicability of any of the conclusions of this thesis, but every effort is made to explicitly note the relevant situational factors, so that the reader may better interpret the results for as they relate to other situations.

<u>Thesis Organization</u>

In Chapter Two, I suggest a simple conceptual model from which to explore resource allocation. I then review the three streams of the management literature on which this research builds. I present general theory on controlling processes and suggest how this is best applied at MWTD to increase yield. Finally, I show that variability reduction is the key goal for controlling the process in this industrial environment .

In Chapter Three, I describe the organizational situation that provided the basis for this study: the site, its history, its markets, and its technology. I then consider the structures present at MWTD: the organization, the incentive system, and the means by which they formulate a strategy and communicate it to the organization. Finally, I present the current resource allocation methodologies in use at MWTD today.

In Chapter Four, I discuss the implications of MWTD's current allocation methodologies. Reviewing first how many of their problems seems to be quite similar to those reported in the literature, I show that the current pattern of allocations is very costly to MWTD. In looking towards improvements over current allocation methodologies, I begin to explore in greater depth one of the reasons for under investment in process control: not paying enough attention to all of the relevant variables. I suggest that MWTD does not fully understand the causes of yield loss and, hence, cannot make necessarily well aimed efforts to improve this yield.

In Chapter Five, I present a framework derived from Quality Function Deployment (QFD) theory as a means to link variability in meeting customer desires to variability in some finite set of process inputs. Then, I apply this QFD derived framework to the 2-26.5 GHz traveling wave amplifier made on the MMIC-A process by correlating existing product data. Having defined a relationship between a key customer desired parameter, gain slope, and a suitable in-process control measure, transfer layer opening, I examine variability in 22 process inputs to the three process steps that contribute most to the variability in the transfer layer opening. I summarize this analysis by showing how variability in these inputs translates to variability in gain slope and show that six process inputs currently dominate in causing the observed variability in the finished circuit. Because of the very high noise present in this environment, I discuss the significant error estimates that come with the experimental results. Finally, I discuss the management implications of the results presented on the MWTD allocation process.

Finally, in Chapter Six, I conclude with a review of the results found in this research. I leave off by showing how this work reflects on the current directions of the literature on resource allocation and suggest possible future extensions of this work.

Chapter 2: Prior work and Hypothesis

In explaining American industries declining competitiveness in world industries, *Made In America* offers several examples of how Japanese companies dramatically out invest their American competitors on process improvement. While US firms invest 2/3 of their research budget in new product development and 1/3 in process development, Japanese firms reverse this ratio. Furthermore, the Japanese have twice the number of process engineers as the US. Clearly, there seems to be a far stronger bias towards investing in process improvement in Japan and it correlates quite well with enhanced competitiveness in world markets.[3] Why then do American firms not mimic this behavior?

Endeavoring to better understand this area, researchers have begun to explore how resource allocations are currently made in practice. From a comparatively limited set of data, one can reach the broad conclusion that current methods do not result in an efficient allocation of resources. The very real situational and structural impediments that lead to such sub-optimization have been well documented on at least a qualitative basis.[4] Finally, much work has been done to improve the actual allocation decision rule by developing or applying more sophisticated mathematical models from decision theory. Unfortunately, many of these models have met with only limited practical success because they largely ignore the situational and structural impediments.

I begin this chapter by discussing a simple conceptual model of the resource allocation process and the constraints involved. While not descriptive of actual practice, this model serves to highlight the key elements necessary for efficient allocation of resources. Then I take a closer look at what other research in the field has revealed about how allocation decisions are currently being made in business and the situational and structural impediments that make efficient allocation difficult.

The literature suggests that one of the prime causes of sub-optimal allocation is the failure of an organization to pay attention to critical variables. This is suggested to be especially true after a major environmental shock. To understand if this is, in fact, the case, I examine one class of allocation decisions, namely those that improve process control. One can only understand the effects of these decision patterns after considering the driving factors of the underlying process technology, namely gallium arsenide (GaAs) semiconductor manufacturing. Here, I show that yield is the best measure of process control in this situation and that changes in yield provide a multitude of benefits to the company, but come only at considerable cost. The pattern of allocations at MWTD in the past has shown a bias away from process control projects. I argue that this decision might be made differently, closer to an overall optimum, if more information on critical variables involved were available to and used by the decision maker.

Finally, to provide a foundation for the considerable focus on process control projects that begins in Chapter Five, I review the basic theory on controlling processes and show that one key piece of needed information is the identification of the causes and magnitude of variability.

2.1 A Simple Conceptual Model of Resource Allocation

The basic challenge of resource allocation is to allocate the resources of the company, primarily manpower and capital, to some collection or portfolio of projects that will bring in the highest possible return to the company. The corporate goal to maximize shareholder value suggests that these allocations should be made to that combination of projects that provide the greatest overall return in terms of profit per resource dollar invested.

(Note that manpower and capital differ somewhat in that one is expensed while the other is depreciated. While most of the literature has focused on the allocation of capital, I broaden the scope to examine both manpower and capital - through the common link of allocation of scarce resources - by noting that both involve cash outlays up front on the part of the firm, even if tax and financial accounting handle the two differently.)

In the simplest perception of how one would efficiently allocate resources, one can conceive of a single corporate "Resource Allocator". To this person or system, flow some large number of options on how to invest the resources of the corporation. With each of these options, come a set of known costs, benefits and risks. From the resource allocator come some subset of these options into which the corporation will invest its resources. Available to the allocator will be a "free market" of resources. This market will include not only the corporation's current assets and staff, but also access to capital markets and an employee base currently external to the corporation. Finally, the allocator would have open to him or her, all knowledge of the environmental factors of the corporation, information such as corporate strategy, capabilities of individual workers, and status of investment projects currently under way. To make his or her decision, the Resource Allocator might use a decision rule such as "Accept all projects with a positive net present value". Conceptually, this model might appear as in Figure 2.1.

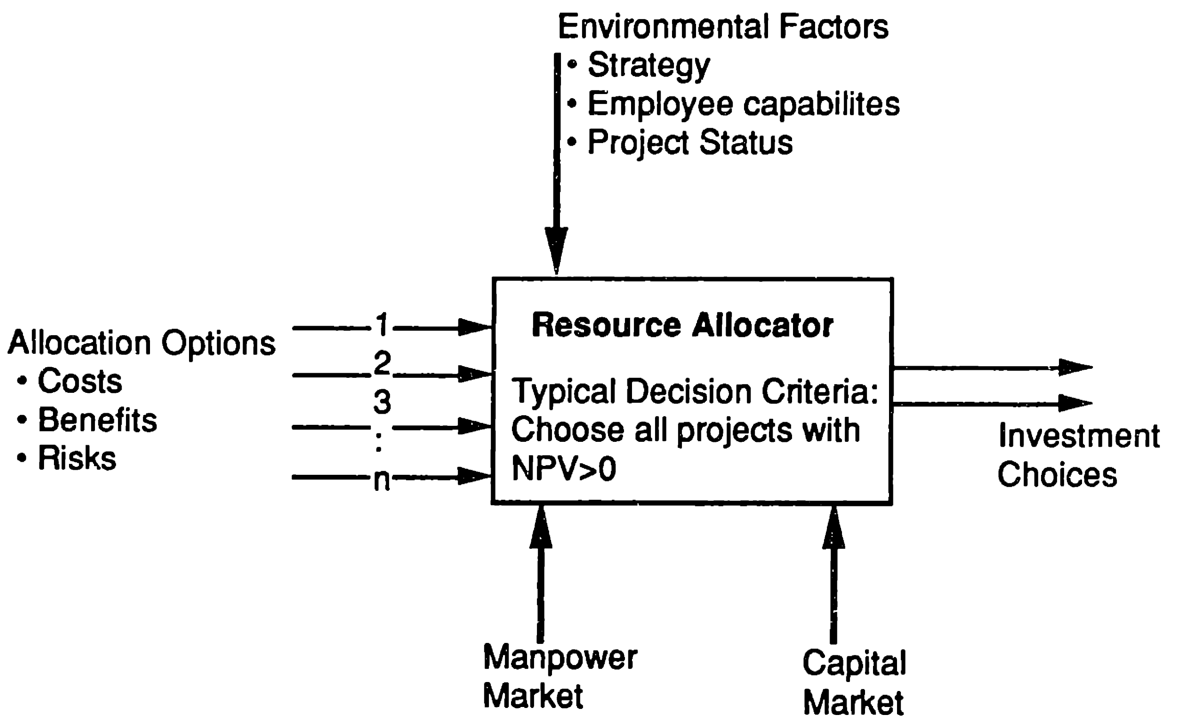


Figure 2.1: Conceptual Model for Resource Allocation

While this model of resource allocations is overly simplistic, it readily suggests many of the primary challenges involved in making efficient allocation of resources:

1) Generating opportunities to invest
2) Understanding of the costs, benefits and risks of each of these options. Expressing the costs and benefits in comparable terms.
3) Being able to freely acquire manpower and capital.

4) Having one "resource allocator" consider all of the relevant environmental information and all of possible allocation options.
5) Dealing with a continuous arrival of projects over time and with interdependence between projects
6) Calculating a meaningful discount rate

Given these challenges, I now consider how resource allocation decisions have tended to be made in practice.

2.2 Review of the Literature on Resource Allocation

There are three primary streams in the literature from which this work follows.

By thinking of resource allocations as investment in the firm's future, academicians and practitioners long ago opened the door to applying Finance theory as an aid in making the decision between which allocations should or should not be made.[5] However, these tools are often too limited to deal with making these decisions within the organizational context, so Management scientists have focused on developing more complex and robust decision tools and elaborate models that can better compensate for the implications of using this model in a non-rational world.[6,7,8,9,10]. Many of these models focus on the construction of a risk-balanced portfolio of projects that maximize return. Various models go to great lengths to deal with particular imperfections, such as time dispersed arrival of project options, or to develop the use of objective or constraint functions as a means of maximizing portfolio return. However, in practice, these complex, mathematical decision models are rarely used. Instead, current practice relies on far simpler techniques such as checklists and project profiles.[11,12,13].

Accepting that allocation decisions are made in a complex, behavioral context, a second stream of research pioneered by Joseph Bower has shown a model analogous to the conceptual one presented earlier, that considers how decisions are made in practice. Bower characterizes the resource allocation process as a multi-attribute choice between options generated at the bottom of the organization in response to goals and strategies imperfectly passed down the organizational ladder. Furthermore, the decision process has been characterized as a search for more and better information on which to base the final decision.[14]

This organizational perspective has been furthered by Quinn in a third and final line of research. Of particular note, he has suggested that after a major shock to the environment in which a business operates, the organization is slow to adapt to the new environment. Organizations in this situation often fail to pay attention to newly critical variables.[15,16]

Taken in summary, the literature has found the resource allocation process to be fraught with difficulty on many fronts:

1) Organizations ignore critical variables after an environmental shock.
2) The strategy formulation, budgeting and allocation processes occur within hierarchically and geographically diffuse organizations.[17,18]
3) Option generation may ignore important options.

4) The multiple criteria used to select between projects can not be summarized in a single measure such as profit.[19,20]
5) All allocation options inherently involve uncertainty in terms of technical, and/or commercial success.[21]
6) Time dispersed arrival of options requires a continuous decision process.[22]
7) Actual decision models do not deal well with the above problems.[23,24]

The literature suggests that most major companies are faced with some or all of the problems outlined above in their attempts to allocate resources efficiently and, although little research has been done to formally verify this, the implied conclusion is that corporate responses typically result in sub-optimum allocation patterns.

Having reviewed current thinking on the general problem of resource allocation, I now provide more background on one of the common problems above "organizations fail to pay attention to critical variables after a major environmental shock." To consider this problem in more depth than has been typical in the literature, I focus on a single class of allocation options: projects that improve process control. I begin by reviewing the forces that drive control of semiconductor manufacturing processes.

2.3 Resource Allocation for Semiconductor Process Control

To fully understand the effects of resource allocations on process control problems, it is necessary to first frame how process control projects differ from other allocation options in the manufacture of semiconductors. I approach this by considering the metrics, benefits, cost and decision criteria that apply to process control projects and separate such projects from other manufacturing investment options.

2.3.1 Measures of Semiconductor Manufacturing Performance

The benefits available from process control projects are somewhat unique to the manufacture of semiconductors. Unlike many manufacturing processes, semiconductor fabrication is often characterized by an extremely low percentage of good circuits when a new process is first introduced. This is particularly true when the semiconducting substrate is gallium arsenide. Furthermore, the production process tends to be composed of several discrete technology processes and may require well over one hundred steps in the manufacture of a single circuit. These characteristics usually make mean yield, the number of good die divided by the number of die started, the key measure of a single process step or the overall process. It is not uncommon for the lot-to-lot (a lot is one production batch of wafers, typically 5-25 wafers), wafer-to-wafer or die-to-die yields to vary widely, (standard deviation of the same order as the mean). Consequently, the standard deviation in yield can be an equal or even more important measure than mean yield. Thus, the goal of most process control projects is to improve mean process yield or, at least, reduce the variance in the yield.

2.3.2 Benefits of Improving Process Control

The benefits of improving mean yield or decreasing yield variability are threefold. First, increases in mean yield are beneficial in terms of reducing costs because more products are produced with little increase in cost. Second, decreasing yield variability provides a more predictable manufacturing process. As has been shown by Hayes, Wheelwright and Clark[25], predictability offers several advantages: reduced finished goods inventory, improved flexibility to respond to changes in demand, and improved flexibility to respond to introduction of new products. It has also been observed that increased predictability reduces the cost of engineering and management support of line "fire-fighting" (responding to unexpected or out of control situations on the production line). Third, as Bohn has shown, the increase in yield and reduction in yield variability improves the speed at which the organization can learn. He extrapolates that in the long run, the speed of learning will drive down overall manufacturing costs.[26]

While the reduction in costs can be turned into a hard dollars measure with a sufficient set of assumptions about current operating conditions, the increase in predictability and faster learning are less easily converted to a dollar measure though most would agree that they add at least some value to the organization. Such benefits must be 'measured', at least qualitatively, in light of how they fit with the enterprise's strategy. In Section 4.4, I further consider all of these benefits.

2.3.3 Costs of Improving Process Control

Given this understanding of the benefits possible from process control projects, I now consider the costs involved. The primary structure of costs involves three resources: capital, manpower, and production line time. Capital is used to purchase new or better equipment that offers greater accuracy or repeatability, if not capacity. (Note that capacity increases, other than by increasing yield, largely fall into a different class of resource allocation options and, hence, are not considered here). Manpower, especially engineering time, is used to run experiments to determine better operating conditions or procedures or to select and install new equipment. The highly variable nature of the semiconductor manufacturing process suggests that most experiments should be run on actual production line equipment, as opposed to distinct R&D or test equipment. Production line time is largely free when a fab (semiconductor fabrication line) is not capacity constrained, but as bottlenecks develop, such time can represent "lost production opportunity".[27] Thus capital, manpower and production line time are the three resources most needed by process control projects. Manpower and production line time might also be needed to identify which process control projects are the most valuable to work on (as is done in this thesis).

2.3.4 Decision Criteria

To allow comparison with other alternatives in manufacturing and elsewhere in the company, one - ideally - frames a process control project in terms of financial return; where return equals additional profit per invested resource dollar. Breaking this down, the return is composed of two terms: the additional profit from a given change in yield and the resources invested to create this change in yield. (Note that I am using the term 'change in yield' loosely here. By this I mean either increases in mean yield or decreases in the variability of this yield over time).

$$\text{Return} = \frac{d\,\text{Profit}}{d\,\text{Investment}} = \frac{\partial\,\text{Profit}}{\partial\,\text{Yield}} \times \frac{\partial\,\text{Yield}}{\partial\,\text{Investment}} \qquad 2.1$$

The first term, profit from a given change in yield can be calculated from the benefits listed above. The yield improvement per resource invested can itself be divided into two components: the yield improvement per process improvement project divided by the investment necessary for that project.

$$\frac{\partial\,\text{Yield}}{\partial\,\text{Investment}} = \frac{\dfrac{\partial\,\text{Yield}}{\text{Project}}}{\dfrac{\partial\,\text{Investment}}{\text{Project}}} \qquad 2.2$$

This might suggest that one should focus on those projects that offer a large impact on yield or those that are very easily accomplished, although the tradeoff between the two factors might find a higher yield improvement per resource somewhere in the middle of the high impact-low cost scale. The last term, resources needed per project, can be calculated by considering the costs outlined in the previous section on a project by project basis.

The final unknown, yield improvement per process improvement project, is not at all straight forward to calculate in this environment. Unfortunately, one of the problems faced by many companies, especially those with yield loss (one can think of each process step having its own yield; 'yield loss' then is the die loss for a single step) spread in many places throughout the process is that there is often little known about the yield improvement possible per change made by allocating resources to a given process control project. It is often difficult to know where yield loss arises in the process, let alone by how much this loss could be reduced through changes to the process.

Accepting now that process improvement is a worthwhile endeavor in this new environment, how should MWTD go about improving the process? Which projects should be attacked first and, looking forward, how does one set in motion a continuous improvement process that continues to solve problems with the highest priority? An answer begins with a review of process control and how it changes the manufacturing environment.

2.4 Basic Theory of Process Control

Theory on control of a manufacturing process begins with a clear definition of what is meant by "process" in this context. Here the term is taken to be any set of actions which together effect some change in one or more inputs, be they materials, procedures, or operator efforts, to create some desired set of outputs. To control such a process is to control the inputs and the actions of the process in such a manner that the nature of the output is as desired. Control of an output rests on the output being "measurable", since one cannot control what cannot be definitively measured. Common ways of looking at the output of a process are the Control Chart, figure 2.2, and Histogram, figure 2.3. Both of these tools are in some use at MWTD.

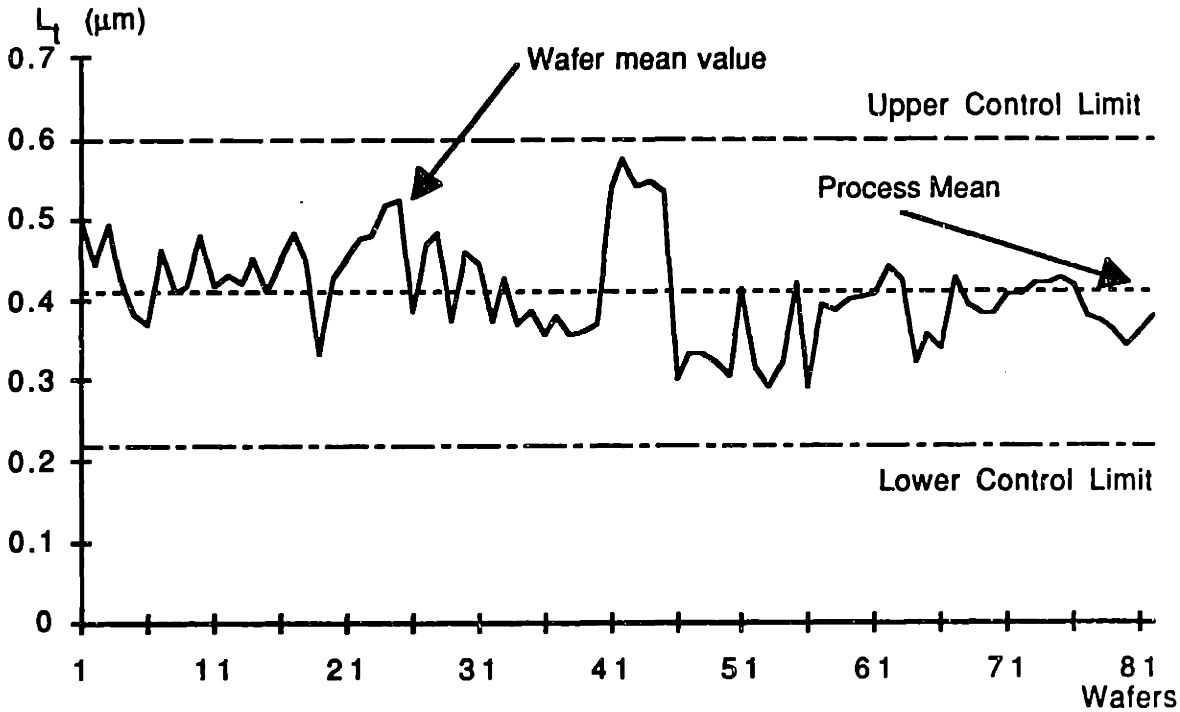


Figure 2.2: Sample Control Chart of L_p (PMMA line length)

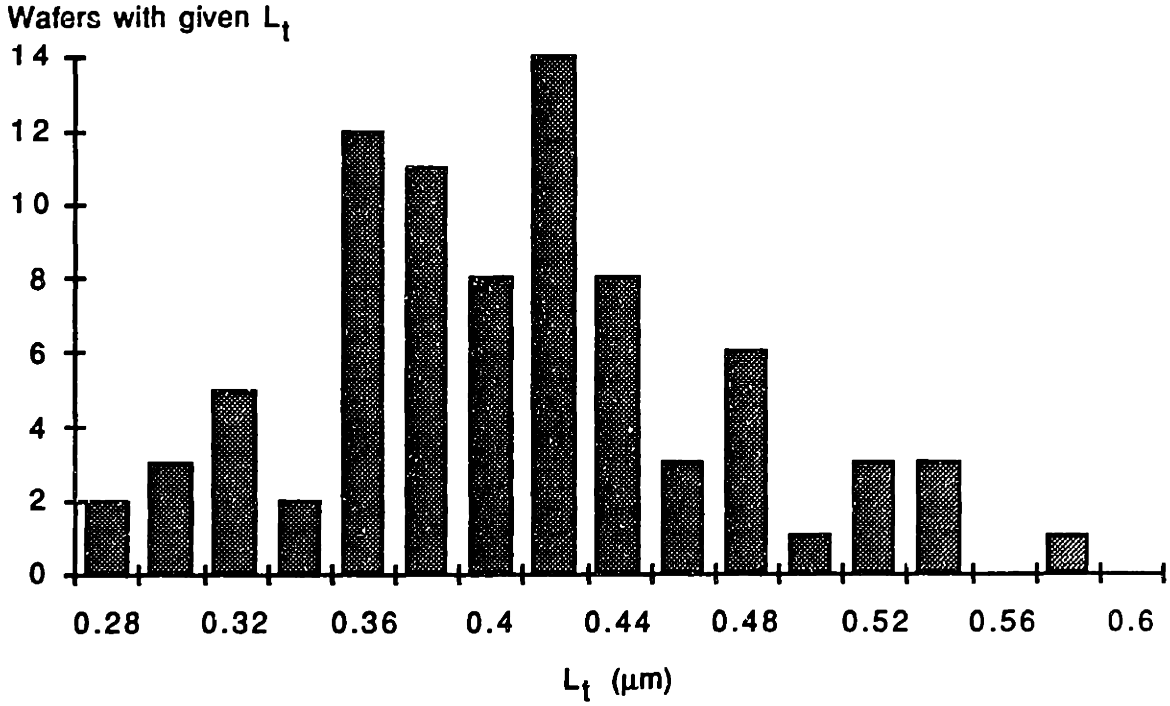


Figure 2.3: Sample Histogram of L_t (Transfer Layer length)

Conventional views hold that the goal of controlling a process is to maximize the percentage of the output that falls within the specifications, i.e., the process yield. This goal is sometimes termed "Conformance to Spec". In this context, yield will be determined by the percentage of parts that fall between the upper and lower spec boundaries with no concern how close the part is to the target value, provided it is within the specification limits.

Taguchi has argued that a more relevant goal is to press for closeness to target, versus some binary in-spec/out-of-spec line.[28] By developing a parabolic "Quality Loss Function", Taguchi suggests a measure of customer satisfaction that attaches greater value to an output that is closer to the customer's desired target value as compared to another output which, though also "in spec", is further from the desired target.

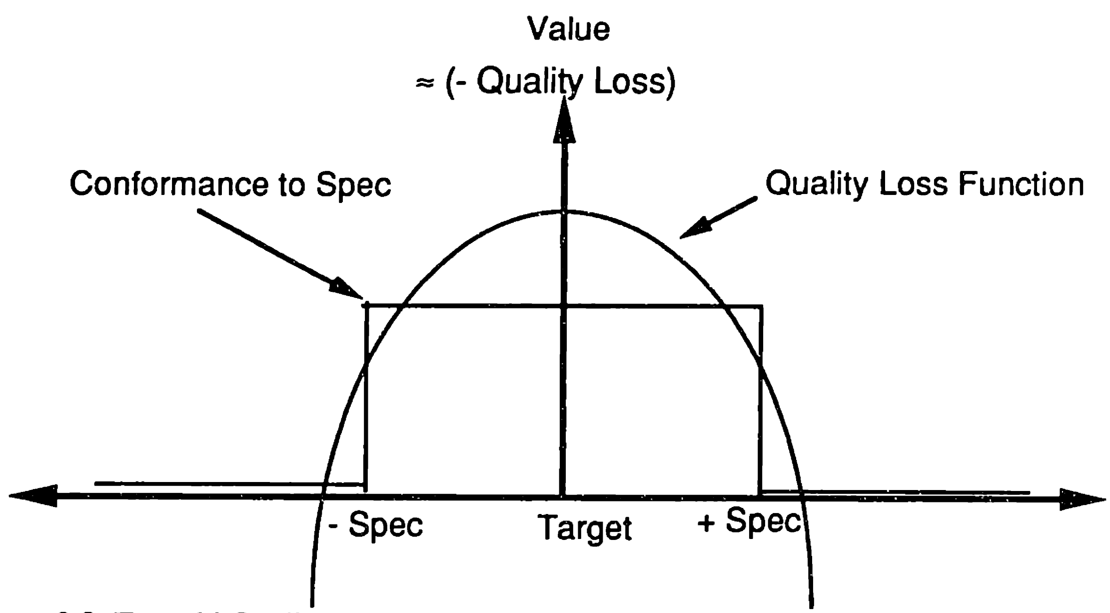


Figure 2.3: Taguchi Quality Loss Function versus "Conformance to Spec"

Yield loss of the MMIC-A process is known to have four major components:

1) Breakage of entire wafers due to the brittle nature of GaAs (other in-fab yield losses are grouped here also);
2) Patterns that have visual discrepancies in the 'printed' circuit pattern despite passing all electrical specifications test. (At MWTD, it is believed that visual defects can adversely affect long term circuit reliability; therefore, these parts are considered 'bad' even if they currently meet electrical specifications)
3) Failure of the circuit to operate, usually due to failure to "print" a complete circuit; and,
4) Failure to meet one or more of the parametric electrical specifications, given that the device operates.

By far the most conceptually challenging of these is parametric electrical yield loss. For this reason, it was the only yield loss mechanism investigated in greater depth. Note though that parametric yield loss is believed to be responsible for only 27.5% of overall die loss.

In controlling process yield, there are three methods of control:

1) Re-target the process mean to more closely match the desired process target value.
2) Narrow the observed distribution of outputs to be more closely centered about the mean.
3) Prevent/React to a statistically unexpected pattern of outputs, e.g., when values fall outside the 3 sigma limits of the process, the possible causes are investigated because this is a low probability event.

Current methods at MWTD pursue primarily categories 1 and 3 above. During the design and development phase of new processes, designers are charged with coming up with that combination of process inputs that produces a desired end product. Typically, this implies targeting process means. Occasionally, because of the continuously evolving nature of a process design and the somewhat informal procedures for qualifying a new process for hand-off to manufacturing, such targeting activities continue in the manufacturing environment. Experience in semiconductor manufacturing shows that once a process is targeted during early manufacturing, variability of the output is the more important metric. Thus, once a process is established in manufacturing, achieving conformance to spec is largely dependent on finding and controlling the central causes of variability in output as the process runs over an extended period of time.

The second process control method that dominates current operation at MWTD is reacting to "problems". Such "problems" occasionally are observed with control charts, but more often are driven by unexpected yield loss. This "fire-fighting" was observed to consume some 33% of available manpower in the process engineering group, precluding investment in more proactive means of process control, such as the second control method outlined in the list above.

2.4.1 Reducing Process Variability

Narrowing the spread of process outputs can be accomplished by a series of several steps:

a) Identifying the causes of process variability.

b) Acquiring resources to allocate to removing or reducing some of these causes. (Such acquisition should, nominally, require showing a superior return when a given process improvement project is compared to other projects).

c) Eliminating or reducing the cause of variability by changing some aspect of the process, its inputs, or its control mechanism: equipment, standard operating procedure, operator adherence to procedure, materials, operating point, or control method.

When MWTD does engage in variability reduction, technical literature and experience-based knowledge of typical causes of variability are utilized. As will be shown in Section 3.3, the resource allocation process does not allow for dynamic adjustments in manpower which is typically the true bottleneck resource in process improvement. This largely limits investment in such variability reduction to projects that can be staffed with existing manpower. However, even these projects are a second priority to line support activities which are needed to meet current production demands. When a variability reduction project is appropriately staffed, engineers seem to be effective at reducing or eliminating the cause of the variability, although difficulty in directly observing the output of a process and a fairly low volume output often complicate showing that a specific change definitively reduces variability in the output. This prevents either positive or negative reinforcement of whether perceived causes of variability are actually the important ones and whether these causes have been eliminated by a given process change.

Current methodology at MWTD suggests that improved information on the precise contributors to variability and their relative magnitudes could improve the use of resources in two ways. First, currently available resources could be definitively allocated to the highest value process improvement project. Second and of even greater value, an understanding of the precise benefits available from attacking a variability source will allow more effective tradeoffs to be made against other types of projects.

2.4.2 Finding the Causes of Process Variability

While there is a large body of literature that considers the experimental methods necessary to identify and quantify the sources of process variability, Clausing[29] and Hayes, Wheelwright & Clark[30] have suggested that such methods too rarely begin with proper focus on meeting customer needs and wants. In trying to set the proper context for an application of experimental methods, I have used the following conceptual guidelines for identifying the causes of process variability:

1) Understand precisely the customer's needs, concerns and wants. Ideally, these are expressed in some explicit set of specifications.
2) Translate these attributes to the process inputs necessary to meet these customer requirements. For example, meeting a customer specification of 0.220 dB/GHz gain slope might require spinning PMMA at 4000 ± 100 rpm for 30 ± 2 sec, among several other process inputs. This implicitly requires that one measures the effect of variability in process inputs on meeting customer requirements.
3) Measure the input variability in all or some important subset of these process inputs. For example, as the process is currently exercised, observe the variation in spin speed and spin time.
4) Multiply the input variability by the effect(s) found in step 2 to determine the relative contribution to variations in meeting customer desires.
5) Use these contributions, together with a knowledge of the customer specifications, to calculate the effect of variation in a particular process input on final customer yield of good product. Rank order these effects in terms of maximum variability in customer requirements.

Following these guidelines answers the final unknown from Section 2.3.4 on allocation decision criteria: yield improvement per process improvement project.

Thus the goal of this thesis was to identify where exactly variation was occurring in this process. Given the nature of the process, this could only be accomplished by a detailed technical understanding of how the device and process physics involved might contribute to variation in devices produced. Knowledge of where the variation in a process is occurring is one key piece of information needed in being able to make more explicit tradeoffs between investing in process control activities and other resource allocation options.

2.5 The Hypothesis of This Research

Working from the foundations and background presented above, I set out to examine five hypothesis:

Hypothesis 1

MWTD's process for allocating resources suffers from many of the problems previously reported in the literature.

Hypothesis 2

Having recently undergone a substantial change to the business, MWTD fails to pay sufficient attention to newly critical variables in allocating resources.

Hypothesis 3

This neglect of critical variables is significantly costly to the division, yet they cannot currently estimate this cost.

In collecting the data to test these three hypothesis, it is necessary to pursue two additional hypothesis.

Hypothesis 4

These critical variables, particularly variability in circuit electrical parameters, can be measured and gainfully incorporated into resource allocation decisions through careful consideration of the manufacturing metrics that will define success in the new environment and detailed analysis of the relevant device and process physics.

Hypothesis 5

Variability in circuit parameters of concern to customers can be traced to variability in a limited set of process inputs. This variability can be measured and its effect on the circuit parameters estimated.

Taken in total, the responses of the data to these hypothesis should suggest how resource allocations are currently being made at one industrial site, the cost and competitive impact of this behavior, and whether improved information can be obtained and used to enhance the current allocation process.

Chapter 3: Current Resource Allocation In MWTD Manufacturing

To explain how resources are allocated at MWTD, I begin with a consideration of the situational and structural environment within which these allocations are made. I consider first the situational environment: the site in general, its history, its standing in the marketplace, and the manufacturing process technology involved. Continuing with the structural environment, I consider the organization, the incentive systems used and the way in which strategy is formulated. Upon this foundation, I attempt to build an understanding of how allocations are currently being made at MWTD. A useful framework in considering MWTD's method is to look at how they address each of the conceptual elements involved in allocating resources:

1) Generating allocation options
2) Understanding costs, benefits, and risks of options
3) Acquiring resources
4) The allocation decision maker
5) The allocation decision criteria

After I cover MWTD's approach to each of these elements in this Chapter, in Chapter Four I show that this approach appears to encounter many of the known problems in allocating resources as reported in the literature. One of the conclusions I draw is that following a major change to the business, MWTD seems to not be paying attention to variables that are critical to the correct guidance of the business in this new environment.

3.1 Situational Environment of HP's Microwave Technology Division

3.1.1 Choice of Industrial Site

This research was conducted over a six month period at the Microwave Technology Division (MWTD) of the Hewlett-Packard (HP) Corporation. This division supplies leading edge solid state components operating in the RF, Microwave and Lightwave frequency ranges on a captive supplier basis to HP's microwave instrumentation divisions.

As an organization, MWTD is charged with providing two things:

1) A continual series of designs for leading edge solid state components operating in the RF, Microwave, and Lightwave frequency ranges. Ideally, these components give HP's instruments a performance advantage over competitors.
2) A manufacturing process capable of producing such devices to meet the needs of HP's Microwave instrument divisions.

After several years of relatively flat sales, the division has set out to dramatically increase transfer revenues (MWTD sets transfer prices to break even on an annual basis; R&D expense is billed separately) over the next five years expecting to grow approximately 150% over current transfers of $ 40 million to $100 million (data has been disguised to protect confidentiality). While the division has a high mix of products, growth projections for the future show that one particular class of components, Monolithic Microwave Integrated Circuits (MMICs), will account for the vast majority of this growth, approximately 80% of tota' growth.

Accordingly, this class of devices was the principal product line studied.

The only current process for manufacturing these devices is known as MMIC-A and it has been producing parts since mid 1988. The manufacture of MMICs involves fabrication of sub-micron gates with precisely controlled dimensions. These dimensions, particularly gate length, directly affect the critical performance parameters of the device, namely cutoff frequency, gain uniformity, and noise. The ability to control these dimensions will be a key determinant of manufacturing success in the coming years.

3.1.2 Division History

The MWTD was created in 1973 to serve as a technology enabling division. By producing devices such as high frequency transistors that were not available commercially, HP's instrument divisions were able to enjoy a performance advantage in the marketplace. Through the late 1970's and early 1980's, growth at MWTD ran about 20% largely due to the technical success of MWTD's products and the similar success of many of the instruments in which these MWTD components were used.

As the 80's progressed, several factors began to change MWTD's ability to provide high end technology at a competitive price: competition from overseas in "cash cow" markets, spiraling increases in the cost of fabrication equipment necessary to produce ever more sophisticated circuits and devices, spending cutbacks at the head of the "food chain" (Department of Defense contracts) which stagnated sales of microwave instruments, an inability to deliver a succeeding generation of FETs, and the nature of the cost allocation of MWTD's costs. These factors conspired to lead MWTD to raise component prices, even while competing external vendors were lowering them.

HP's microwave instrument designers, allowed to choose freely among all vendors and charged with final cost of the instrument, designed in less MWTD parts further deteriorating the base for allocating the enormous overhead costs. This led MWTD to further raise prices to cover non-allocated overhead.

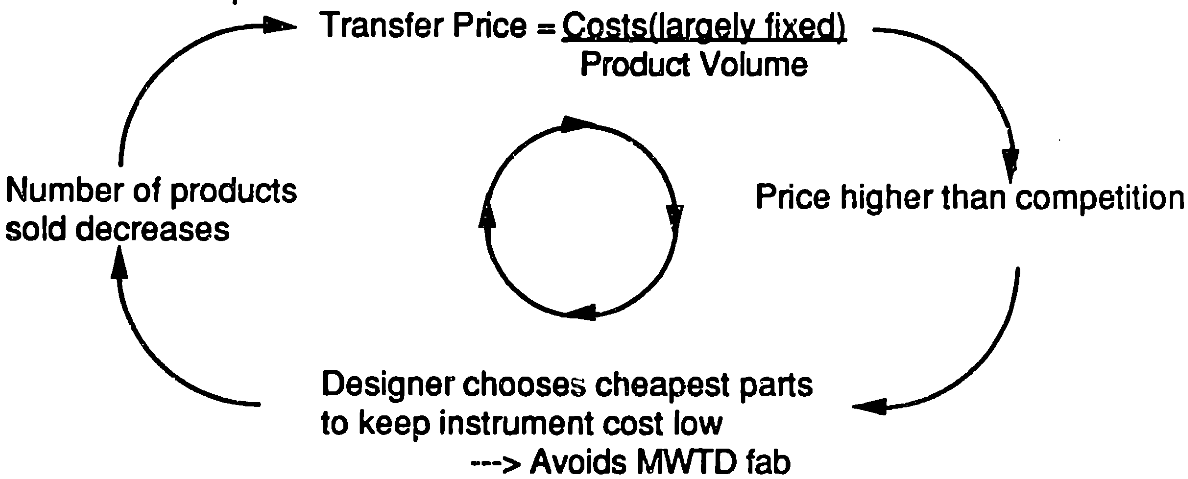


Figure 3.1: The "Death Spiral" of setting transfer prices to fixed costs over a diminishing product volume

Traditionally, MWTD billed <u>all</u> costs to customer divisions through an elaborate transfer pricing scheme. In the mid-1980's, MWTD was -for the first time - allowed to build up a slight residual ("residual" is the amount of MWTD costs not billed to customer divisions through transfer pricing on devices produces. Costs include cost of goods sold, depreciation charges, inventory change, administrative costs, but not R&D expense. The goal is to keep residual equal to zero by setting transfer prices accordingly) instead of continuing to raise part prices; however, even then, the residual was allocated back out to the customer divisions at a different organizational level. This pattern of attempting to assure full allocation of costs to customer divisions by raising prices or billing a residual would, if continued, cause a 'death spiral' (see Figure 3.1); nonetheless, this pattern of behavior continued into 1988 when the increasing residual finally forced MWTD to react. At this point, MWTD remained a technological leader on most products, but was not cost competitive on the more established products. Furthermore, MWTD management acted under the perception that the stall in revenue growth would not provide sufficient growth paths for its employees, nor allow for sufficient new blood to enter the system by external hiring.

In mid-1988 the seriousness of this situation demanded a response from management. The reaction arose from an extended meeting of the division's management team. It seemed clear to them that current business trends would only widen the residual and yet the corporation was demanding that all divisions contribute their share towards the overall corporate growth objective (In the case of MWTD, this implied rapidly eliminating the residual). Furthermore, on a technical basis, they felt that it was necessary to continue investing heavily in capital equipment if they were to be capable of producing leading edge components in fulfillment of the division's charter.

The only available path seemed to be greatly enhancing revenue; however, MWTD was constrained by their marketplace, the instrument divisions, which were essentially set at a slow growth rate. After considerable debate, it was decided to pursue sales to external customers. Because MWTD had never sold on the external market and, more so, because MWTD's primary goal was to provide key enabling technologies to the instrument divisions, this was not a popular resolution and even two years later, it does not have the support of the entire division.

With a somewhat unpopular solution, the challenge became making this rapid growth in volume and development of external markets happen. The process by which this is being done is now the overriding theme at MWTD and reveals much about the current operation of the division. The slow down in revenue and this plan to respond have done a great deal to remove MWTD from what appears to be a sense of complacency that had set in after many years of concentrating wholeheartedly on their technology and being quite successful at providing HP's instrument divisions with an unchallenged technical advantage, albeit without much significant competition.

One is left with a picture of a division that has recently undergone a major environmental shock. While the shock began with a significant change in the marketplace in 1986, the division did little to respond until 1988. As I will show, while their initial response to this change seems well aimed, their ability to pursue this radically new vision may be hindered by the structural processes that brought them success in a prior environment.

3.1.3 Market

MWTD manufactures a wide range of leading edge, solid state components operating in the RF, Microwave and Lightwave frequency ranges. Typical components include MMICs, RFICs, FETs, diodes and capacitors. Within any class of products, MWTD manufactures both leading edge components and several older generation components still being used in older generations of HP instruments. While MMICs currently make up only a small fragment of total sales, it is predicted that they will account for most future growth.

Until recently, the entire market for MWTD products has consisted of other HP divisions within the Microwave Instruments Group, particularly Networks Measurement Division and Signal Analyzer Division.

Components from MWTD are used in instruments such as Network Analyzers and Signal Analyzers. Principally, these products vary in terms of the frequency of operation and amplification of signal available. These products are sold commercially to a variety of users in the commercial electronics, telecommunications, and defense electronics industries. While HP avoids selling directly to the government, so that it can avoid the associated mandatory full financial disclosure, a majority of the Microwave Communications Group's dollar volume of instruments is sold to Department of Defense primary or secondary contractors.

MWTD will soon expand their list of HP divisional clients to include new ones that will package, market, and/or distribute MWTD's parts to the expected developing base of external instrument customers.

As mentioned previously, MWTD is not without competitors in its efforts to sell circuits and devices to the various HP customer divisions (and to other instrument manufacturers). Included among these are Texas Instruments, NEC and Anritsu.

3.1.4 Manufacturing Process Technology

While the literature has generally not covered the particulars of the technology involved in observing the resource allocation method, I will show in Chapter Four that many of the critical variables ignored by management after an environmental shock are strongly intertwined with the particulars of the technology involved. Thus, to understand the process of resource allocation at MWTD and what information it does or does not include, it is important to first understand the details of the manufacturing technology involved.

MWTD operates "production lines" established around one of several technologies, such as GaAs (a III-V compound) or Silicon. Each of these technologies is capable of producing one or more product types, such as integrated circuits (ICs) or resonators, by using different routings along the production line. (see Figure 3.2) The flagship line is the GaAs fabrication line.

Process Technologie	Component
III-V Compounds	FETs
Acoustic Waves	Diodes
Magnets	ICs
Thin Film	Modulators
Silicon	Filters
	Resonators
	Spheres
	Thin Film Circuits
	Si Components

Figure 3.2: MWTD's Process technologies and Components produced

Of the several processes that compose the III-V technology, this researched focused on the most recently introduced process, MMIC-A. The MMIC-A process is an eleven mask process (implying that a wafer makes approximately eleven cycles in the production line, each one involving a trip through the mask center) capable of creating MESFETs with 0.42 μm long gates. These devices are the active building blocks of a circuit. Together with other, simpler elements such as resistors and capacitors, they can be combined to form a wide variety of circuits all integrated onto one piece of semiconducting GaAs. Some common measures of the complexity of a semiconductor fabrication process are presented in Figure 3.3 for the MMIC-A process.

• integration level (relative number of transistors on a single chip)	low scale
• number of mask levels	11
• type of lithography	contact, lift off
• smallest feature size, nominal	0.42 μm

Figure 3.3: MMIC-A Process Specifications

For a fuller, description of the process, see Section 5.4

<u>Process Development</u>

In its role as a supplier of technology, MWTD must strive to constantly introduce state of the art products and manufacturing processes capable of producing them. Typically, processes are developed together with at least one circuit that can be used as a test vehicle. Early development work is often done in an associated part of HP Corporate Labs. Once feasibility is proven, the development is transferred to the R&D group at MWTD where it begins life on a formal New Process Introduction Cycle. (see Figure 3.4) It is important to note that the work at Labs, in R&D and in Manufacturing typically is done on three different sets of process equipment.

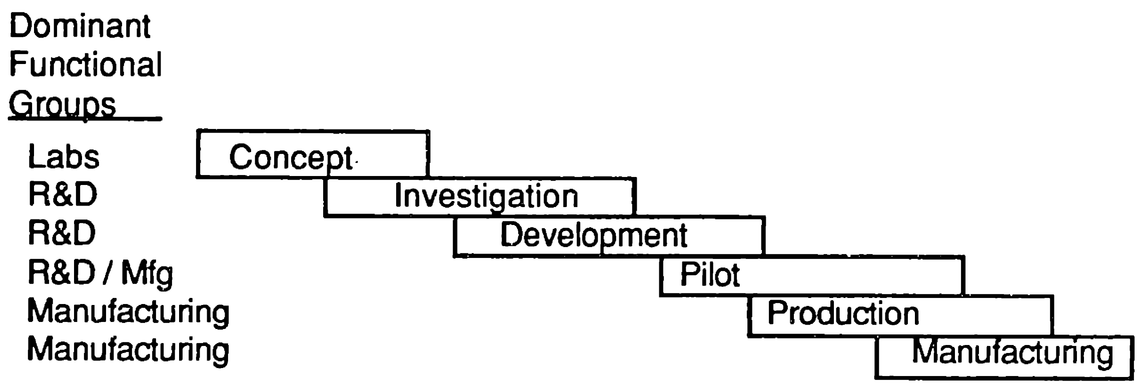


Figure 3.4: HP's New Process Introduction Cycle

The move from one phase to the next is determined largely by fulfillment of set objectives. Only recently, have these objectives begun to include goals for manufacturability, typically in the form of process yields. Manufacturing becomes progressively more involved as the phases progress and assumes lead responsibility after the pilot phase.

3.2 Structural Environment of HP's Microwave Technology Division

3.2.1 Organization

HP generally organizes as a corporation of largely independent divisions each with its own development, manufacturing and marketing staffs. One such division is MWTD. These divisions are grouped into related lines of business. For example, MWTD and five instrument divisions form the Microwave Communications Group. (see Figure 3.5)

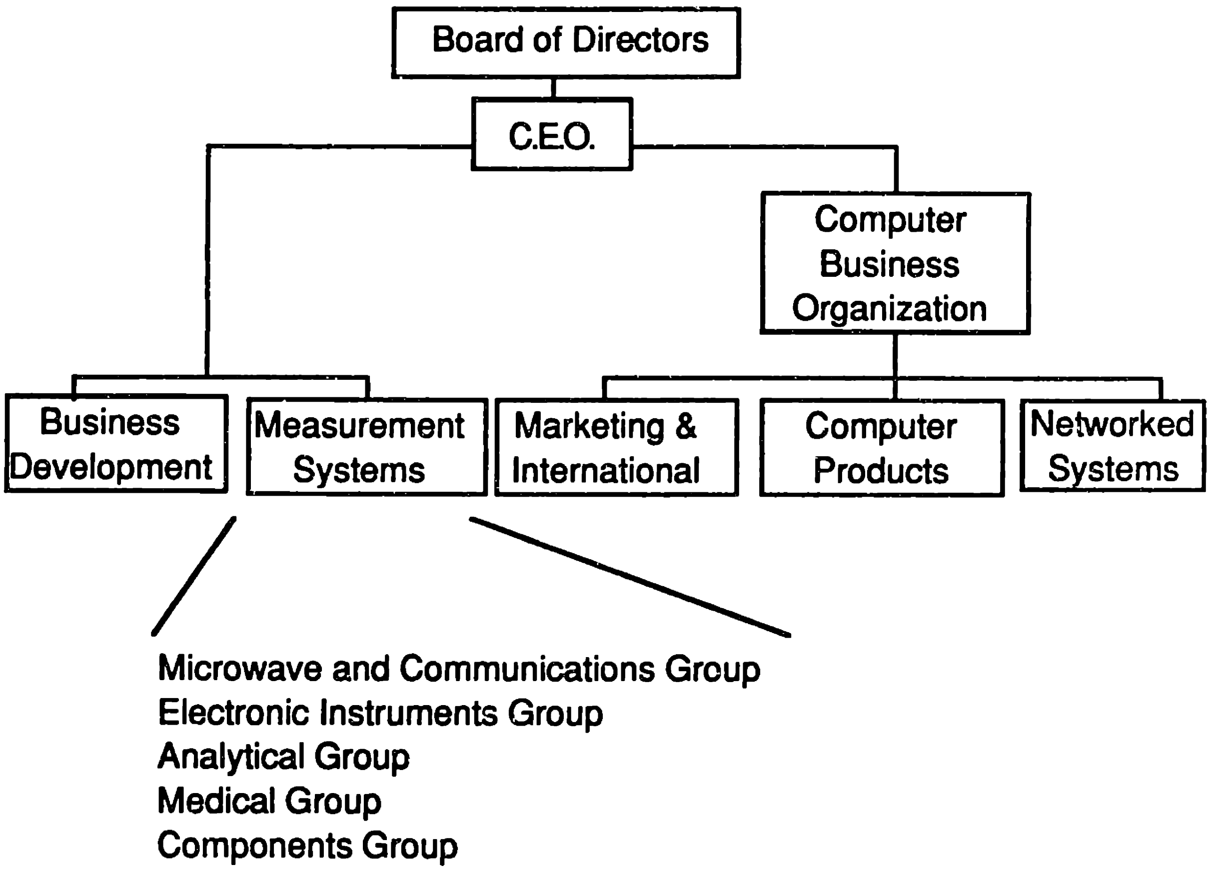


Figure 3.5: HP Corporate Structure

As the business needs call for it, certain functions are shared between two or more divisions. This is quite common with the Sales organization, early product research (HP Labs), divisions which together offer a "systems" solution (primarily computer systems divisions), and, increasingly, divisions that fabricate components for consumption by several HP divisions. In particular, this is done with integrated circuits, printed circuit boards, assembly of surface mount components and sheet metal. MWTD is one such captive supplier. While much of this sharing is new in the last five years, MWTD has acted as a common, captive technology supplier since 1973.

Internally, MWTD follows much the same organization as all other HP divisions with two notable exceptions. First, the marketing organization has traditionally been highly technical and somewhat weaker due to all customers being other parts of HP. Much of the Marketing function has fallen, de facto, to the design group which often interfaces directly with engineers from customer divisions. Second, the development engineering group is a semi-distinct entity apart from MWTD because product and process development is funded directly out of the development engineering budget of the various instrument groups rather than through the transfer price of delivered components.

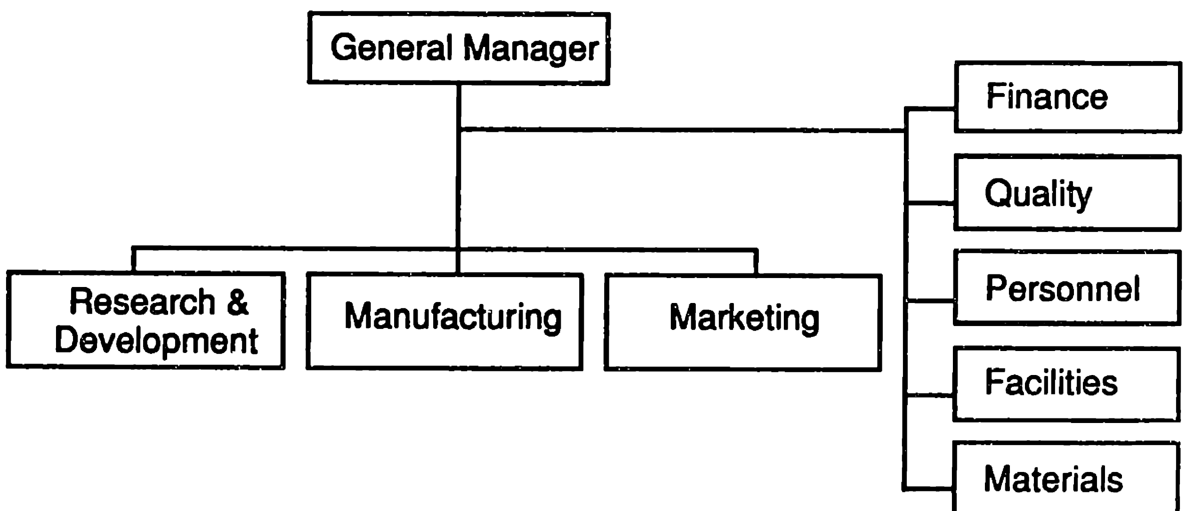


Figure 3.6: MWTD Functional Organization

The manufacturing function is divided into three groups. One that fabricates semiconductor devices. A second group that fabricates other types of electrical components such as thin film circuits, surface acoustic wave devices and YIG oscillators.

Included in this second group is the associated engineering support staff. The third group is semiconductor manufacturing engineering which supports the semiconductor production line: introduces new products and processes to manufacturing, and works to improve the current production processes. Due to the highly technical nature of the process, manufacturing engineering is a rather large and powerful group. Unlike traditional manufacturing organizations, the logistics/materials function reports directly to the general manager.

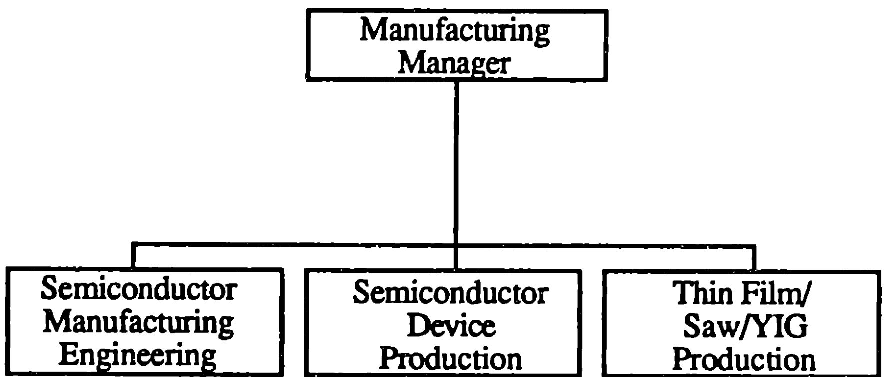


Figure 3.7: MWTD Manufacturing Group Organization

The semiconductor manufacturing engineering group is divided into six smaller operating level groups. The Process Engineering group, which is the central focus in this study and one key group in the effort to improve in-fab process control, has the responsibility for maintaining, improving and adding to the processes involved in fabricating a finished wafer. The engineers in the Process Engineering Group each have responsibility for one set of associated process technologies. Forming a matrix with these technologies, see Figure 3.7, the Device Engineering group has the responsibility for the fabrication of completed devices and is divided according to device type. This group is also of importance in improving in-fab process control. Thus, for any given problem, there is likely to be an interested device engineer who's device is not yielding well and an interested process engineer who's process is causing the yield loss.

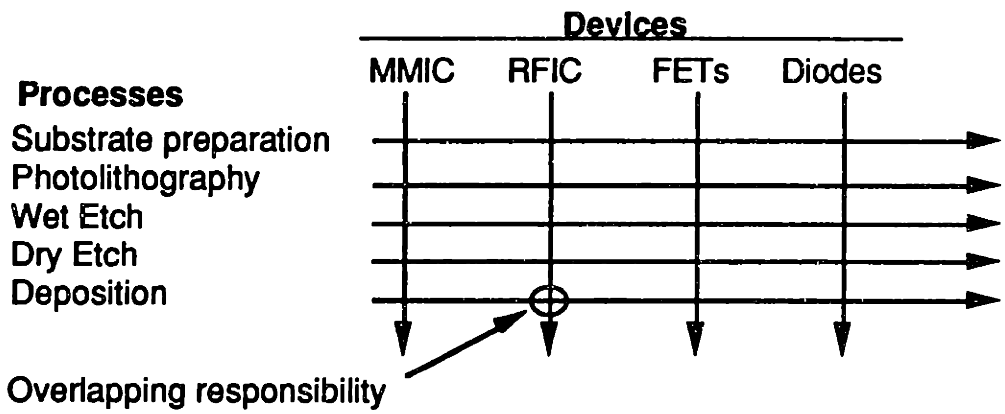


Figure 3.8: Overlapping responsibilities of Device and Process Engineering

The other four groups within semiconductor manufacturing engineering offer testing and packaging support, Silicon process support, and support in establishing and utilizing technical databases that aid all of the other engineering groups.

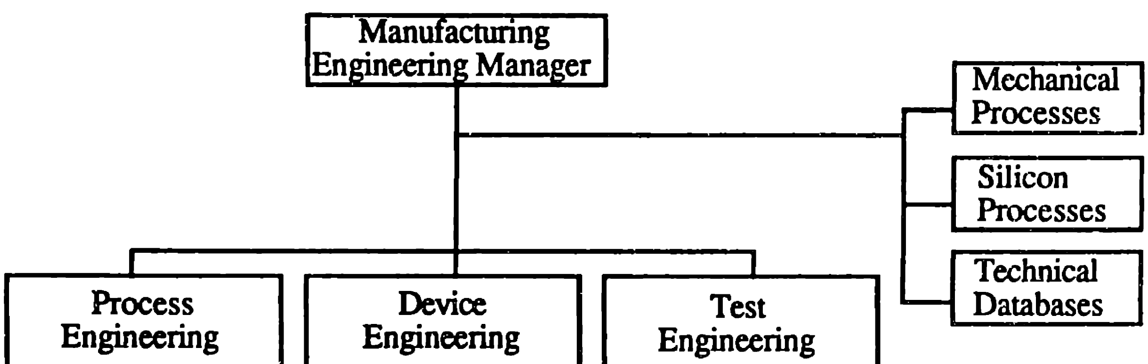


Figure 3.9: Semiconductor Manufacturing Engineering Group

3.2.2 Strategy Formulation

The process of setting division level strategy has changed in recent years due to two factors. First, the growth of the residual (and associated other problems) and the decision to fight these problems through a major increase in revenue have focused the division very clearly on one 'vision' for the future. Second, the adoption by HP of a new strategy planning method, known as Hoshin, has modified the prior model for strategic planning.

Formulating the Vision

Once the division management had agreed, albeit with internal dissension, that revenue growth and external sales were the keys to future success, the question became how much revenue growth was necessary and how were the goals to be achieved. Current projections of shipments to internal customers were $60 Million for 1993, which represents 11% growth per annum, but only 2% manpower growth. Having little information on the potential size of the external market and not knowing how their current products might fare in that market, they had little to go on in setting a revenue target: $100 million was settled on as a target number mostly because it sounded like a reasonably aggressive goal, but also one that might be attainable.

Once agreed on "$100 million in 1993" helped to awaken and unite the entire division as to the primary challenge to MWTD and the time frame in which this goal must be accomplished.

With a clear target in terms of revenue established, the management team set about finding a methodology to make this goal become reality. The planning process by which they undertook this is called Hoshin Kanri, Japanese for "Strategy Management". This method focuses on setting two or three breakthrough goals which are essential to the success of the organization and require significant changes in the process or structure of the business. The Hoshin planning process also provides a framework for cascading these goals and strategies to reach them throughout the organization. At each succeeding level of the hierarchy, that function or group is expected to repeat the process coming up with two or three breakthrough goals for their group and strategies to meet these objectives. Because managers are still learning to use it, the process is not yet woven into the texture of the organization and it is unclear how 'top-down' this process actually is in practice. There is some confusion because the objectives are supposed to be truly 'breakthrough' in the sense that they would not be accomplished in the normal course of business without the extra focus that this process will provide. Thus, these objectives are different than those included in traditional strategic planning. On the flip side, these objectives must be tempered with a degree of realism so that they can, in fact, be reached. Such realism is interjected by managers who receive an unrealistic hoshin goal as they complain and force the goals to be altered.

The three breakthrough objectives that management focused on were:

- development of external customer sales
- introduction of key new products/processes on schedule
- Establishment of a path to control and improve operating processes.

Clearly the second and third objectives above were being pursued in at least some measure before the Hoshin process was used. After the first year of using the Hoshin process, management agreed that they had failed, in retrospect, to emphasize truly 'breakthrough' objectives during their initial use of the Hoshin process.

As an example of the cascading, one can look at how these division level hoshin goals affected the Process Engineering Group. The current strategy imposes two particular challenges to the Manufacturing function:

1) Handling far higher volume without a substantial investment in capital equipment
2) Handling a multitude of new products

The plan to handle these challenges in manufacturing relies on three changes within manufacturing:

1) Faster new circuit and process introduction
2) Reduce cycle time through the fab with existing equipment by improving scheduling
3) Greatly improve yield

The third goal was of central importance to the process engineering group, because yield increase would require tighter process control.

3.2.3 Incentive Systems

An implied norm at MWTD, and seemingly throughout HP, is to reward technical prowess. Until recently, HP operated a "dual ladder" in which one could continue to progress at an equal rate while remaining wholly technical or moving into management. This system was discarded on the basis that a strong technical person who can leverage his or her talents by also managing others is necessarily of greater value to the company than a strong individual contributor. Nonetheless, one can achieve substantial, even if not equal, pay while remaining wholly technical.

Pay increases for engineers are based on rankings against all other engineers in the division. Rankings are done by the management team as a whole although the direct supervisor carries a seventy percent weighting in the final outcome. Differing views on what defines technical prowess from an engineer make the outcome of the ranking process difficult to predict. Traditionally, R&D has been considered more challenging and prestigious than manufacturing, although this attitude is starting to wane lately. Each supervisor exercises considerable latitude in choosing the attributes he or she wishes to reward. The Process Engineering supervisor weights heavily good experimental design, use of statistical quality control techniques, timely response to "customer" (production line) needs, permanent fixing of problems to eliminate future firefighting, and the accomplishment of these responsibilities with minimal supervision.

3.3 The Allocation of Resources at MWTD

3.3.1 Generation of Allocation Opportunities

Opportunity generation follows much the same path laid out by Bower. Through the Hoshin planning technique, strategies and goals that achieve those strategies are passed down through successive layers of the organization until reaching the operating groups. But then actual project options tend to arise at the bottom level from one of three sources.

The Process Engineering Supervisor is one key originator. In choosing projects, the supervisor looks for problems that are currently creating the highest yield loss. Because yield numbers throughout the process are not reported on a regular basis, the supervisor must rely on fragmented bits of information and his or her experience to develop his or her own sense of which process problems are most important. This is not an easy task as the group supports several different process technologies, each of which has considerable process variability.

A second class of originators, the device engineers, are often better able to spot major process problems because they follow a single device through all of the processes.

A third project generator, the operators on the production line, interact directly with the process engineers to solve or explore problems they notice during the course of production.

Thus the pattern of opportunities generated is motivated by the functional goals set forth under Hoshin planning, yet most of the originators seem to suggest options based on where <u>problems</u> are observed. The only major exception to this is the process engineering supervisor who, at times, pursues a broader agenda of process control projects according to his or her sense of where the greatest long term opportunities for reducing yield loss reside.

3.3.2 <u>Understanding costs, benefits and risks of options</u>

Much as described by Bower, MWTD does not rely on explicit and directly comparable assessments of costs, benefits, and risks. Rather they rely on 'measurements' of perception and measurements made along scales that are not directly comparable.

Costs are loosely estimated by the engineer in charge of the process area involved. Typically, these 'costs' include man-weeks of engineering effort involved and capital costs, if any. It is rare to consider the cost of production line time as most engineers assume this to be an elastic commodity available at no cost provided one "works around" scheduled production flow.

Only two types of benefits are generally considered at MWTD. The more typical seems to be "reduction in yield loss". Typically, an inexact measure of yield loss due to a particular cause is used to justify any project that might reduce the yield loss. For example, a year old calculation of wafer breakage is used to justify a host of breakage reduction projects. A second type of benefit that is considered is "reducing cycle time", where a process that appears to be currently holding up production is worked on to remove the cause of the bottleneck. While this often implies increasing process yield, the benefit motivation is to reduce cycle time by getting production flowing again and the increase in yield is typically limited to restoring the prior process yield value.

Typically, technical risks are the only type of risk considered, i.e., can the problem be found and eliminated. For projects that require primarily manpower investment and little capital, the assessment is typically left to the engineer and his or her supervisor. The possible outcomes tend to be binary: I can do it; I cannot do it. On major capital equipment decisions, the assessment is made by an entire team of non-management personnel who spend 2-3 months evaluating several dimensions of the purchase.

3.3.3 Capital Allocation

The primary determinant of resources allocated is the current allocation to that task or functional area. Beyond this common determinant, the acquisition of capital and manpower proceed under very different systems at MWTD.

Capital Budgeting

Capital Budgeting follows a relative standard pattern as described by Bower. Budgets bubble up from the operating group and can be pared at each management level. From year to year, capital budget amounts for a given group are quite steady or slowly increasing. Only when a distinctly larger amount of capital is budgeted is rigorous justification likely to be necessary; nonetheless, once a budget is approved, the likelihood of later rejection of an accepted budget proposal is small.

Capital Requisition

Having finished the budgeting process, no capital has actually been allocated. This requires a second step of justifying any single purchase. Much of this "justification" comes in the form of an elaborate and formalized scenario for investigating a capital purchase. A non-management team representing the various functional groups involved in a purchase spends 2-3 months investigating in detail how and what to purchase. While this teams tends to assure the successful choice and - usually -

implementation of capital items that are purchased, their primary function is not to evaluate the cost/benefit return on such a purchase except in so far as the equipment contributes to the technical accomplishment of strategic goals. While a traditional "hurdle rate" calculation is required, because many of the capital investments enable future technology or the ability to manufacture such technology, projects need not - and often do not - show a positive financial return. It is usually enough to show that this project contributes to the division's mission.

This mentality may emerge, in part, from the corporate philosophy that the business should be run lean on manpower, but with abundant capital to support that manpower. Thus, management believes it is not a primary objective to greatly limit capital expenditures.

3.3.4 Manpower Allocation

The literature does not consider the issue of manpower allocation in much depth, but in a production environment reliant on many different and highly technical pieces of equipment, having sufficient manpower to order, install, and integrate the equipment cannot be ignored, yet at MWTD manpower planning and capital are only remotely linked. In general, current staffing levels are targeted to remain constant, which is not unexpected given the overall stagnation in instrument group revenues.

Increases in headcount can be achieved in three ways: by justifying a new hire, by internally redistributing workers, or by full scale reorganization. Justification is subject to management chain approval and is the first thing to slow when revenue or profit growth stalls. Redistribution is rare, happening only when an employee leaves the organization or a specific project can be supported at the direct tradeoff expense of slowing another project currently under way. Finally, reorganizations while not infrequent, are also not common. Recently, they have occurred about every two years. Typically, they are done to support a change in the primary strategic direction of the business. While difficult to assess, it is also apparent that political aspirations of the various managers involved play an important role in determining group size and reporting structure following a reorganization.

This limited scheme for re-allocating manpower suggests a limited dynamic response to major changes in what a given group needs to accomplish. Further, the lack of a strong link between manpower allocation and capital allocation is evident when one considers the large amount of capital equipment that sits for several months or even years before an engineer has time to install it.

3.3.5 The Allocation Decision Maker

HP, as a company, actively tries to push decisions making of all kinds down to the lowest possible level of the organization. The one force that can alter this is a strong willed manager who wishes to retain control for whatever reason. Particularly because of the very relaxed atmosphere at HP, and more particularly at MWTD, such a manager - if he or she is directed and vocal - can greatly alter where decisions are made.

Within this context, much of the decision to pursue, delay or ignore a process improvement project is made between the process engineers and their supervisor. Because of the current supervisor's particular experience relative to his subordinates, he tends to make much of the decision. As projects bubble up from below, the supervisor can quickly accept or dispense of an idea with this decision largely deciding the fate of the project. If a supervisor 'accepts' a proposal, he or she is implying that his or her current staffing level can handle it.

If ideas pass this first level test, capital is typically already budgeted and thus largely approved. The project is carried upward in the organization, typically with the supervisor acting as champion. If the project involves only a commitment of manpower, it typically requires approval only from the manufacturing engineering manager. Even this 'approval' might be more accurately called 'notification' when the proposal calls for relatively short product duration.

If the project involves a commitment of capital, the process is more involved. The project will be considered all of the way up the management chain until reaching the appropriate final sign off level. At any management level, the project can be rejected, delayed, remanded for additional work or explanation, or passed on, yet it is relatively rare for decisions that meets the financial criteria to be rejected, although delays in the management chain are quite common.

3.3.6 The Allocation Decision Criteria

At this point, a description of the allocation decision criteria is largely repetitive - its key aspects having been incorporated in the elements of MWTD's resource allocation process already described. The final decision can be pictured as making one or more trade-offs subject to certain constraints. The most important tradeoff in this capital free, manpower constrained environment is typically competition for manpower. The tradeoff is typically made in favor of the higher yield improvement project, provided current production is reasonably on schedule. If the production schedule falls too far behind, this quickly elevates the priority of projects related to the production problem. It is also important to note that specialization of engineers further constrains assignments. Typically, each process engineer handles all projects within his or her process specialty.

Three factors determine much of the deployment: existing needs of the production line, the supervisor's perception of where change is needed, and agreement by engineers and the supervisor as to what to work on.

As with any manufacturing process, keeping the existing process performing to current standards requires engineering support. Necessarily, these calls for support come sporadically and demand immediate attention if the line is to be kept running. Furthermore, because engineers are given responsibility for a clear subset of processes, they are motivated to support any calls for assistance as expediently as possible. This support requirement creates difficulty in that the "fire-fighting" activity becomes de facto the number one priority and the stochastic arrival of support calls tends to interrupt work on projects with a longer term focus.

Engineering time beyond basic line support is spent in three areas: overhead activities, new circuit/process introductions (NCI and NPI), and process improvements. Overhead activities, such as training, non-focused meetings, and record-keeping, consume approximately 25% of an engineers time. Like support activities, overhead activities tend to get done de facto. Thus, the two activities that lead to improvements over current methods tend to be considered third priorities, to be fit in after overhead and support activities are completed.

Time spent on NCI, NPI, and process improvements is divided by agreement between the supervisor and engineer. At the time this study was conducted, the supervisor knowingly used a heavy hand in making these decisions because he felt that the group was too new and the engineers to novice to have a correct sense of priorities. As this study was ending, the supervisors was beginning the process of allowing the engineers to assume a larger part in this role feeling that he had sufficiently communicated the direction and trained the engineers in the criteria that should be used for setting project priorities.

Towards improving yield, the supervisor acted according to his perspective of where yield loss was occurring in the process. In particular, he felt that the key yield loss mechanisms were breakage and visual defects. This perception is not easily substantiated or disputed due to a lack of organized data. Accordingly, projects were chosen and priorities set according to this perception of where yield loss occurs.

Chapter 4: Effects of the Current Method of Resource Allocation

In Chapter Three, I presented how MWTD currently allocates resources to process improvement projects. This discussion was set within the context of MWTD's situational and structural environment. When these views come together, a perception develops that MWTD encounters many of the problems with resource allocation that have previously been noted in the literature. In particular, the division has gone through a cathartic change in its business within the last two years, yet the resource allocation system has not changed to meet the needs of the new business climate. One particular weakness is that the resource allocation process appears to incorporate only some of the information which is relevant in MWTD's new environment. Beginning in this chapter and continuing in Chapter 5 and 6, I explore in greater depth what information is being missed and of what use and value such information could be to the division.

After reviewing the key points of MWTD's current allocation process, I show that their methods encounter many of the problems previously cited in the literature. Following, I begin to focus more narrowly on the problem of ignoring critical variables. Here I show that while MWTD is aware of many of the variables essential to their current business, they appear to be overlooking a few key ones regarding the value of engaging in process improvement. Finally, I show that the current pattern of investment in process at MWTD, leaves considerable value unrealized and may hinder MWTD's ability to compete long term.

4.1 Summary of MWTD's Current Method of Resource Allocation

The downward trend in MWTD revenue demanded management response at the time of the major organizational redirection in mid-1988. This new direction - of considerable note internally because of the decision to sell MWTD circuits outside of HP, but also ground breaking in the high revenue targets it represents - poses new internal challenges to MWTD. Of importance to process improvement, achievement of the plan for MWTD success in this new direction will require a four times increase in throughput that must be achieved during the next four years, in a fab that is already at or near capacity, with minimal capital investment to increase capacity, and while reducing the current cycle time through this fab by 50%.

In Chapter Three, it was shown that MWTD's resource allocation process is typified by the following scenario:

1) Process improvement project options are originated by the process engineering supervisor on the basis of his perception of where the greatest yield loss is occurring. The supervisor, acting according to his view of the division strategy (as delivered to him via the Hoshin process), recognizes the need to improve yields by a factor of two in the coming year and by another factor of two soon thereafter. Because regular and reliable information on yield loss is not available, project options can respond only to <u>perceptions</u> on yield loss.

2) Costs, benefits and risks are only loosely estimated, usually by the process engineering supervisor or a process engineer. The primary consideration is driven by perceived benefit in terms of increased yield or removal of a production blockade. Cost estimation is confined to capital costs as calculated during an elaborate capital justification phase and educated guesses about engineering manpower required.

3) Capital is acquired through a justification process that uses a team composed of members from all the functional groups affected by the purchase. Over a period of two months, they investigate many aspects of the purchase with the objective of insuring technical success of a capital purchase. While their investigation does make estimates of the financial returns from the project, a low financial return does not appear to affect project approval because most projects are justified based on technical or strategic necessity.

4) Manpower rarely changes to meet the needs of a specific project, outside of the development engineering group. Rather a group supervisor must work within the confines of current staffing levels.

5) Ultimate decision approval on process improvement projects is maintained at a very low level. At MWTD, much of this responsibility rests with the process engineering supervisor.

6) The ultimate criteria for approval of a process improvement project is typically available manpower and perceived yield loss. "Firefighting" on the production line remains a top priority for process engineers. Once this duty is met, the

supervisor divides their remaining time between process improvement and new process introduction based on how he perceives the respective needs.

4.2 Analysis of MWTD's Current Method of Resource Allocation

MWTD's resource allocation process appears to suffer from the same set of problems that have previously been described in the literature. These similarities can be summarized in five points.

1) Because MWTD relies on its first level people (operators, engineers, and supervisors) to generate most process improvement options, it is likely that most of these ideas are well founded in terms of the needs of the production line; however, this system does little to insure that all important options are considered, or even generated. The definition of which projects are sufficiently important to warrant further consideration tends to emphasize improvement of areas that have recently caused highly observable trouble. These areas might be very different than those possessing the highest long term value.

2) MWTD's current system only lightly assess the costs, benefits and risks associated with various project options. In particular, benefit measurement is largely limited to reducing the believed, although not measured, yield loss of a process step and thereby increasing the throughput without additional investment in capital. Other possible benefits, such as increased manufacturing predictability or faster learning, are

not often considered. Meanwhile, costs, in terms of manpower required, are only guessed at.

3) Due to the loose measurements of costs, benefits and risks, financial justifications are realized to be incomplete descriptions of the merits of a project; consequently, such justifications, while routinely done, are not rigidly adhered to; instead, other criteria are relied on. The primary criteria seems to be a benefits assessment of reducing recent, large yield loss mechanisms; however, this is contrasted against a severe manpower limitation. Project staffing for process improvement must largely come from current group staffing, and it remains a second priority to line support activities.

4) Because the current allocation system does not explicitly measure costs, benefits and risks in a manner that allows objective evaluation, MWTD has adapted by pushing decision making power on allocations to a very low level, typically the first line supervisor. This method would seem to allow decisions to be made by one who has experiential knowledge of the various project options that can, to a large degree, compensate for the vague information that has been formally gathered on various options. Indeed, as I shall show in Chapter 5, this may well be the case. However, such a system has two distinct limitations. This decision method relies on the skills, knowledge, and biases of the particular supervisor. It was apparent at MWTD that the switch from one Process Supervisor to another invited a radical change in the projects that were pursued. Second, a particular group supervisor does not have

equal experiential knowledge of other types of allocation options; hence, the supervisor cannot always accurately tradeoff resources between differing types of allocation options, e.g., some times the process engineering supervisor cannot accurately tradeoff process improvement against a variety of new product development options.

5) Because of this last problem, inability to accurately trade off different types of project options, the division and the corporation reserve one tool to manage this tradeoff: allocation of manpower to various groups. But the implementation of this tool is such that manpower is not dynamically allocated, rather it is largely static at current staffing levels. Hence, the organizations dynamic response to new allocation options or competing options is limited.

These five points evidence that MWTD encounters many of the problems outlined in the literature in their attempts to efficiently allocate resources. At this point I begin to focus more narrowly on one central problem addressed above: not observing, measuring, and incorporating certain critical variables - particularly those that describe costs, benefits, and risks - into the allocation decision.

4.3 Ignoring Critical Variables

While the literature suggests that companies rarely incorporate all relevant information in allocating resources, it elaborates that companies that have recently undergone major organizational changes are especially prone to ignore critical variables. I contend only that some critical variables are being ignored by MWTD. Many, if not most, important variables appear to be actively observed and incorporated into management decisions at MWTD. Thus my emphasis here is on seeking to improve what are already generally efficient management methods.

The challenge of understanding more about the information that should be incorporated into process improvement allocation decisions can be sub-divided: what are the critical variables in this new environment in which MWTD finds itself; and, among these variables, which are being paid sufficient attention to and which are being ignored or insufficiently understood. I begin with a consideration of the primary benefits to be derived from improved process control in this manufacturing situation. A consideration of each of these benefits readily show how the environment has changed which variables are important and points to the cost impact of under investing in process control.

4.3.1 Benefits of Improved Process Control

Referring to the section on benefits of improved process control in Chapter Two, there are five benefits to consider: more circuits through a fixed capital base, reduced line support cost, improved customer service and/or reduced finished goods inventory, increased flexibility to deal with changes in product or demand, and faster future 'learning'.

Producing More Circuits with a Fixed Capital Base

Capital costs are by far the most substantial portion of production costs in semiconductor fabrication. (While depreciation charges may not always reflect this, aging equipment should/must constantly be replaced with expensive, current generation systems). Such costs are largely fixed provided there is sufficient capacity to meet demand. Historically, MWTD has maintained excess capacity simply because the production volumes that could be supported by even a single machine of a given type were always far greater than the required throughput. However, internal MWTD studies[31] and Lawton[32] have both argued that the current fab line is approaching capacity due to demand growth. This idea is evidenced by the rapidly increasing backlog for MMIC parts. Furthermore, the planned increases in volume over the coming three years will only aggravate this problem. Thus, given current operation, increases in demand, particularly the large ones expected, can only be met by substantially increasing yield or increasing capacity. Since increasing capacity requires substantial capital investment in new equipment (which both reports have shown to be the primary bottleneck resource in meeting demand), it is -ideally- avoided.

Provided demand is inelastic, yield increases can be valued using the shadow price of increased circuit sales or, alternatively, increased instrument sales. Because yield improvements provide additional parts at little or no incremental cost, the shadow price is just the part selling price; for a typical fab:

Typical MMIC price	\$ 200 /part
Throughput at 4% yield	600 parts /month
Expected benefit	\$ 360,000 /yr /1% yield increase.

Figure 4.1: Expected return from increased capital throughput based on transfer price (typical industry data)

However, this is probably an underestimate of the true return. The MMIC-A circuits sold by MWTD are technical 'levers' for many instrument designs. Without these parts, many instrument designs could not be accomplished. Upon this basis and because of the sizable and growing backlog of parts owed to HP instrument divisions, one can make a second estimate of return based on the average instrument price of instruments that use MMIC-A products.

Typical instrument price	$30,000 /instrument
Instrument throughput at 4% yield	150 instruments/mnth
Expected revenue benefit	$ 13,500,000 /yr /1% yield increase

Figure 4.2: Expected return from increased capital throughput based on instrument price (typical industry data)

Finally, a binary 'cost avoidance' must also be considered. As long as demand can be met with current equipment by increasing yield, MWTD avoids the expense of acquiring additional pieces of equipment, at $100,000 to $1,500,000 per piece of equipment, to increase capacity. Clearly, only bottleneck capacity has to be expanded, but Lawton[33] has shown that the bottleneck is a roving one and hence at least three pieces of equipment would have to be acquired to truly expand capacity. While capacity expansion is certainly a valid means to meet demand in this case, the additional strategic imperative of halving the cycle time suggests that yield improvement is the better alternative.

Equipment needed	3 instruments
average burdened prico	$ 500,000
expected capacity increase	50% over current 600 /mn
Cash outlay required	$ 1,500,000
Cost/equivalent 1% yield increase increase	$ 750,000 /1% yield
Annual Cost (10 yr life, r=10%) before taxes	$ 122,000/yr /1% yield increase

Figure 4.3: Expected capital avoidance from increased capital throughput (typical industry data)

MWTD is keenly aware that there is value to be derived in fabricating more good circuits on a fixed capital base, yet in assessing this value, they emphasize capital purchase avoidance benefits and a measure of additional profit based on circuit transfer prices. While this is a valid concentration on division performance, it ignores contribution to group level profit. As MWTD parts reach higher levels of integration and greater degrees of complexity, as is true with circuits made on the MMIC-A process, the leverage effect on instrument value increases and group level profit becomes a better metric of MWTD performance then transfer prices that are loosely based on the cost of goods sold.

Engineering Line Support Cost

Currently process engineers spend one-third of their available time on supporting the production line. Their activities here focus on responding to out of control or unexpected situations. These situations sometimes result in a "yield crash" where some group of processes produces very low yield wafers. Because lead times are long (1-3 months), this unexpected yield loss throws off the precisely scheduled fab and puts deliveries to customers at risk. The seriousness of these situations demands immediate management intervention. It is quite typical for a group of ten or more people (schedulers, supervisors, and managers) to spend several hours reviewing and rescheduling every lot in the fab. Then many of these same supervisors and managers will spend more time working with the engineers trying to understand what is necessary to recover the process. Thus, line support activities require prompt and considerable attention on the part of engineers, supervisors, and managers. This implies a significant cost to MWTD.

Because of the many process beyond MMIC-A that are supported by a common group, it is difficult to assess the impact of MMIC-A yield improvements on overall support costs, nonetheless, an understanding of the total cost of support provides a useful indicator. It is worth emphasizing that MMIC-A is by far the most variable process and, hence, a large measure of support costs should be attributable to it. For estimation purposes, I assume that half of these support costs are attributable to MMIC-A. It cannot easily be estimated how much this support cost will change for a given increase in yield, yet, as a rough indicator, I assume that support costs are proportional to yield loss. Note that useful time equals 75% of total time because approximately 25% of an engineers activities are spent on overhead activities such as training, communication meetings, filling out paperwork, etc.

# Device & Process Engineers	8 engrs
Annual Cost/Engineer (burdened)	$100,000/yr
Percent of useful time on line support	44%
S/T Engineering support cost	$352,000 /yr

Man hours/week of management line support	100 hrs/wk
Average hourly cost of management	$ 50/hr
S/T management support cost	$250,000 /yr

Total MMIC Support Cost @ 96% Yield loss	$301,000 /yr
Expected Benefit (from MMIC support cost)	$3,100 /yr /1% yield increase

Figure 4.4: Expected cost savings from reducing line support cost (typical industry data)

While management has recently taken note of the large percentage of engineering time spent on line support, thus far they have sought to reduce this figure only by highlighting it to engineers, but without suggestion that better control of the process could help to reduce this figure. There is no evidence that the cost of management support required is at all influencing allocation decisions.

Improved Customer Service and/or Reduced Finished Goods Inventory

One way of dealing with high yield variability is to maintain a large finished goods inventory that buffers the customer from the variations in process yield. Unfortunately, with current growth in customer demand coming right as the fab becomes capacity constrained, MWTD finds itself unable to build any finished goods inventory, and, in fact, has a growing backlog of orders to fill. As shown above, an increase in yield could help to fill these orders. While generating additional profit as calculated above, this yield increase would also build customer goodwill. While difficult to explicitly value, goodwill is certainly one factor used by instrument design engineers in choosing between an MWTD part and one produced by an external vendor. Notably, part delivery was one of only two complaints voiced regularly by MWTD customers.

MWTD clearly recognizes the importance of meeting customer due dates, and looks on yield increase and yield stability as necessary for meeting these dates. These deliveries were not previously a problem because MWTD could just build parts to stock and the inventory holding costs could easily be passed on in the form of higher transfer prices.

Faster Future 'Learning'

Both process improvement and process/product development are based on learning. Bohn has shown that the speed of learning is an underlying cost driver in most manufacturing environments, and has presented a model that suggests that learning speed is highly dependent on environmental noise (as measured by standard deviation/mean). Clearly, the noise level is quite high in this environment. It can be reduced by changing the standard deviation of process output - by better controlling the process. The value of faster learning arises in two ways. Yield improvement within manufacturing can occur at a faster rate, and new circuits can be brought to market sooner.[34]

Yield improvement within manufacturing is largely a function of the clarity of experimental results and the availability of high leverage (large improvements possible) projects. Process yield improvement in many semiconductor fabs has been described as occurring in three phases: process introduction, rapid learning, and process maturity (see Figure 4.5). During process introduction, there are many high leverage improvement project possibilities; however, the choice of which project to pursue and the validation that the change is a useful one are clouded by the extremely high noise level (high process variability). Thus process improvement, in terms of yield increase, tends to be slow during this phase and the justification for investing in process improvement tends to be hidden from immediate view. Over time and if the facility invests enough to work through this slow phase, they reduce noise to a critical level of acceptability. This gives way to a period of rapid learning during which high leverage opportunities are quickly taken advantage of and yield increases rapidly. As the highest leverage opportunities are consumed, growth is maintained by the decreasing noise level and associated increasing experimental clarity. Eventually, the high leverage opportunities are exhausted - despite decreasing noise levels - and yield growth begins to level off as the process reaches maturity.

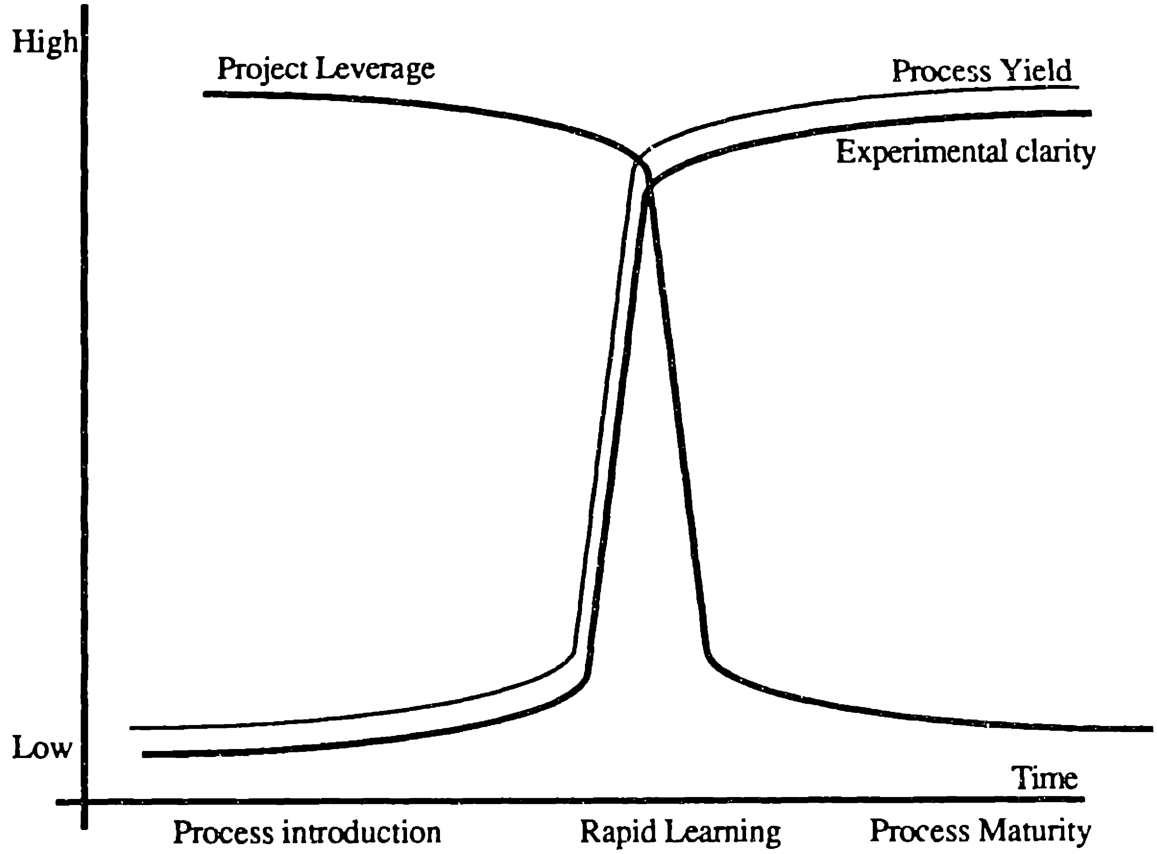


Figure 4.5: Project Leverage and Experimental Clarity over time

Based on this model, foresight suggests investing to reach the period of rapid learning more quickly than the "average" pattern of investments (see Figure 4.6).

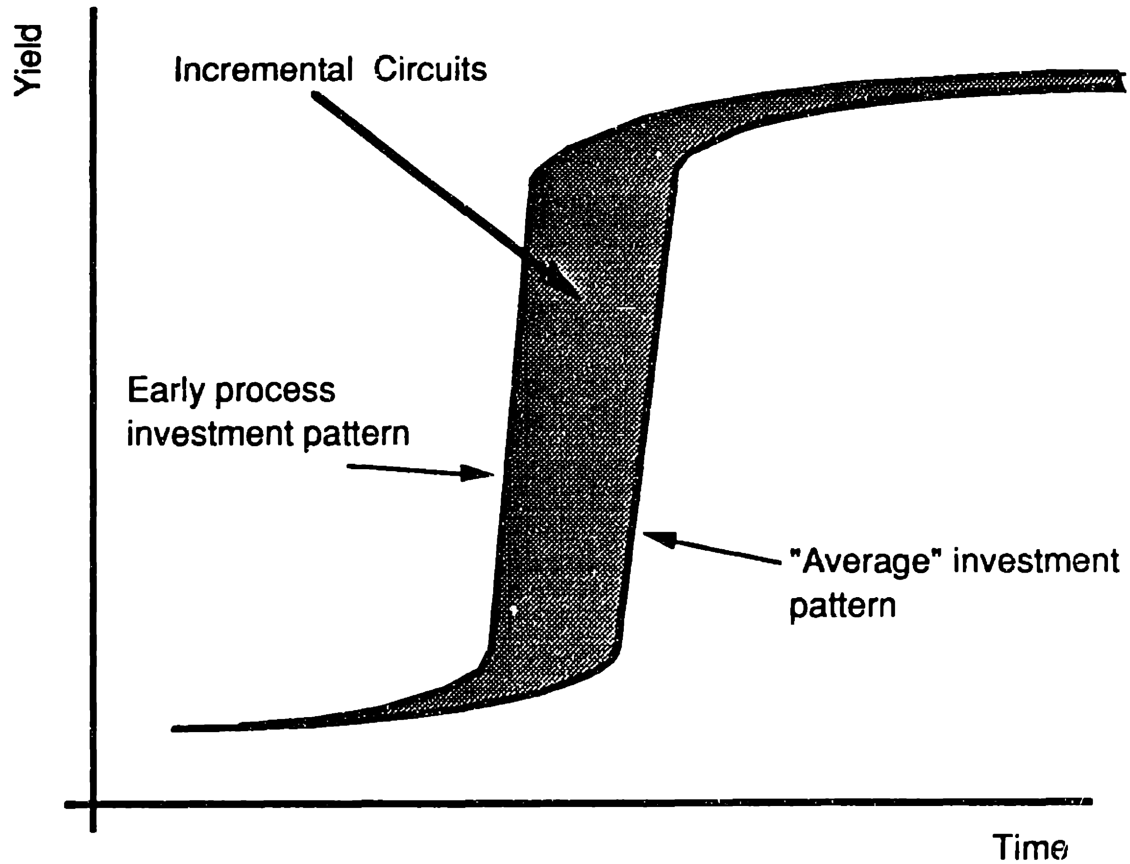


Figure 4.6: Incremental circuit production due to early investment in process improvement

MWTD's current situation, with wide variations in process measures such as gate length, suggests that they are still well within the "process introduction" phase, yet the rate of learning, or, by analogy, the rate of increase in mean yield, will increase over its current level of 0.02% per wafer run (see Figure 4.8), and long term advantage will be gained by increasing process investment now to get through the process introduction phase and capture the incremental output. Beyond the obvious cost savings this brings due to additional parts being available for sale, the increased yield can also contribute by reducing time to market.

While I argued earlier that cost savings could be garnered by increasing yield to meet rising demand without increasing the capital base, I now argue that improved yield and increased yield predictability can actually increase overall demand by helping new circuits, and thus new instruments, be brought to market faster and thus gain a greater market share. Earlier product introduction might also allow more monopolistic pricing, further increasing revenue.

Circuit time to market was seen to be highly dependent on the length of time required to fabricate a test device and the information quality from each such test run. In turn, the length of time to fabricate a test device is determined by the fab cycle time. Process improvement can affect cycle time in three ways:

1) Increases in yield allow less wafers to be started to meet a constant output of finished product die. Little's Law shows that this results in a decrease in cycle time. Although, to avoid double counting, an increase in yield can produce only one of the two benefits discussed: either increased die throughput with constant wafer throughput as discussed under the benefit of producing more circuits with a fixed capital base; or, reduced cycle time due to reducing wafer throughput while holding die throughput constant (or some combination of both as allowed by Little's Law). (See Figure 4.7).

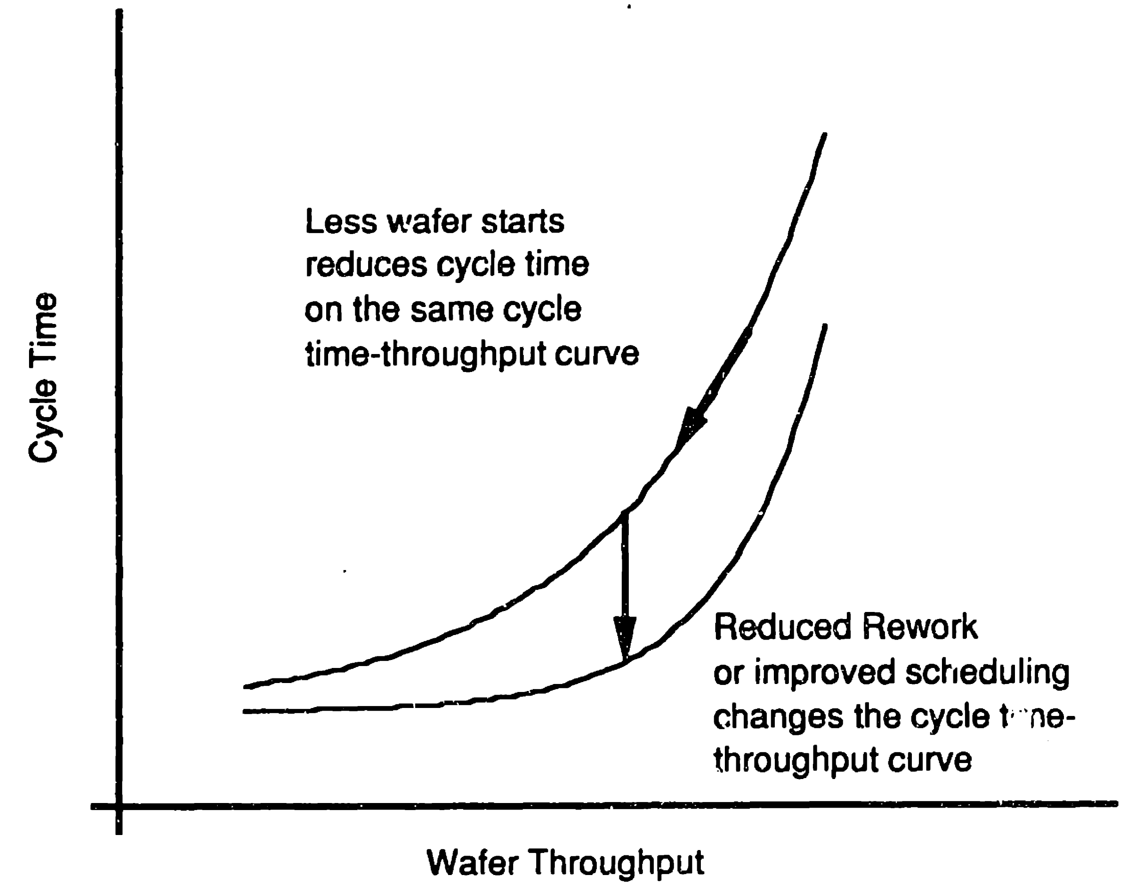


Figure 4.7: Little's Law and improvements in cycle time

2) Increased yield reduces rework which further diminishes the load on production equipment, enhancing this decrease in cycle time. (See Figure 4.7).

3) Increases in predictability of yield allow more deterministic scheduling of wafers in the fab, again reducing cycle time by preventing unexpected roving bottlenecks. (See Figure 4.7).

The information quality of a test run is largely a function of process variability and can be enhanced - as mentioned above, for learning within manufacturing.

Lawton has suggested that a reduction in MWTD fab cycle time can have a dramatic effect in improving circuit time to market and instrument time to market with a resulting HP revenue enhancement. While Lawton estimates a revenue increase of some $120-$250 million for a 50% reduction in the cycle time of the entire fab, this work focuses on reducing MMIC-A cycle time only and thus the effect should be somewhat muted.

Increased Flexibility to Deal with Changes in Product or Demand

As mentioned before, the new competitive arena for microwave instruments is demanding faster delivery of new products. This requires that MWTD be able to bring new products into production quickly and rapidly ramp to mature volume. This demands flexibility in MWTD's manufacturing system. By maintaining high yield predictability, MWTD is better equipped to respond to requested changes in a deterministic manner. While such an ability is too difficult to assess quantitatively, it clearly makes an organization more competitive. Because the value of flexibility has only arisen with the change in business environment, MWTD is still adapting and has yet to fully consider how process improvement can create flexibility to deal with changes in product or demand.

These various benefits can be summarized in a single table along with the drivers of this benefit. A comparison of which of these drivers were important in the old operational environment with those that are now important shows how the critical variables have changed. Furthermore, noting which of these variables MWTD pays attention to shows where critical variables are being ignored.

Benefit	Driver	Imprtnt. Before	Imprtnt. Now	Paying Attentn.
Capital Deprec.	Circuit cost	no	yes	yes
	Instrument cost	no	yes	no
	Capital Avoidan.	no	yes	yes
Line Support	Engineer time	yes	yes	some
	Managemt. time	yes	yes	no
Customer Serv.	Custmr. goodwill	yes	yes	yes
Learning	Faster Ramp	no	yes	no
	Time to market	no	yes	no
	- Cycle time	no	yes	no
	- Good info	no	yes	no
Flexibility	Demand changes	no	yes	some
	Product changes	no	yes	some

Figure 4.8: Summary of benefits of process control and whether they are paid attention

Thus, improved process control, as measured by yield and yield variance, can provide a strong contribution to MWTD in a wide variety of ways. An explicit assessment of this contribution is difficult for some of the benefits. While I have presented indications of the magnitude of these values, and - where convenient - considered the dependance of this benefit on a specific increase in yield, a complete benefits per yield improvement calculation is left to future work.

It is apparent from Figure 4.8 that while MWTD has adapted somewhat to the change in environment, they appear to not be paying attention to all of the benefits from improved process control. Before trying to better understand how yield can be improved in Chapter Five, it is enlightening to briefly examine the effects - beyond the missed benefits just outlined - of the current pattern of allocations to process improvement at MWTD.

4.3.2 Effects of the Current Method of Resource Allocation

As expressed above, the goals of an effective system for allocating resources to process control projects rest on providing a desired, predictable yield from the production process and providing meaningful estimates of the improvement in yield that can be expected from a given investment of additional resources in process controlling activities. Here I address the question of whether the allocation system at MWTD, laid out in the preceding chapter, has accomplished these goals.

Looking at the final yield of the MMIC-A process, of that half of the wafers that actually make it to final test, figure 4.9 shows a high variability in final yield. A five wafer moving average of this data shows a definitive upward trend in yield, but predictability of yield for a given wafer has not come with this upward trend in average yield. Instead the wafer-wafer yield continues to vary on a magnitude comparable to the average yield.

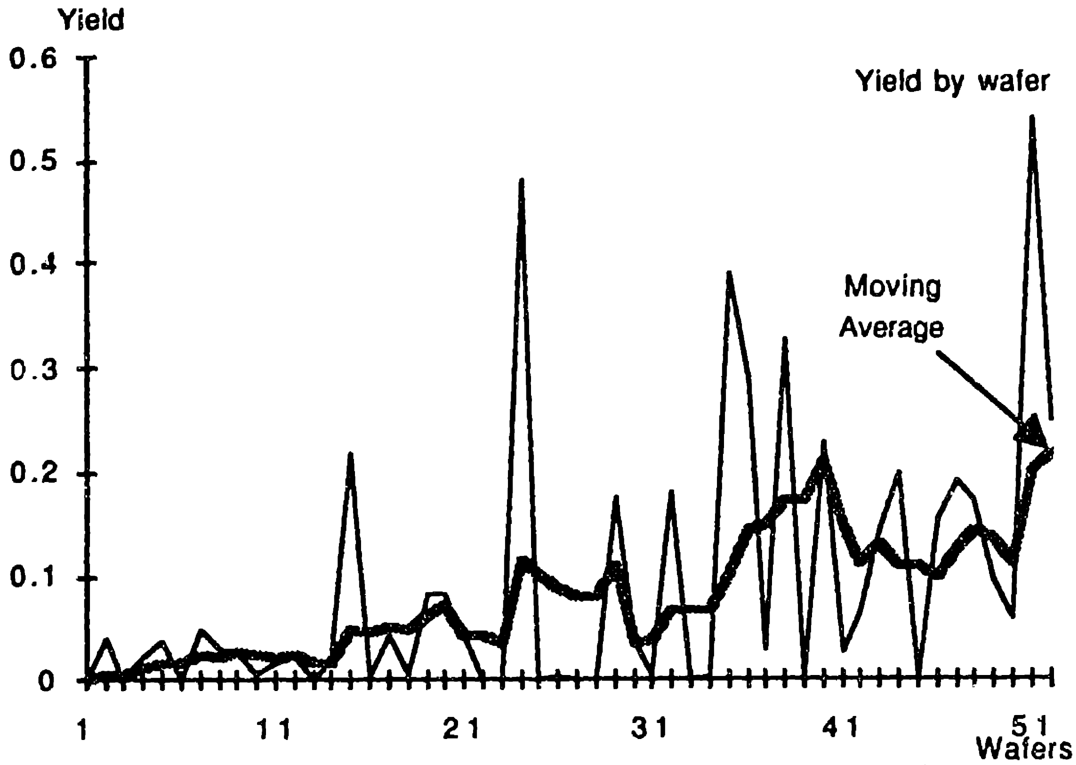


Figure 4.9: Wafer Yield at Final Test over Time (Data disguised, but proportional)

Turning to the goal of providing meaningful estimates of the improvement in yield that can be expected from a given investment of additional resources, one can certainly observe that overall yield is trending upwards at a rate of approximately 0.2% yield per wafer started. Presumably this increase is a direct result of the process targeting and controlling activities already undertaken at MWTD. The mere observance of upward trending yield might be a sufficiently acceptable result in many manufacturing environments.

However, a problem arises when one wishes to consider whether to invest more or less resources to affect the current rate of change in yield. If one more engineer or a million dollars in capital could be allocated to improving the process, by how much more could the yield be increased? Given this yield increase per amount of resources invested, should the engineer or capital be used to improve control, or might they bring a better return elsewhere in the organization? Such questions are implicitly answered by the current allocation process at MWTD. Capital budgeting and capital requisition processes provide capital at levels that approximate previous investments in that part of the company, and manpower is allocated at rates equal to prior staffing levels. There is little evidence that these responses represent efficient levels of investment.

Finally, in the current system, there is nothing to suggest that the process improvement projects that were invested in provided superior return to the other possible process improvements. All that can be stated conclusively is that considerable gains are possible through process improvement that focuses on a finite set of variability causes identified in Chapter Five, but these gains are not, generally, pursued at MWTD.

In summary, with such variation as now exists at MWTD,

1) Management cannot plan reliable deliveries to customer divisions without maintaining a considerable finished products inventory which increases product cost, or, in the current capacity constrained environment, deliveries cannot be readily met and backlog increases rapidly.
2) The flexibility of the production system to respond to changes in volume or product type is diminished.
3) Management cannot invest in process improvement with a good estimate of resulting yield increase. When a set yield increase is central to the division's strategic plan, such lack of knowledge places achievement of strategic goals at considerable risk.
4) Process improvement projects that are worked on may not imply an efficient use of resources: the benefits derived may be lower than other allocation options.

It is difficult to show conclusively that the current <u>total</u> pattern of resource allocations is either inefficient in maximizing return or insufficient to meet the strategic goals of the organization. Time will answer these questions well enough. The more limited but also interesting challenge possible here is to better understand the critical variables that should be measured and incorporated to improve upon the current, very human and subjective system for allocating resources efficiently.

As I will suggest in Chapter Six, the measurement and incorporation of these critical variables would tip the balance more in favor of process improvement than current practice. However, it is important here to recall the conclusions of Chapter 3. The very real inefficiencies in the resource allocation system at MWTD, inefficiencies that are present in many/most industrial allocation systems, might very well maintain the current pattern of allocations even if clear information favoring investment in process improvement could be shown. Thus, this information while it may be a necessary prerequisite to change is certainly not a sufficient one.

Chapter 5: Variability in a GaAs MMIC Production Process

In Chapter Two, I presented conceptual guidelines for identifying the causes of process variability (repeated in Figure 5.1). In this chapter, I apply those guidelines to a particular traveling-wave amplifier circuit made on the MMIC-A process. I begin in section 5.1 by considering what will be necessary to apply these guidelines to the MMIC-A process. I continue in section 5.2 with a brief overview of the experimental approach and results. In section 5.3, I cover the translations from customer attributes to device physical parameters. In section 5.4, I share the majority of the experimental work done in this research. Included here is a description of the gate fabrication process; the experimental approach for and results from measuring the effect of process input variation on device physical parameters; the experimental approach for and results from measuring the variability of process inputs; and the results of combining the input variabilities with input effects to estimate resulting variability in a critical in-process control monitor, PMMA line length. Finally, in section 5.5, I consider the implications of these results. I consolidate the data into identifications of the process causes of variation in customer attributes and the resultant yield loss and examine how this information might influence the resource allocation decision.

5.1 Applying the Conceptual Guidelines to the MMIC-A Process

While conceptually simple, the guidelines for identifying the causes of process variability, summarized in Figure 5.1, encounter several substantial problems when applied in the MWTD manufacturing environment.

1) Understand precisely the customer's needs
2) Translate these to the process inputs necessary to meet these customer requirements.
3) Measure the input variability in all or some important subset of these process inputs.
4) Multiply the input variability by the effect of such variability on customer desired parameters, found in step 2, to determine the relative contribution to variations in meeting customer desires.
5) Use these contributions, together with a knowledge of the customer specifications, to calculate the effect of variation in a particular process input on final customer yield of good product. Rank order these effects in terms of maximum variability in customer requirements.

Figure 5.1: Conceptual guidelines for identifying the causes of process variability

MWTD is generally successful at accomplishing the first of these guidelines. Because the circuits that MWTD sells are a technically enabling feature in the design of HP's microwave instruments, circuit and process designers in MWTD tend to have a strong knowledge of their customer's needs and desires; this is atypical of many, if not most, manufacturing concerns.[35]

However, in accomplishing the second guideline, translating these customer requirements into the necessary process inputs, one encounters the largest challenge of finding the process causes of variability in this environment. The translation cannot be done analytically because the device physics involved is not fully understood. Even models that incorporate second order effects are not sufficient to explain the relationship involved. Turning to the standard alternative of experimentation offers only mediocre aid in this environment: the number of process inputs (over 1000) is simply too great when the environmental noise (one gets a sense of the noise level from the facts that overall process yield is below 10% and variability is spread throughout the fabrication process) and the resulting cost of experimentation are considered (typical costs in a GaAs environment are $200 - $1000 for the wafer coming into the fab and several thousand dollars for a finished, tested wafer; furthermore, turnaround times on finished wafer experiments are one to three months). These conditions recommend against direct experimentation of the effects of variation in process inputs on producing the desired customer electrical specifications.

These same experimentation effects hamper measurement of variability in process inputs, although to a lesser degree because the experiment is free from the noise induced in subsequent process steps. Additionally, for many inputs, such as resist sensitivity or exposure dose delivered, MWTD is not currently equipped to measure these values in a convenient and accurate manner. Further complicating the experimental option, as the MMIC-A process presses the limits of current IC technology, the accuracy and repeatability of the measurement equipment used to gather data are no longer negligible. Additional consideration of translating customer requirements into process specifications is clearly necessary.

One method that has been suggested is to use a design tool known as the "house of quality", which is part of a management approach known as quality function deployment (QFD).[36] This method is deceptively simple in principle, suggesting that the customer's desires can be backed into the production requirements by making smaller causal links to several intermediate points of information. Exhibited generically, the QFD approach would appear as in Figure 5.2.

```
Customer   ---->
Attributes
           Engineering      ---->
           Characteristics
                         Parts           ---->
                         Characteristics
                                      Key process   ---->
                                      Operations
                                                  Production
                                                  Requirements
```

Figure 5.2: Generic "House of Quality" causal links

QFD rests on the notion that sufficient analytical or experimental work has been or can be done at each of several steps to get around the inability to make the large experimental or analytical leap directly from customer attributes to production requirements.

QFD is in line with an experimental decomposition approach where large problems are broken into several connected, but smaller ones. Presumably each of the smaller problems can be solved more easily, with less experimentation and with less noise in the experiments. Then an overall answer can be assembled by combining the findings. However, both QFD and experimental decomposition rely on understanding the effects taking place at each stage. These effects are rooted within the circuit, device and process physics involved. As applied to a GaAs process, such as HP's MMIC-A, the QFD translation would appear as in figure 5.3.

Customer
Specificat. --->
Circuit
Elect. Perf. --->
Device
Elect. Perf. --->
Device
Physical --->
Process
Inputs

Figure 5.3: "House of Quality" causal links for the MMIC-A process

To understand how in-process variability affects final customer performance, one need only consider the translation at each step on either an analytical or experimental basis and mesh these individual relationships to arrive at the desired customer-to-process relationship. This idea served as the overall model, in this study, for determining the effects of in-process variability on variation in customer attributes. Before I begin applying this QFD approach in section 5.3, I briefly overview the experimental approach and key results.

5.2 Overview of Experimental Approach and Results

An experimental plan was set up to study the process causes of variability in the two key customer attributes of a particular traveling-wave amplifier circuit made on the MMIC-A process: Gain slope and gain magnitude, with control of gain slope being the particular emphasis of this research. It is shown that gain slope is largely determined by the device parameter, input capacitance (C_{in}) which, in turn, is a strong function of gate length (L_g). Because actual gate length cannot be conveniently measured during the process, a related feature of the topography, transfer layer opening or transfer length (L_t), is relied on as an in-process control monitor. This research shows that variability in L_t is set by three process steps: PMMA spin/bake, align/expose, and PMMA develop. Twenty-two inputs to these three process steps are identified as causing variability in L_t and three types of variability are studied for each input: die-to-die, wafer-to-wafer, and batch-to-batch. Out of all the inputs, die-to-die variation is shown to be dominated by variation in intra-mask CDs, variation in developer temperature and the presence of particles on the wafer during contact gate lithography. Wafer-to-wafer variation is found to be dominated by developer exhaustion and variation in developer temperature. Batch-to-batch variation, though nominally present due to many inputs, is presumed to be largely "controlled away" by targeting develop time for each batch; however, it is suggested that the targeting process itself introduces substantial batch-to-batch variation.

5.3 From Customer Attributes to Process Parameters

The sequence of translations that connect customer attributes to process inputs are outlined in Figure 5.4, together with the key variables used as metrics at each step and the method used to accomplish the particular translation. Also included is whether existing process data or new experimental data was used. "Existing process data" refers to the considerable process data, of many types, that is kept by MWTD for the most recent twelve months. Whenever possible, existing data was used to reduce the cost of obtaining information.

Step	Key Variables	Method of translating to prior step
Customer Attributes	**Gain Slope** Gain	
Circuit Electrical Perform.	**Gain Slope** Gain	Incorporated by MWTD during design Assumed perfect translation
Device Electrical Perform.	**Capacitance** Transconductance Current R_g R_s Breakdown Voltage	Experimental Correlation Existing data sets
Device Physical Parameter	Gate Length **Transfer Length**	Analytical Experimental Correlation Existing data sets
Process Inputs	**Exposure** **Develop** **Mask** **PMMA Spin**	Matrix Experiments New data sets

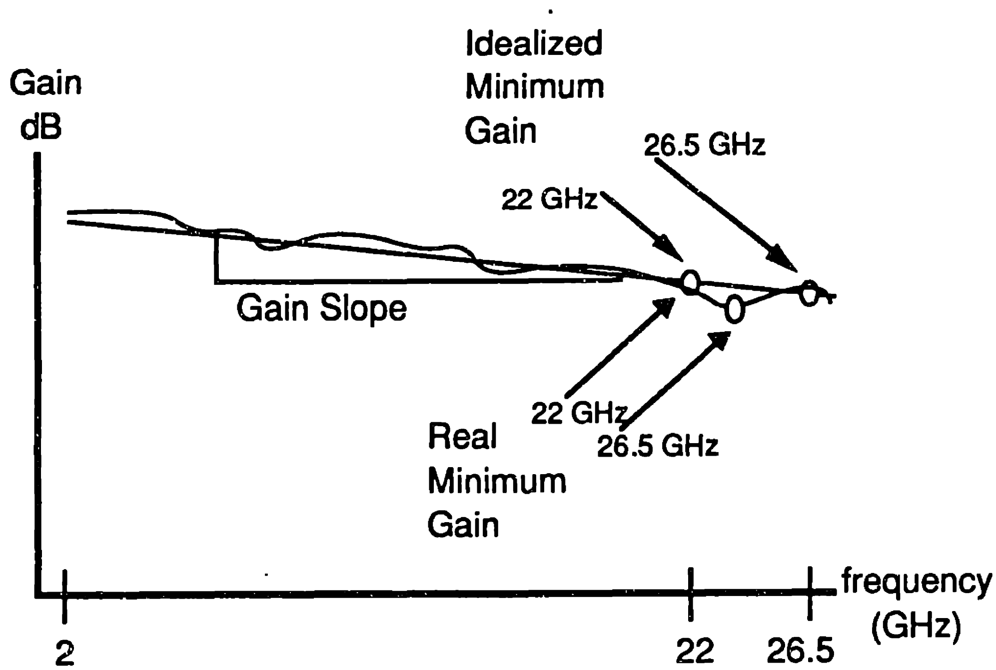
Figure 5.4: Translations from customer attributes to process inputs

Each of the translations, shown in figure 5.4, is discussed together with a consideration of the method used. The final step in these translations, the effect of process inputs on device physical parameters is considered in section 5.4, along with variability in process inputs, to allow for a fuller presentation of the experimental work done.

5.3.1 Customer Requirements to Circuit Electrical Performance

As mentioned previously, MWTD has a strong sense of customer needs in terms of electrical specifications. The strength of this bond originates from the internal supplier nature of MWTD. The circuits that MWTD supplies to the instrument divisions are one of the key determinants of instrument performance. This encourages the customers to be actively interested in the electrical performance of MWTD's products. Thus far, customers divisions have been able to "negotiate" for tighter gain slope specs from MWTD. The fact that two of the primary customer divisions are located at the same site or very nearby helps strengthen this connection. It is also quite common for customer divisions and MWTD to hold joint technology reviews. Interviews with customer engineers suggest that MMIC-A products fully meet customer requirements in all areas of electrical performance. The only areas of dissatisfaction expressed by the customer divisions were the development of long term reliability problems and the inability of MWTD to consistently meet demand in a timely manner. As discussed in Chapter 4, in the current capacity-constrained environment, the ability to reliably meet demand will largely be a function of yield and yield variance, the causes of which are studied in this research. While circuit reliability is undeniably a key "customer requirement", I confine myself to the goal of meeting desired electrical performance when a part is first delivered to the customer.

In considering precisely which parameters are important to customers, I focused on a particular circuit produced on the MMIC-A process. One such product is a 2-26.5 GHz traveling-wave amplifier. This is a leading edge circuit destined for use in many HP instruments. It is believed to be the only production device in the world capable of delivering high, flat gain over such a broad frequency band.[37] The electrical parameter of primary interest to customers on this product are gain magnitude and gain slope. (Refer to Figure 5.5). The product is really a family of four products, differentiated by the frequency band they cover, either 2 - 22 GHz or 2 - 26.5 GHz, and the maximum allowable gain slope.



The straight line is the idealized Gain Slope, while the jagged one is the actual gain vs frequency plot showing ripple in the gain slope.

Figure 5.5: Gain vs. Frequency for a traveling-wave amplifier. Definitions of Gain Slope

For this traveling wave amplifier, gain and gain slope are defined as follows,

Gain = Gain (at 22 GHz)
Gain (at 26.5 GHz)

$$\text{Gain Slope} = \frac{\text{Maximum Gain - Minimum Gain}}{\text{Maximum rated frequency - Minimum rated freq.}}$$

$$= 0.050(\text{Maximum Gain - Minimum Gain}) \text{ for 22 GHz}$$
$$= 0.041(\text{Maximum Gain - Minimum Gain}) \text{ for 26.5 GHz}$$

where the maximum gain and minimum gain are measured over the applicable frequency range. Nominally, maximum gain occurs at 2 GHz and minimum gain at 22 or 26.5 GHz. However, a pole frequency often creates a minimum frequency before reaching the high end rated frequency, as is the case for the 2-26.5 GHz trace shown in Figure 5.5.

These are but two of some twenty electrical parameters of interest to the customer, yet it is commonly accepted that these two parameters do the most to leverage the technical advantages of an instrument design and are the most difficult for MWTD to deliver. Hence, electrical yield loss is mainly determined by these two parameters, accounting for over half of the 27.5% of die lost to parametric yield loss.

Because of the general strength of the MWTD-customer connection and the absence of any performance based customer problems, I assume that the translation between customer desired attributes and circuit electrical performance is a perfect one. In other words, any circuit meeting MWTD's specifications for this amplifier will fully satisfy the customer's needs for electrical performance of the product. Because of the perfect and direct translation that I have assumed, the parameters of importance for circuit electrical performance are the same: gain and gain slope.

5.3.2 Circuit Electrical Performance to Device Electrical Performance

Orr at MWTD has shown that the two key circuit parameters discussed above are primarily dependent on only two device parameters of the MESFET used to fabricate the amplifier circuit: transconductance (G_m) and input capacitance (C_{in}).[38]

Referring to Figure 5.6,[39] by definition,

$$\text{Transconductance} = G_m = \frac{\partial I_d}{\partial V_{sg}} \cong \frac{\varepsilon v_{sat} w_g}{d} \tag{5.1}$$

C_{in} is approximately equal to C_{gs} (gate-source capacitance) plus C_{gd} (gate-drain capacitance); however, C_{gs} typically dominates, so

$$C_{in} = \frac{\partial Q_g}{\partial V_{sg}} \cong \frac{\varepsilon L_g w_g}{d} \tag{5.2}$$

where

ε	=	permittivity of the active region
v_{sat}	=	saturation velocity
w_g	=	gate width
d	=	depletion depth
L_g	=	gate length

and assuming that the device is operated in the velocity saturation regime.

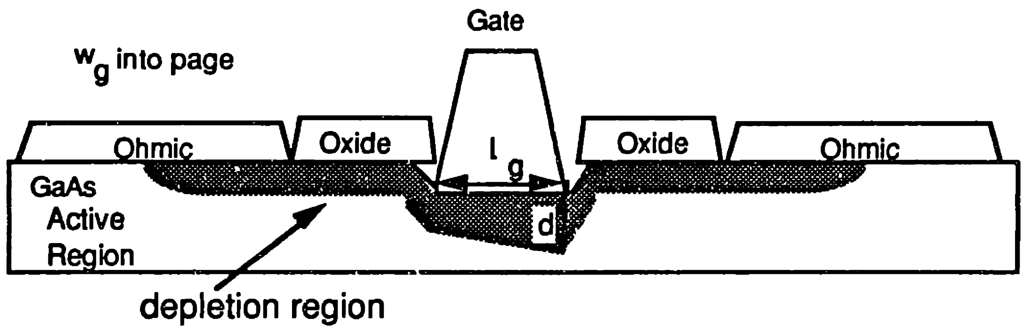


Figure 5.6: MESFET side profile[40]

Dependence of Gain Slope:

In particular, gain slope is set by roll off of gain at high frequency which is attributed by Orr to the $C_{in}*R_{load}$ pole frequency.[41] After appropriate inductive matching in the circuit design, the circuit parameter R_{load} is dominated by the device parameter R_{in}, which equals source resistance (R_s) plus gate resistance (R_g) but is dominated by source resistance. A review of process data shows that R_{in} does not vary significantly with gate length[42]; Hence, gain slope is largely dependent on C_{in}. A linear correlation between Gain Slope and C_{in} agrees well with this conclusion. Orr found:[43]

$$\text{Gain Slope (dB/GHz)} = 0.225 - 0.2182 * C_{in} \text{ (pF/mm)}$$

$$\text{Mult-R} = 0.786 \qquad 5.4$$

Dependence of Gain:

Amplifier stage gain equals $G_m * R_{load}$ when the device is operated in the saturated regime. With R_{load} largely constant as stated above, circuit gain is set by G_m. A linear correlation between circuit gain and transconductance agrees well with this conclusion. Orr found:[44]

$$\text{Gain} = \text{Mag } S_{21} \text{ (dB)} = 0.57 + 0.0142 * G_m \text{ (mS/mm)}$$
$$\text{Mult-R} = 0.839 \qquad 5.5$$

It is important to realize that the active device in the circuit under consideration, a 600μm gate width MESFET, is located within the circuit and it cannot be probed (electrically measured) by itself. Because of this problem, circuit designers include a test pattern on the same mask that is used to print the circuit. Within this test pattern are a variety of devices designed to reveal information about the circuit, about devices in the circuit, or about performance of the process. A 600 μm gate width discrete MESFET, in fact, was designed to be as similar in performance as possible to the actual MESFET in the circuit. Thus, these correlations relate performance of a test device to performance of a physically nearby, but separate circuit built using MESFETs with similar layouts. Finally, it should be noted that data was screened to remove 'outliers' before running the above correlations.[45]

5.3.3 Device Electrical Performance to Device Physical Parameters

While the correlations above are useful in terms of showing a causal link between the device under study and the circuit being fabricated from this device, they suffer from the disadvantage of only adding information late in the process. Because the 600μm MESFET is a multi-fingered FET it requires processing through the "second metal" mask step (see Figure 5.7) before the device is completely fabricated and can be tested.

The delayed arrival of information from testing a completed device, relative to the fabrication of the gate, makes it useful to find a measurable attribute of the device that can be tested soon after the sensitive part of the device is fabricated. (Refer to Section 5.4.1 for a description of the MMIC-A process). While appropriately designed test patterns can be and are used as in-process monitors just after gate fabrication, they suffer from necessarily approximating the device under study rather than being an identical device or layout. An alternative is to consider the physical parameters of the device that determine electrical performance and use one or more of these as an in-process monitor.

Device theory suggests that C_{in} will be a function of gate length (L_g), gate width (W_g), and depletion depth (d) (see Eqn. 5.2). The gate width of these devices is 600 μm (600 μm total width over five gate 'fingers') and this dimension is easily controllable within less than 1 μm over this length. The depletion depth is intimately tied to the channel doping which is comparatively easy to control with the Molecular Beam Epitaxy process used to grow the active layer of the device, and permittivity is a property of the semiconducting material and does not vary. Gate length is approximately 0.42 μm and can be easily controlled only to within one standard deviation of 0.05 μm. Thus, L_g emerges as the critical parameter in need of tight control and would be an ideal in-process monitor of proper MESFET construction.

Unfortunately, the actual gate length cannot be observed visually on an in-process basis because of the recessed contact area to the active region (see Figure 5.8.7). Accurate visual observation would be possible only by destructively cleaving the wafer and measuring a side profile of the gate. While Azzam and del Alamo have suggested[46] in-process electrical measures of this gate length, such measures are not in use at MWTD and are not quick to implement.

The need for an in-process measure has forced MWTD to rely on the imperfect but closely related value of the transfer layer opening (L_t) (see Figure 5.8.3). Immediately after measurement of this opening, the gate metal is evaporated into the opening, after which the surrounding structure is lifted off by developing the polyimide leaving the finished gate structure. While process theory suggests that L_t should have a strong correlation to L_g, this relation has not been shown conclusively at MWTD. Differences might arise from non-perpendicular arrival of the evaporated gate metal. Either too large an aperture in the evaporator or device profile sidewalls that are too sloped can allow non-perpendicular metal arrival. Despite these imperfections, L_t is currently the best available in-process monitor of proper gate fabrication.

Given the discussion above of L_g, L_t, and C_{in}, a strong relationship between L_t and C_{in} would also be expected; however, Orr's correlation of C_{in} and L_t is weak:

$$C_{in}\ (\text{pF/mm}) = 0.51 + 1.6014 * L_t\ (\mu\text{m})$$

$$\text{Mult-R} = 0.342 \qquad 5.6$$

A similar, although stronger, correlation directly to the gain slope was also observed by Orr:

$$\text{Gain Slope (dB/GHz)} = 0.32 - 0.8978 * L_t\ (\mu\text{m})$$

$$\text{Mult-R} = 0.635 \qquad 5.7$$

These correlations of L_t suggest a mean target value of 0.42 µm with a maximum acceptable transfer layer length of 0.45 µm. The minimum gate transfer length is set by a different parameter, pinch off voltage (V_{po}). The device engineers felt that values of L_t less than 0.35 µm would result in wafers having V_{po} problems.

A similar methodology could be used to relate G_m to its relevant parameters, but that challenge is left to future research.

5.4 Experimental Setup

Having established that control of gate length - or, more specifically, transfer length - is of prime importance in controlling the circuit parameters desired by the customer, I turn to the issue of how the fabrication process for these devices actually works to produce these electrical parameters by creating devices, particularly MESFETs, with a distinct set of physical attributes, most notable among these being L_g.

5.4.1 The MMIC-A Gate Fabrication Process[47]

The MMIC-A process requires the eleven masking levels outlined in Figure 5.7. The critical gate region, however, is fully formed after the fourth masking layer and changes little during subsequent processing. Initial processing begins by growing a doped GaAs active region on a semi-insulating GaAs substrate. Oxide is deposited by a CVD process to provide field passivation. Initial masking steps pattern the oxide for deposition of ohmic contacts, allow for proton isolation of active devices and pattern the oxide for sputter deposition of a thin film resistor. At this point, wafers begin the critical processes that lead to gate formation.

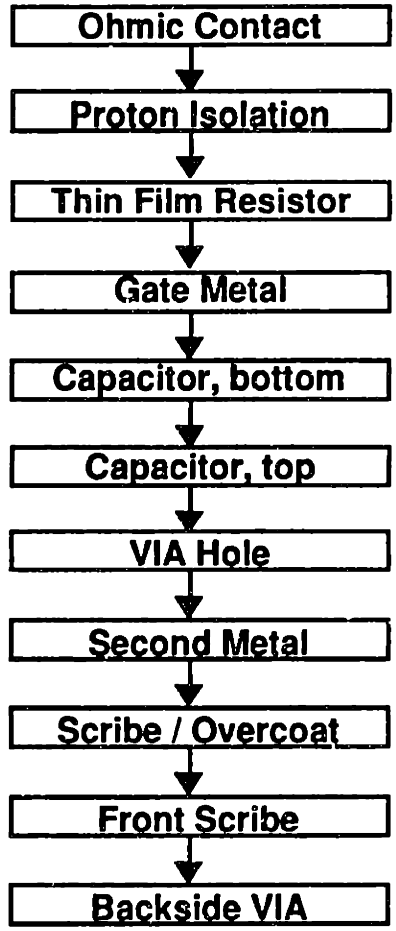


Figure 5.7: MMIC-A process flow

Gate processing begins by spinning on a 0.6 μm polyimide layer to planarize the wafer surface (other surface features are ~0.2 μm off the active layer) and to provide a lifting medium for the transfer layer after the gate metal is deposited. After the polyimide is baked to provide stabilization, a 0.5 μm PMMA layer, is spun on to act as an imaging resist (see Figure 5.8.1).

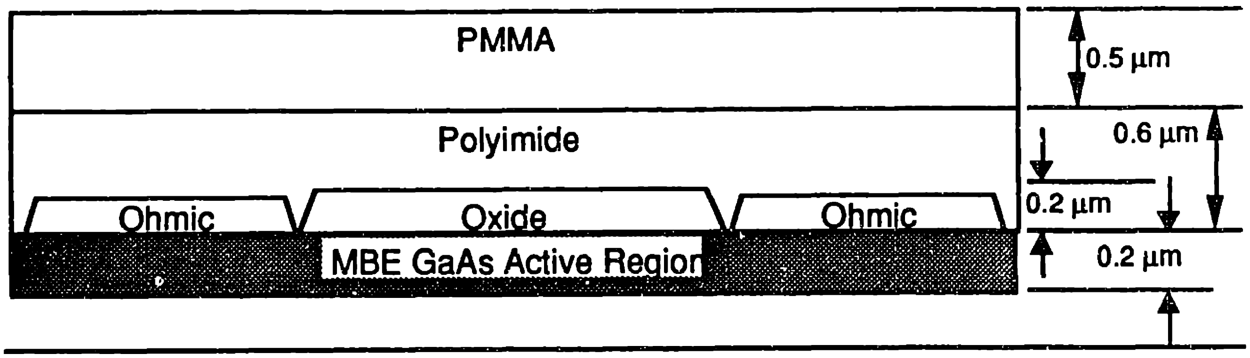


Figure 5.8.1: Gate region profile after spinning on Polyimide and PMMA

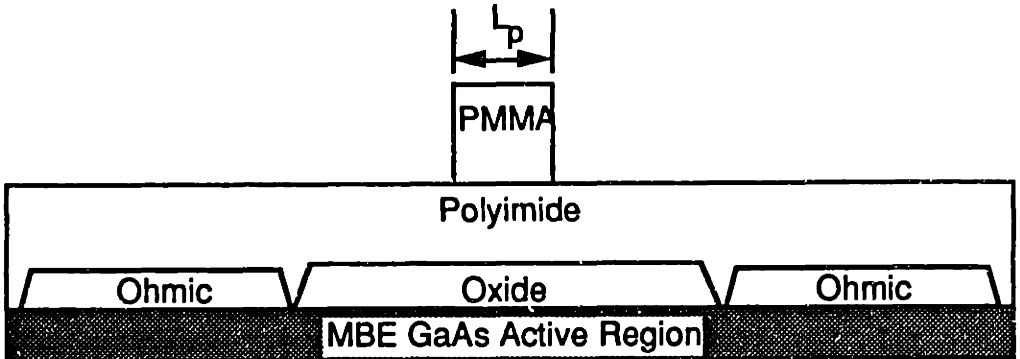


Figure 5.8.2: Gate region profile after developing PMMA

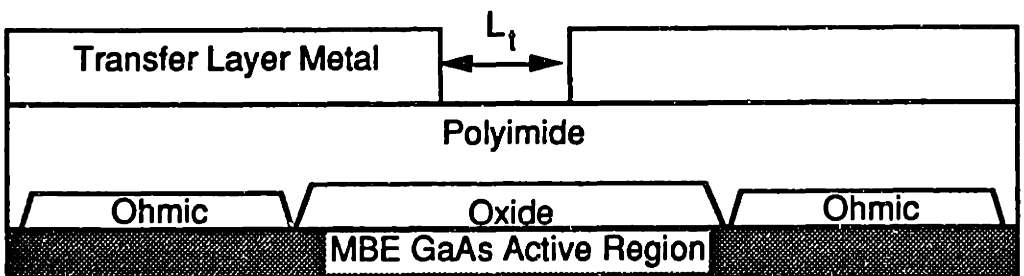


Figure 5.8.3: Gate region profile after evaporating transfer layer and lifting PMMA

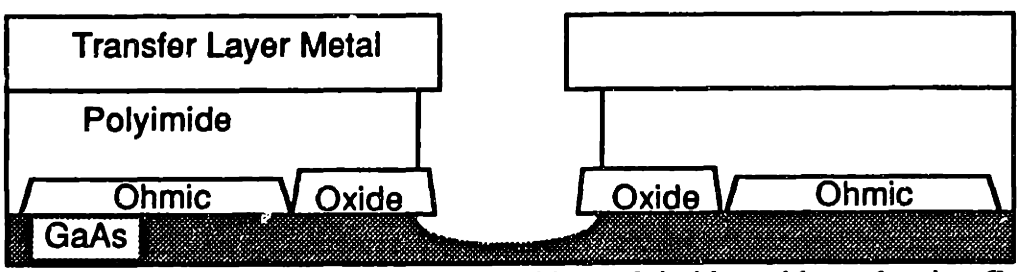


Figure 5.8.4: Gate region profile after etching polyimide, oxide, and active-GaAs

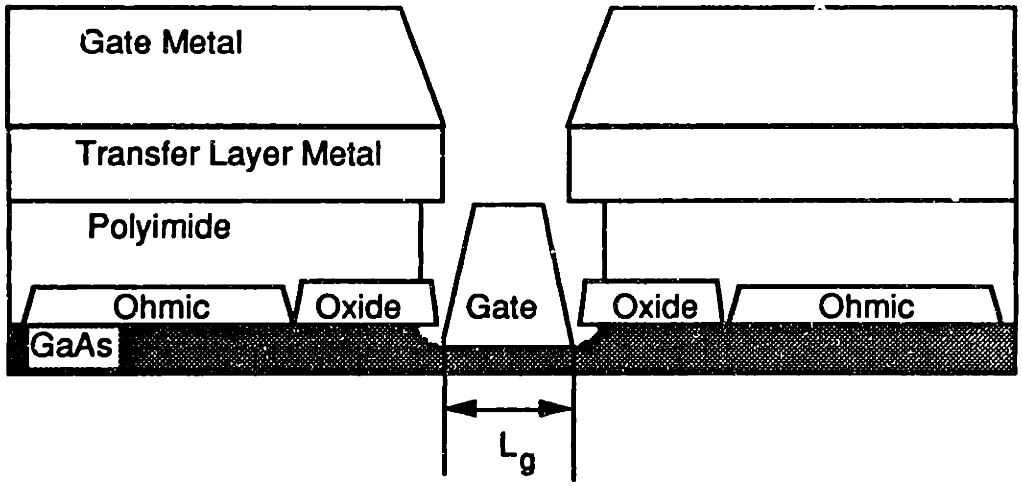


Figure 5.8.5: Gate region profile after evaporating gate metal

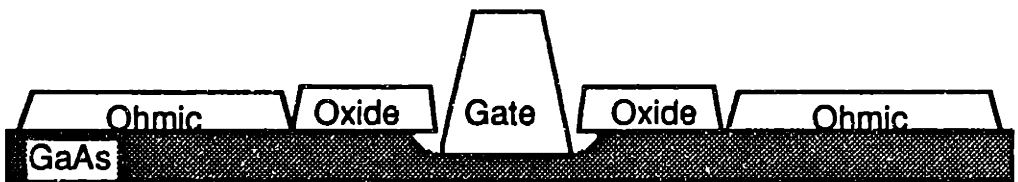


Figure 5.8.6: Gate region profile after lifting polyimide

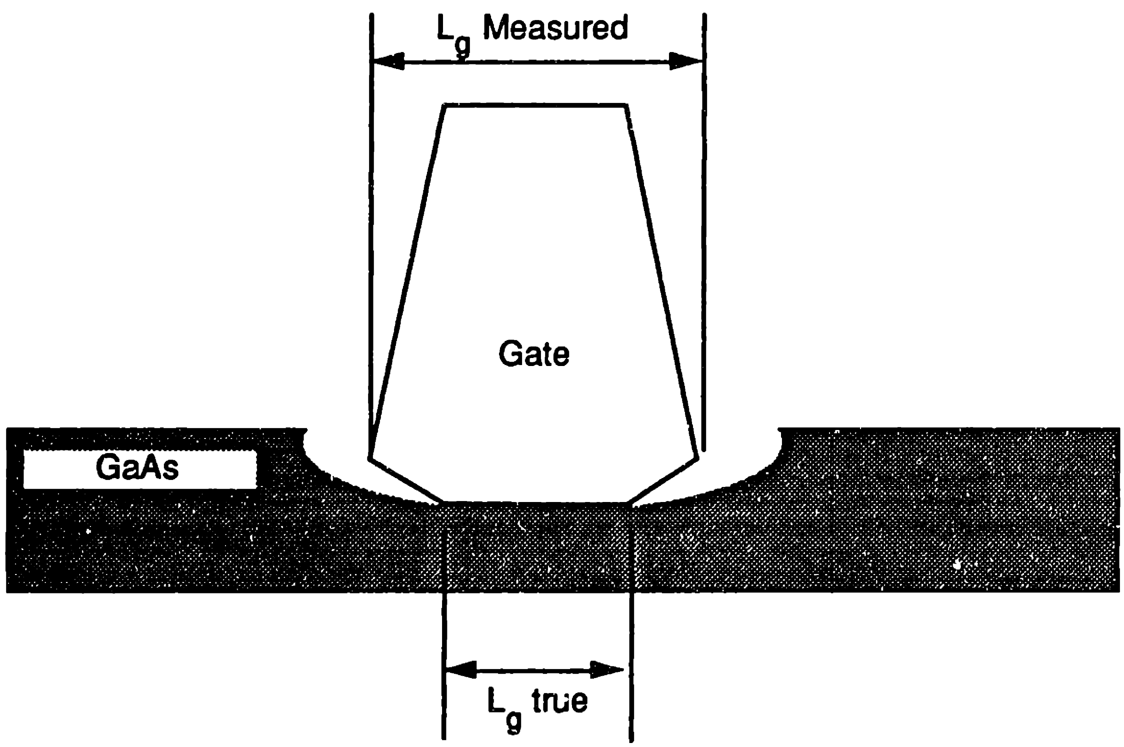


Figure 5.8.7: Close up of finished gate metal

After a stabilization bake, the PMMA is aligned and exposed with deep-UV illumination on a contact lithography system. Exposure dose is kept semi-constant by calibrating the illumination level before every batch is run. Calibration is accomplished using an exposure analyzer to measure intensity and then dividing this into the total desired dose to get the necessary time.

Developing is done in a 50:50 mixture of MIBK:Iso to remove the exposed PMMA and leave a PMMA line of length L_p. Batch-to-batch variation in developing is reduced by targeting the develop time before each new batch is processed. Targeting is done using a silicon "dummy" wafer. A single dummy is developed for a set period that approximates, but undershoots the necessary develop time. The resulting PMMA line is then examined with a SEM. Based on the line length observed and an experimentally determined function of change in L_p with increased develop time, the dummy is redeveloped. This process continues until L_p on the dummy is within the acceptable limits of 0.30 µm to 0.38 µm. Nominal L_p equals 0.35 µm. Then the total develop time, nominal develop time is 120 seconds, of the dummy is used to batch develop the actual product wafers. It is assumed that this calibration control accounts for batch-to-batch variability in prior processing as well as batch-to-batch variability present in the develop process. (See Figure 5.8.2)

Subsequent processing seeks to transfer the current critical feature, the PMMA line, to an approximately equal line of gate metal. To accomplish this, a metal transfer layer is deposited, which because of the PMMA thickness and sharp profile, is non-conformal. A transfer layer opening, of length L_t, is left when the PMMA is dissolved away and the transfer layer now acts as a conformal mask. (See Figure 5.8.3). L_t is the parameter shown in the preceding section to be useful as an in-process monitor.

Reactive Ion Etching (RIE) is used to transfer this pattern anisotropically thro3 379 rst the polyimide and then the oxide. An isotropic wet etch process is used iteratively to etch the semiconductor and target the drain current, I_{dss}, by controlling channel depth with the etching process. (See Figure 5.8.4). The gate metals are sequentially evaporated into this trench with the effective gate length controlled by the location of the sidewalls, the steepness of the trench profile and the aperture in the evaporator. (See Figure 5.8.5). Finally, the polyimide is lifted in NMP to remove the transfer layer and the excess gate metal. This yields the final gate profile. (See Figures 5.8.6 and 5.8.7)

5.4.2 In-Process Control Monitoring

As seen in Figure 5.8.7, the shape of the gate prevents direct observation of the actual gate length from a top view. As discussed previously, this implies that gate length cannot be easily measured in process. Rather, one is forced to use the transfer layer opening, L_t, and assume that it acts as a good indicator of final gate length.

Wafers fabricated on the MMIC-A process show a high variability in L_t. (See figure 5.9) This process has a mean of 0.411 μm and a standard deviation of 0.069 μm.

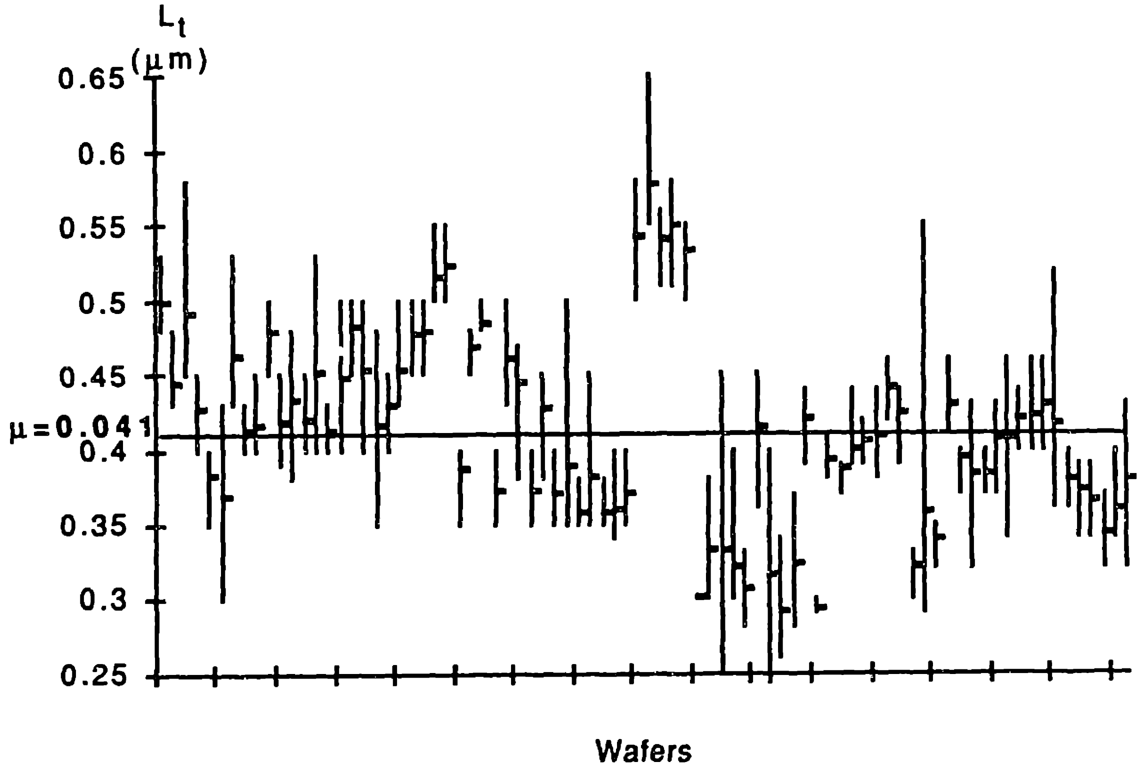


Figure 5.9: Variation in L_t over time.
Graph shows range and mean by wafer for 82 wafers.

The process description suggests that the variability in transfer layer opening should equal the variability in PMMA line length, L_p, plus the effects of depositing the transfer layer and lifting the PMMA line. Provided the side profile of the PMMA line is repeatable over time, these additional effects should be small and the variability of L_t should be approximately equal to the variability of L_p. A review of process data shows that this is indeed the case (see Figure 5.10).

$$L_t\ (\mu m) = 0.003 + 0.837\ ^*\ L_p\ (\mu m) \qquad 5.8$$

Mult-R = 0.918

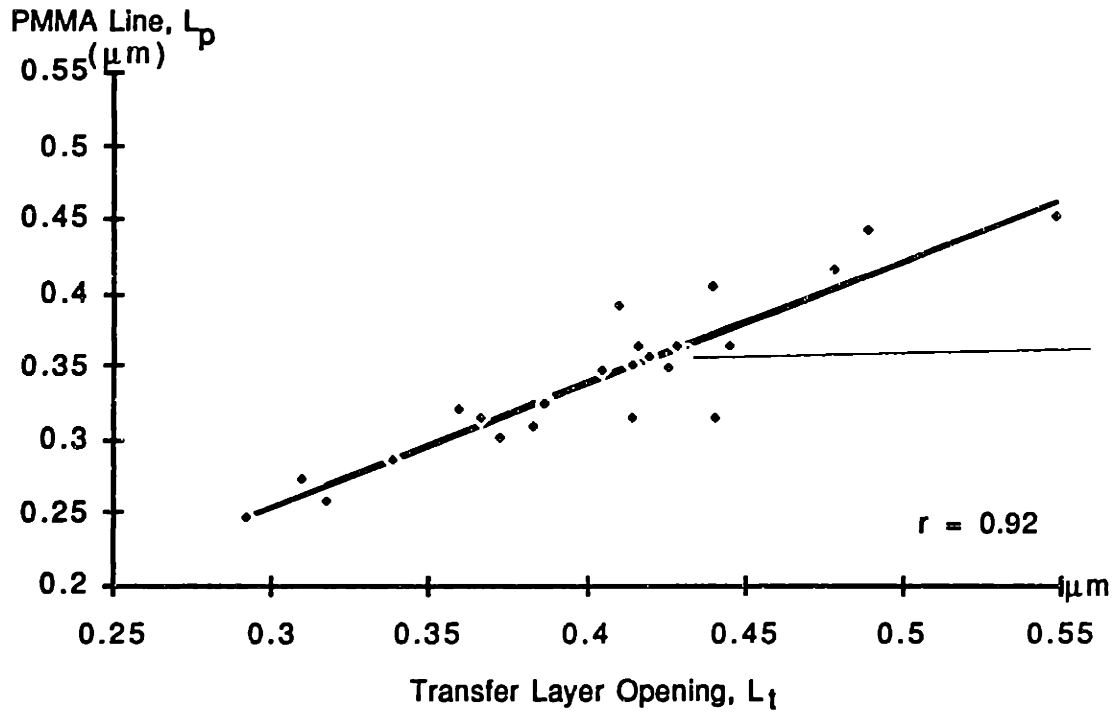


Figure 5.10: PMMA Line Length versus Transfer Layer Opening

Hence, the problem of understanding and controlling the variability in Gain Slope has been simplified by containing the origins of most of this variability to the processes that fabricate the PMMA line: spin/bake PMMA, align/expose, and PMMA develop. The spin/bake polyimide and all preceding processes should not play a substantial role because the polyimide planarizes the wafer for the PMMA and should not interact with the expose or develop processes.

In adopting L_p as an in-process control monitor, it is necessary to modify the observed mean and standard deviation in L_t to account for the relationship noted above. This suggests a mean $L_p = 0.347$ µm. Standard deviation can be calculated by

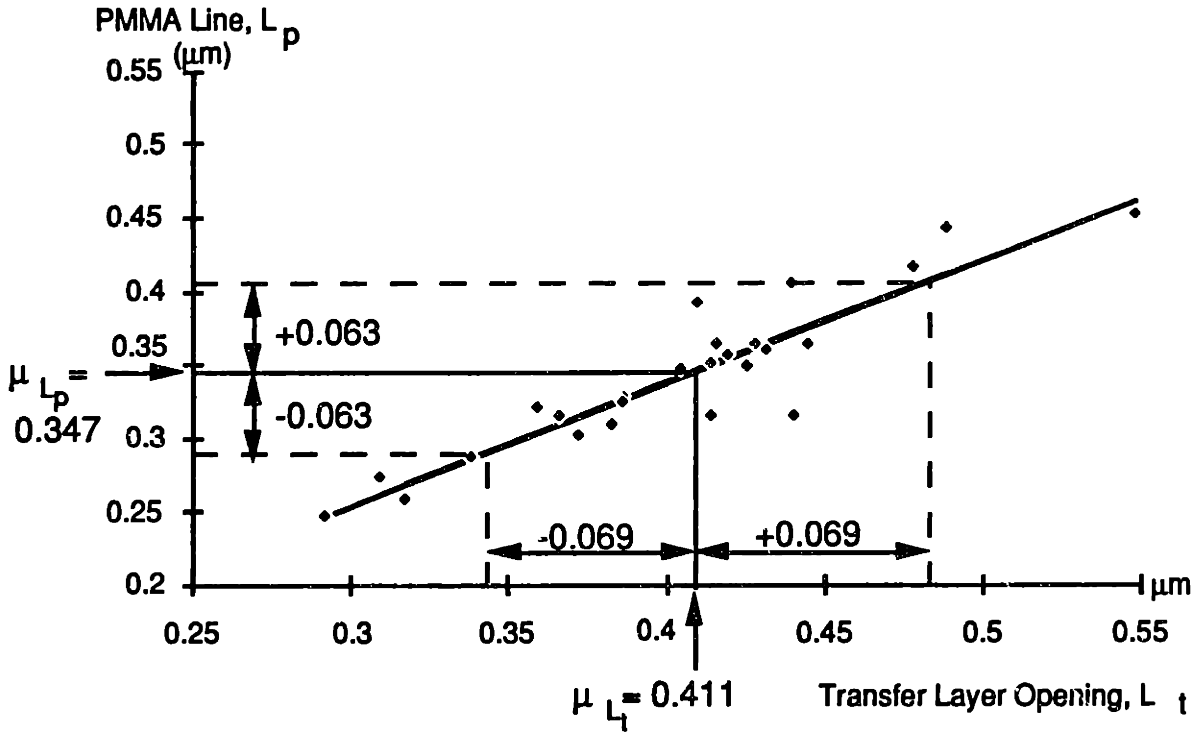
$$\sigma L_p = m\frac{\sigma L_t}{r} \qquad 5.9$$

where

m = slope of L_p / L_t

r = correlation coefficient of L_p vs L_t

Using the results of equation 5.8, this suggests a standard deviation in L_p of 0.0629 µm. (see Figure 5.11)



5.11 Effect of variation in Transfer Layer Opening on L_p

5.4.3 Sources of Process Variability

Previous work on semiconductor process control suggests several factors that are likely to be critical for control of each of the three process steps outlined above.[48] This served as a basis for choosing the possible sources of variability. However, current practice still regards a standard deviation on the critical dimension of 0.2 µm as state of the art in process control for a silicon manufacturing environment. This is primarily because most semiconductor fabrication research is done for digital silicon processes which are less sensitive to the linewidth variability than this GaAs analog process, where linewidth directly affects gain slope. The MMIC-A process requires a standard deviation that is almost an order of magnitude lower than current silicon process technology, ≈ 0.02μm, in order to maintain high parametric yields in traveling-wave amplifiers. Thus, in considering what might cause PMMA length variation, it was necessary to consider some of the secondary effects which are negligible in other types of semiconductor processing.

From this, a list of twenty-two possibly relevant causes of variation was generated, as shown in figure 5.12.

Process Step	# #	Process Input
PMMA Spin	1	Spin speed
	2	PMMA viscosity
	3	Spin time
	4	Wafer off center during spin
	5	PMMA volume
	6	Room temperature
	7	Time since ash process
Align and Expose PMMA	8	Light Intensity
	9	Poor contact due to particles
	10	Exposure target method
	11	Intra-mask CD variation
	12	Compensating for mask accuracy
Develop	13	Develop mixture concentration
	14	Develop time
	15	Developer loading effect
	16	Number of develop cycles
	17	Developer strength aging
	18	Develop bath temperature
	19	Wafer temperature
	20	Humidity
	21	Agitation
	22	Develop time targeting method

Figure 5.12: Causes of variability in PMMA line length, by process step

Each of these 22 inputs was investigated to determine:

1) The variability of the input in the standard process
2) The effect of change in the input on device physical parameters

After an overview of experimental methods, those six inputs that were found to cause the largest variation in transfer layer length (and thus gain slope) are discussed in detail in section 5.5; the remaining inputs, the experiments used to estimate their effects and the results found are summarized in Appendix 1.

5.4.4 Description of Experiments

Experiments were designed to test for two values:

- the existing variation in each of these inputs to the standard process
- the effect of such variation on L_p

Wherever possible, currently available process data, relevant literature findings or manufacturer information was used to minimize the cost of experimentation. The experimental plan pursued the two sets of values in question through different methods and then combined the two sets to estimate the variability in L_p due to each respective process input.

5.4.4.1 Estimation of Existing Standard Process Variability of Process Inputs

"Input variability", the variability observed in one of the 22 inputs, was estimated in one of four ways:

1) Observe the Process in Action (Abbreviated: "Process" in Figure A1.1) - this was used wherever possible (for 17 of the 22 inputs). The production line was observed while taking care to minimize disturbance of the system. For six inputs, it was impractical to measure the process in action due to inconvenience in making measurements during the course of production or due to the lack of suitable measurement equipment.

2) Use Designed experiments ("Expt" in Figure A1.1)- For three inputs, spin speed, spin time, and exposure intensity, experiments were designed to simulate what the variability would be during typical production.

3) Use Literature or Manufacturer's estimate ("Liter" in Figure A1.1) - For one input, PMMA viscosity, the manufacturer's data on typical viscosity variation was used.

4) Make an Analytical Estimation ("Analyt" in Figure A1.1) - For one input, exposure dose, an analytical estimation was made using manufacturer's data on bulb emission variability, PMMA frequency sensitivity, and exposure analyzer frequency sensitivity.

For each input, three types of variability were measured:

- Die-to-die - the average variation in the die mean value of a process input across die on the same wafer
- Wafer-to-wafer- the average variation in the wafer mean value of a process input across wafers in the same batch
- Batch-to-batch - the average variation in the batch mean value of a process input across several batches

This study then gives rise to 66 data entries: 22 process inputs * 3 types of variability for each input.

In all cases, the one sigma (one standard deviation) value is used to describe "variability". Many of these variabilities turn out to be zero; often because that type of variability - for a particular process input - would make no sense. For instance, die-to-die variability in spin speed is zero because the entirety of the wafer "sees" the same rotation speed. In regressing data, these three types of variability were treated as independent of one another. Input variability error was estimated based on the repeatability of the measurement instrument used.

5.4.4.2 Estimation of Effect of Process Input Variability on L_p

These effects were measured in one of four ways:

1) Use Designed Experiments (Abbreviated: "Expt" in Figure A1.1)
Designed experiments were used for 14 of the 22 effects. Typical experimental design called for using three settings of the input under study. The Middle setting was chosen to approximate the current mean setting observed in the process and Low and High input settings were chosen to provide a large swing in the output under test to fully explore the experimental space around the current process operating point. The mid-point was replicated three to five times to estimate experimental error.

Whenever variables were thought to lack independence, for instance exposure dose and develop time, a matrix experiment was used to allow for estimating the cross effects.

For one input, develop time targeting, a different type of experimental design was used: an entire lot of silicon dummy wafers was processed as a batch until the develop step. Then each wafer was processed separately in the manner typical for targeting the develop time. The variation in recommended target time from these wafers was used to calculate an estimator of the targeting process's effect on PMMA line length.

2) Research Manufacturer's Literature ("Liter." in Figure A1.1) - For 4 of the 22 effects - PMMA viscosity, room temperature during PMMA spin, develop bath temperature, and humidity during develop - effects stated in the manufacturer's literature were used.

3) Correlate input data to PMMA line length data ("Correl." in Figure A1.1) - For 3 of the 22 effects - particles on the wafer, mask repeatability, and mask accuracy - correlations were made from available process data to determine the effects of various input settings on PMMA line length. This was only done when considerable process data was available that recorded both the values of the input under study and the corresponding values of PMMA line length.

4) Analytically calculate the effect ("Analyt" in Figure A1.1) - For 1 of the 22 - wafer temperature before develop - an analytical derivation was used to examine the rate of wafer to bath heat exchange. This calculated rate of heat exchange was multiplied by the effect of variation in developer bath

temperature (determined from manufacturer's literature) to compute the net effect of variation in wafer temperature before develop on the resulting PMMA line length.

These experimental method are summarized in Figure A1.1, in Appendix 1.

5.4.4.3 Combining Experimental Results to Estimate the Variability in L_p Due to Each Process Input

The measured values of input variability and input effect were multiplied to estimate the respective impact of current variability in process inputs on variation in L_p. These results can be used to order the inputs into a chart (see Figure 5.18) that then suggests which inputs offer the most potential reduction in L_p variability. Nominally, this pareto chart can be used to allocate resources to process improvement; however, it does not take into consideration the amount of resources necessary to reduce a given input variability.

5.5 Discussion of Key Results

In this section, the results of the search to identify the causes of variability in L_p are presented. In section 5.5.1, the observable variation in L_p is compared to that estimated by summing the experimentally determined effects of each of the 22 process inputs. This comparison is made for total variability and for each of the three types of variability: die-to-die, wafer-to-wafer, and batch-to-batch. Continuing in section 5.5.2, the primary causes of L_p variability are highlighted. It was found that variability in L_p is primarily due to six process inputs. Accordingly, each of these inputs is discussed in some depth in sections 5.5.3-5.5.8. Finally, the managerial implications of these results are discussed in section 5.6.

5.5.1 Observed vs Estimated Variation in L_p

It was discussed previously that L_t demonstrates considerable variability when observed over time. (see Figure 5.13) In Figure 5.11, equation 5.9 and section 5.4.2, it was shown that this variability can be used to estimate the variability in L_p. The results of this translation are presented in Figure 5.14, which plots the standard deviation of L_t as a function of wafer number.

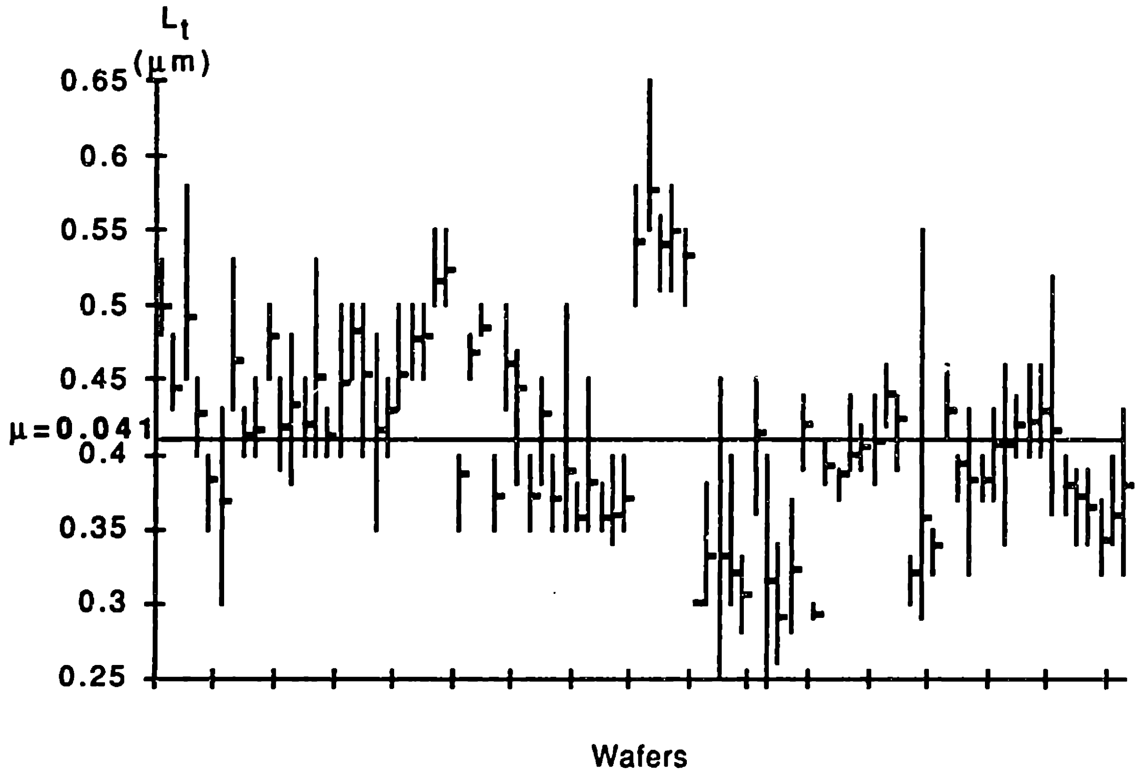


Figure 5.13: Observed process variability in Transfer Layer Length of Traveling-Wave Amplifiers made on MMIC-A process. (Repeat of Figure 5.9) Graph shows range and mean by wafer for 82 wafers.

Using the same data set that shows total variability in Figure 5.13, one can separate out the three independent types of variability in L_t: die-to-die, wafer-to-wafer and batch to batch. The separated variabilities are presented in figures 5.15, 5.16 and 5.17. Equation 5.9 is again used to estimate respective variabilities in L_p from the L_t data, the results of which are shown in figure 5.14.

Using the method of Kline and McKlintock (see Appendix 2 for more detail on this method), these variability components combine to equal total variability according to equation A2.2

$$\delta R = \left[\sum_{i=1}^{N} (\delta m_i)^2 \right]^{\frac{1}{2}} \qquad \text{A2.2}$$

where δR equals total variability and δm_i refers to the three types of variability.

	Observed		Estimated by Experiment
	L_t	L_p- Translated	L_p
Total	0.069	0.063	0.049
Batch-to-Batch	0.057	0.052	0.031
Wafer-to-Wafer	0.025	0.023	0.030
Die-to-Die	0.029	0.026	.024

Figure 5.14: Comparison of observed and experimentally predicted variability in L_p

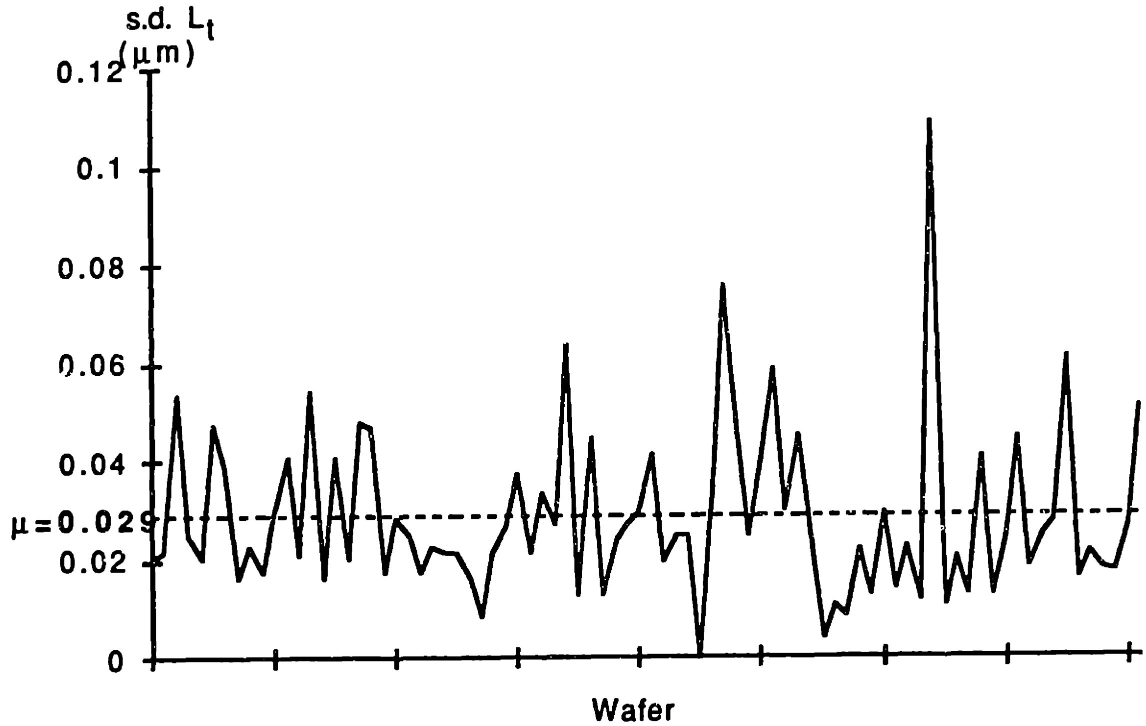


Figure 5.15: Observed Die-to-Die variability in Transfer Layer Length of Traveling Wave Amplifiers made on MMIC-A process

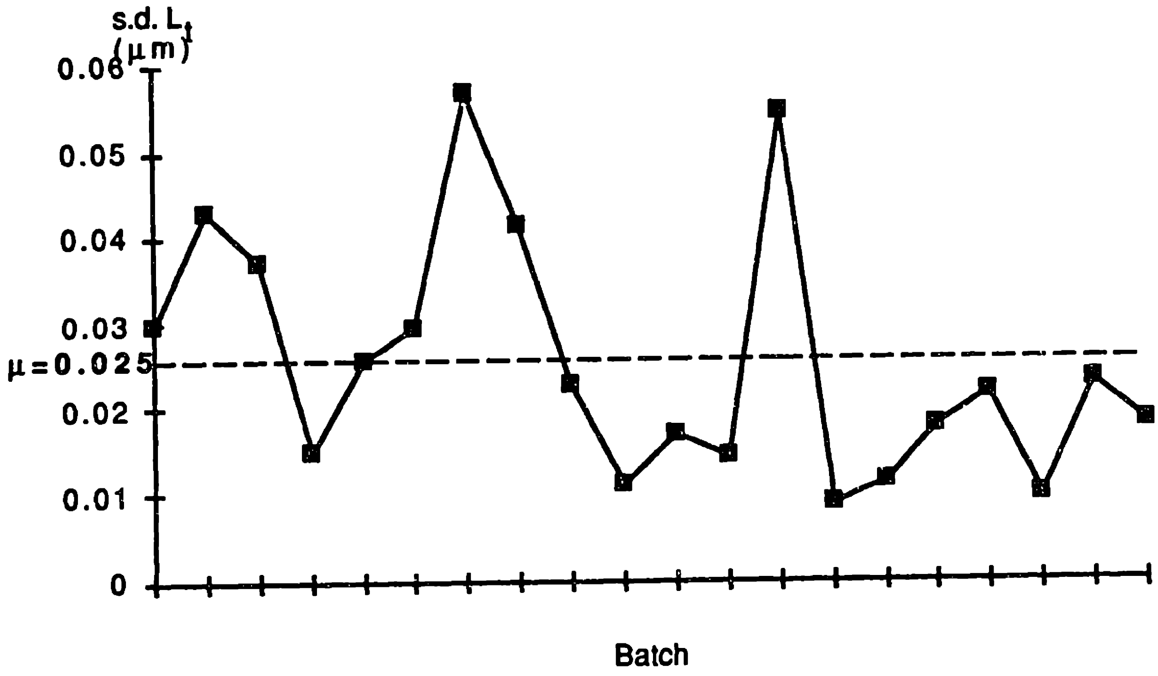


Figure 5.16: Observed Wafer-to-Wafer variability in Wafer Mean Transfer Layer Length of Traveling Wave Amplifiers made on MMIC-A process

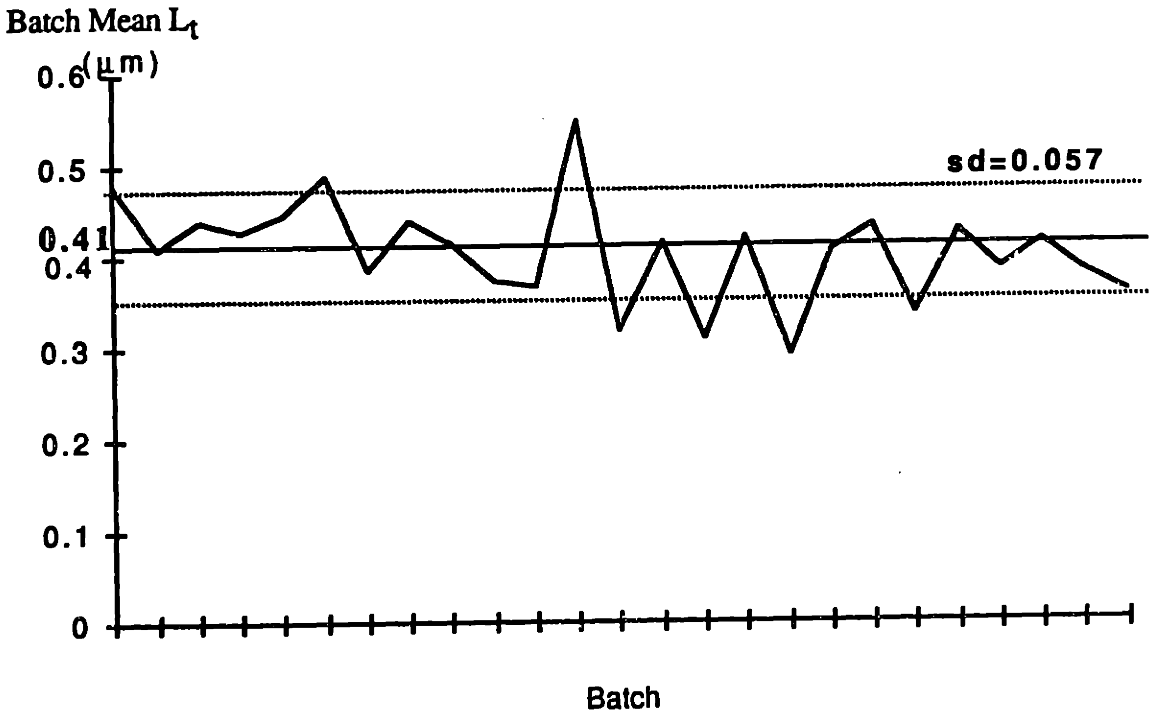


Figure 5.17: Observed Batch-to-Batch variability in Batch Mean Transfer Layer Length of Traveling Wave Amplifiers made on MMIC-A process

These observed values of variation in L_p can be compared against the sum of the variability estimated to be induced by each of the 22 process inputs studied. Assuming that these 22 inputs are independent to first order, the sum of these effects proceeds according to equation A2.2 with dR equal to total estimated variability in L_p and dm_i equal to the variability induced by a specific process input. The results of this summary are presented in figure 5.14 to allow comparison with the L_p variation that is expected based on observed variation in L_t.

These values compare favorably with the observed variation in L_p, typically accounting for 60-130% of the observed variation. In general, one would hope that the experimentally estimated values would underestimate the observed value to allow for possibly relevant process inputs that may not have been considered in these experiments. In any case, the differences observed are easily explained by the relatively large error margins on most experimental results and hence the experimental results seem all the more plausible as a true explanation of how variability is induced in L_p. (See Appendix 2 for a fuller description of experimental error.)

5.5.2 Primary Causes of Estimated Variation in L_p

This section discuses which inputs contribute most to the estimated variability in L_p. This analysis is repeated for each of the three types of process variability: Die-to-die, Wafer-to-wafer, and Batch-to-batch.

Six process inputs were found to dominate variability in L_p:

- the presence of particles on the wafer during contact gate lithography,
- variation in intra-mask CD,
- variation in mean mask CD,
- exhaustion of the developer due to loading effects,
- variation in developer bath temperature, and
- the method used to target the develop time.

Experiments and results for each of these six inputs are discussed in detail in the following sub-sections. Included in this review are a brief description of the input, how the experimentation or analysis was done, and the key results with a discussion of their reasonableness and their implications to the process. Summary data for the remaining sixteen inputs can be found in Appendix 1.

An overview of the variability in L_p induced by each of the process inputs is presented in Figure 5.18. This graph suggests several important conclusions regarding the process origins of variability in PMMA line length.

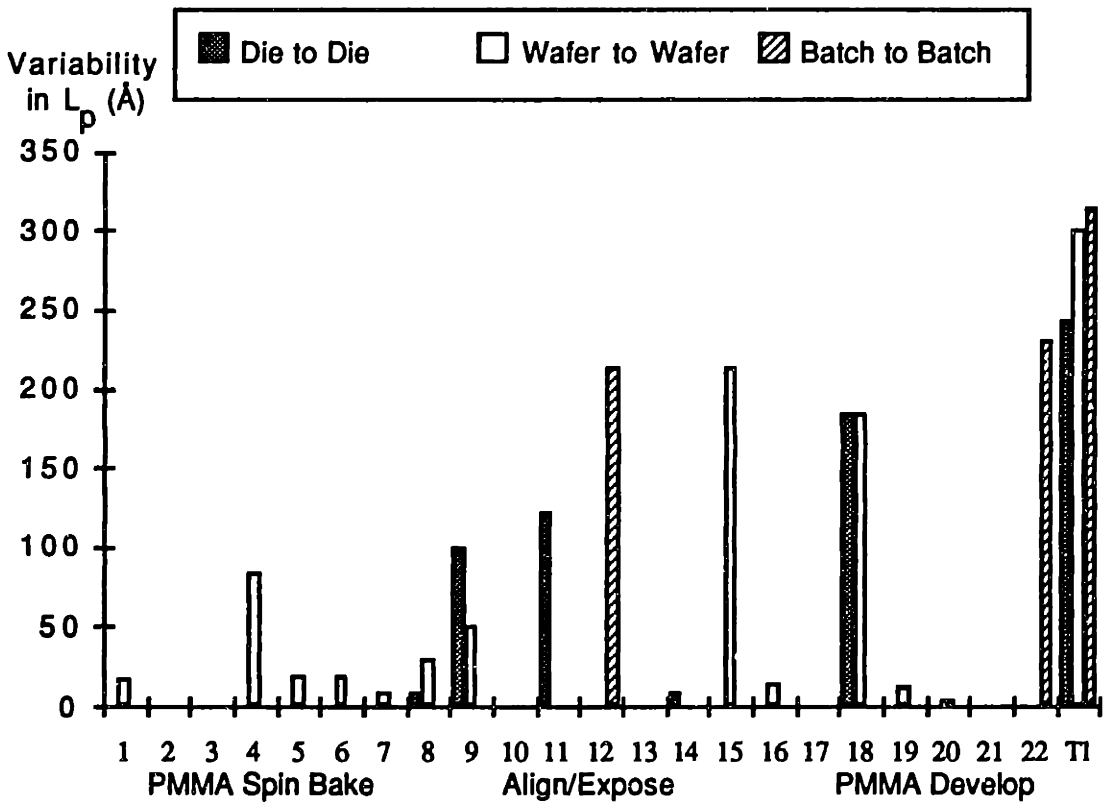


Figure 5.18a: Variability in L_p(Å) due to each of 22 process inputs
(Three types of variability measured for each input.
Where no bar appears variability was found to be not applicable.)

#	Process Input	Die-to-Die	Wafer-to-Wafer	Batch-to-Batch
1	Spin speed	0	18	0
2	PMMA viscosity	0	0	0
3	Spin time	0	0	0
4	Off-center spin	0	83	0
5	PMMA volume	0	19	0
6	Room temperature	0	20	0
7	Time after ash	0	8	0
8	Light intensity	9	30	0
9	Poor contact	100	50	0
10	Exposure target method	0	0	0
11	Intra-mask CD variation	123	0	0
12	Mask accuracy	0	0	214
13	Develop mixture concentration	0	0	0
14	Develop time	9	0	0
15	Developer loading effect	0	213	0
16	# of Develop cycles	0	15	0
17	Developer strength	0	0	0
18	Developer bath temperature	183	183	0
19	Wafer temperature	0	12	0
20	Humidity	0	4	0
21	Agitation	0	0	0
22	Develop time targeting method	0	0	230
	Total	242	301	314

Figure 5.18b: Variability in L_p(Å) due to each of 22 process inputs
(Table has same values presented graphically in Figure 5.18a)

Conclusions

1) Most of the variability in L_p is due to only six of the inputs. For each type of variability, these six account for > 90% of the variability in L_p (with one major caveat on batch-to-batch variation discussed in item #4 of this list). (See Figure 5.19) Hence it is appropriate to constrain the focus to these six variables to simplify matters as has been done in Figure 5.20

	Experiment. Predicted Total Variat.	Contribution of top six inputs	Contribution of other inputs	$\frac{(Top\ 6)^2}{(Total)^2}$
Batch-to-batch	314 Å	314 Å	0 Å	100%
Wafer-to-Wafer	301 Å	285 Å	97 Å	90%
Die-to-Die	242 Å	242 Å	12 Å	100%
Total	498 Å	488 Å	99 Å	96%

Figure 5.19: Variability explained by six primary causes

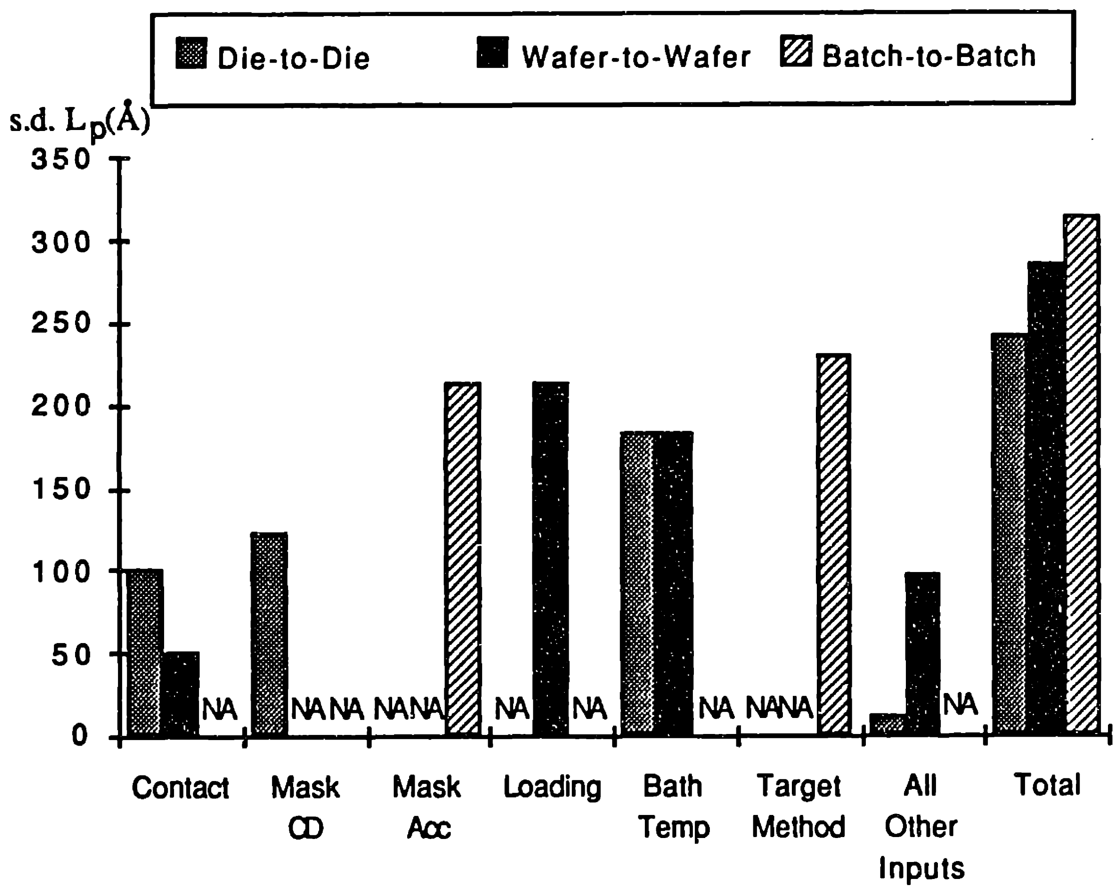


Figure 5.20: Variability in L_p(Å) due to top 6 sources among 22 process inputs "NA" refers to variability type not applicable for that process input

2) Die-to-die variation accounts for a one standard deviation spread in line length of 242 Å out of the total 260 Å observed die-to-die variation in PMMA line length. (see Figure 5.14) This spread is dominated by three components which, in order of importance, are:

- variation in developer bath temperature (s.d. = 183 Å),
- variation in intra-mask CD (s.d. = 123 Å)
- the presence of particles during the contact gate lithography step (s.d. = 100 Å).

Together, these three inputs explain 99+ % of the observed die-to-die variability.

3) Wafer-to-wafer variation accounts for 301 Å out of the 230 Å observed wafer-to-wafer variability. (see Figure 5.14) This over explaining is most likely a result of experimental error. Wafer-to-wafer variation is dominated by two components:

- exhaustion of the developer due to loading effects (s.d. = 213 Å),
- variation in developer bath temperature (s.d. = 183 Å).

Together, these two inputs explain 90% of the observed wafer-to-wafer variability.

4) Batch-to-batch variation represents the largest of the three components, accounting for 310 Å out of the 520 Å observed batch-to-batch variation in L_p.

Nominally, batch-to-batch variation should be 602 Å (see Figure 5.21); however, the method used for targeting develop time is effective in 'tuning' the develop process to this nominal batch variation and essentially "controlling away" most causes of batch-to-batch variability - however, the control process introduces significant variability itself because it relies on a few readings from a single dummy wafer that is subject to the considerable die-to-die and wafer-to-wafer variation just discussed. Hence, this control wafer is subject to considerable bias. Additionally, variation in the accuracy of mean mask CD makes the entire process less sensitive to other causes of variability and is not "controlled away" by develop time targeting.

Estimates of the batch-to-batch variability that would exist without the develop time targeting process were made using the observed batch-to-batch variation in each of the 21 process inputs - all 22, except develop time targeting - and multiplying times the effect of the respective input on variation in L_p. The results of this are shown in figure 5.21. Note that this estimate of batch-to-batch variability - without time targeting (602 Å) - is much closer to the observed batch-to-batch variability in L_p (520 Å, see Figure 5.14) than the estimate of batch-to-batch variability that incorporates develop time targeting (314 Å, see Figure 5.14). Because batch-to-batch variability is not well explained by the experiments conducted, this questions how well the time targeting method is actually working in practice.

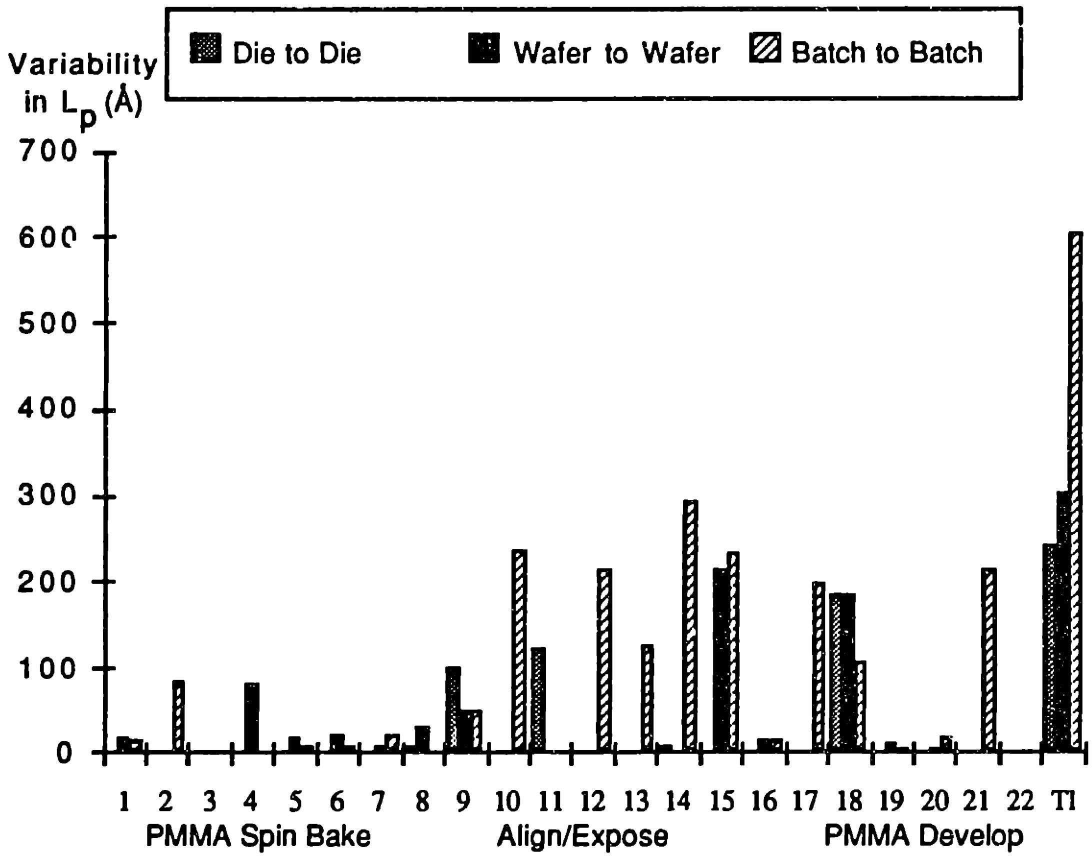


Figure 5.21a: Uncontrolled Variability in L_p(Å) due to each of 22 process inputs
(Three types of variability measured for each input.
Where no bar appears variability was found to be not applicable.)

#	Process Input	Die-to-Die	Wafer-to-Wafer	Batch-to-Batch
1	Spin speed	0	18	14
2	PMMA viscosity	0	0	85
3	Spin time	0	0	0
4	Off-center spin	0	83	0
5	PMMA volume	0	19	8
6	Room temperature	0	20	9
7	Time after ash	0	8	20
8	Light intensity	9	30	0
9	Poor contact	100	50	50
10	Exposure target method	0	0	234
11	Intra-mask CD variation	123	0	0
12	Mask accuracy	0	0	214
13	Develop mixture concentration	0	0	126
14	Develop time	9	0	293
15	Developer loading effect	0	213	231
16	# of Develop cycles	0	15	15
17	Developer strength	0	0	198
18	Developer bath temperature	183	183	107
19	Wafer temperature	0	12	5
20	Humidity	0	4	17
21	Agitation	0	0	213
22	Develop time targeting method	0	0	0
	Total	242	301	602

Figure 5.21b: Uncontrolled Variability in L_p(A) due to each of 22 process inputs (Table has same values presented graphically in Figure 5.19a)

5) No inputs to the PMMA spin process were found to be major sources of variability (See Section A1.1 for details).

Results of input variability and effect experiments are tabulated in Figures 5.20 and 5.21, together with the combination of these two results into an estimate of induced variability in L_p. These tables are explained in the following subsections.

		Unit	9	11	12
	Unit		# Particl.	Å	Å
Input Variation	Die-to-Die	---	2	170	0
	Wafer-to-Wafer	---	1	0	0
	Batch-to-Batch	---	1	0	650
Effect on L_p	$\partial L_p/\partial$ Input	Å/input units	50	1	0.328
L_p variation	Die-to-Die	Å	100	123	0
	Wafer-to-Wafer	Å	50	0	0
	Batch-to-Batch	Å	50	0	214

Figure 5.22: Summary of Results for Primary Align/Expose process inputs

		Unit	15	18	22
	Unit		wafers	°C	Å-s.d.
Input Variation	Die-to-Die	---	0	0.29	0
	Wafer-to-Wafer	---	4.8	0.29	0
	Batch-to-Batch	---	5.2	0.17	230
Effect on rate	∂ rate/ ∂ Input	Å/s /input units	0.37	5.26	1
Effect of rate on L_p	$\partial L_p/\partial$ rate	Å/ (Å/s)	120	120	1
L_p variation	Die-to-Die	Å	0	183	0
	Wafer-to-Wafer	Å	213	183	0
	Batch-to-Batch	Å	231	107	230

Figure 5.23 Summary of Results for Primary Develop process inputs

5.5.3 The presence of particles during the contact gate exposure

Air borne and fluid borne particles are of constant concern in semiconductor processing due to the deleterious effects such particles have on the lithography process. In VLSI technologies, such particles prevent complete transfer of the mask image by blocking some portion of the transmitted light. In low scale integration, contact lithography technologies such as MMIC-A, this effect also occurs and contributes to reliability failure and visual defects; however, such particles can also interfere with the contact between mask and wafer preventing accurate imaging of the mask image onto the wafer, thereby causing variation in L_p. (see Figure 5.24) It is this latter effect that is considered here.

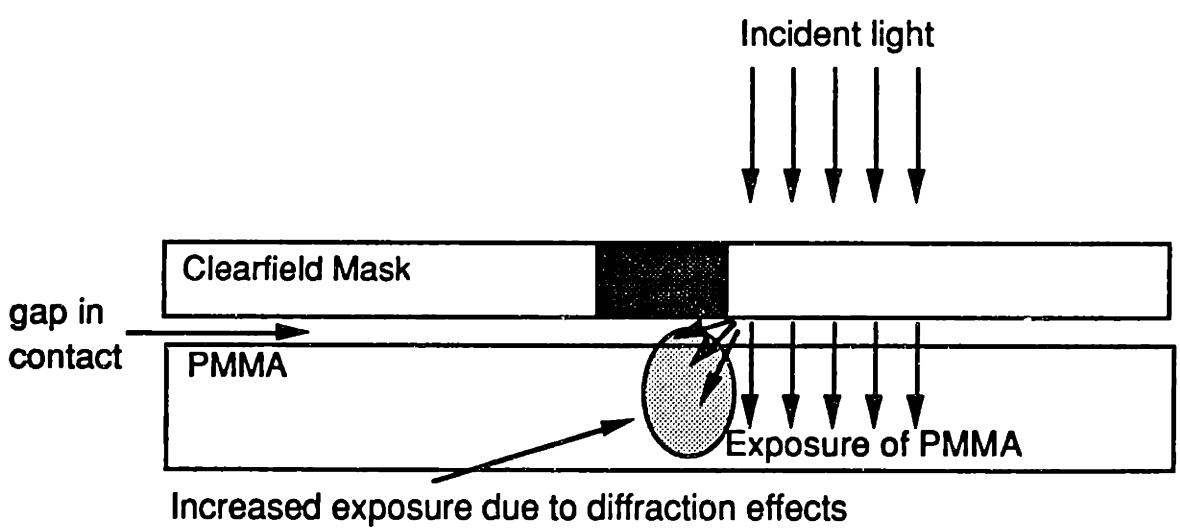


Figure 5.24: Diffraction effects due to mask and wafer not being in good contact due to particles

The presence of such particles is confirmed by Newton rings, which are observed when a wafer is brought into contact with the mask and a non-planar point, typically a particle of foreign matter, exists between the two surfaces. These points can be correlated, loosely, to the observed PMMA line length in the same region as the particle on a given wafer (see Figure 5.25).

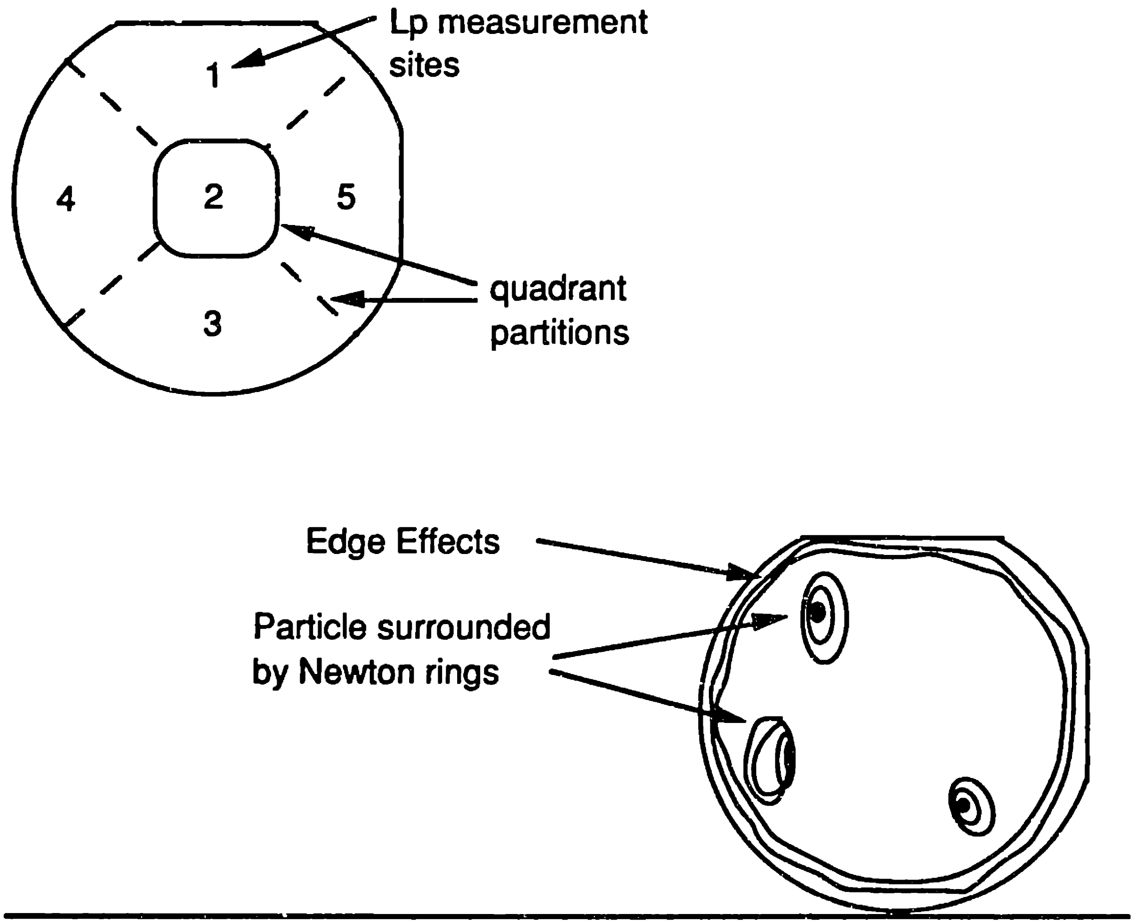


Figure 5.25: Newton rings on a wafer and use of quadrants to examine effect of particles on wafer.

Wafer maps, shown in figure 5.25, drawn by hand by the operator before wafer exposure for 36 production wafers were used to count the number of particles in each quadrant of the wafer and these values were used to calculate variability in the number of particles present. Die-to-die, wafer-to-wafer and batch-to-batch variation in particles on the wafer were found to be 2, 1, and 1 particles, respectively. PMMA line length values for each quadrant were normalized based on the wafer average PMMA line length. These two values were correlated to determine the effect of a given number of particles on the localized PMMA line length, 50 Å per particle (see Figure 5.26).

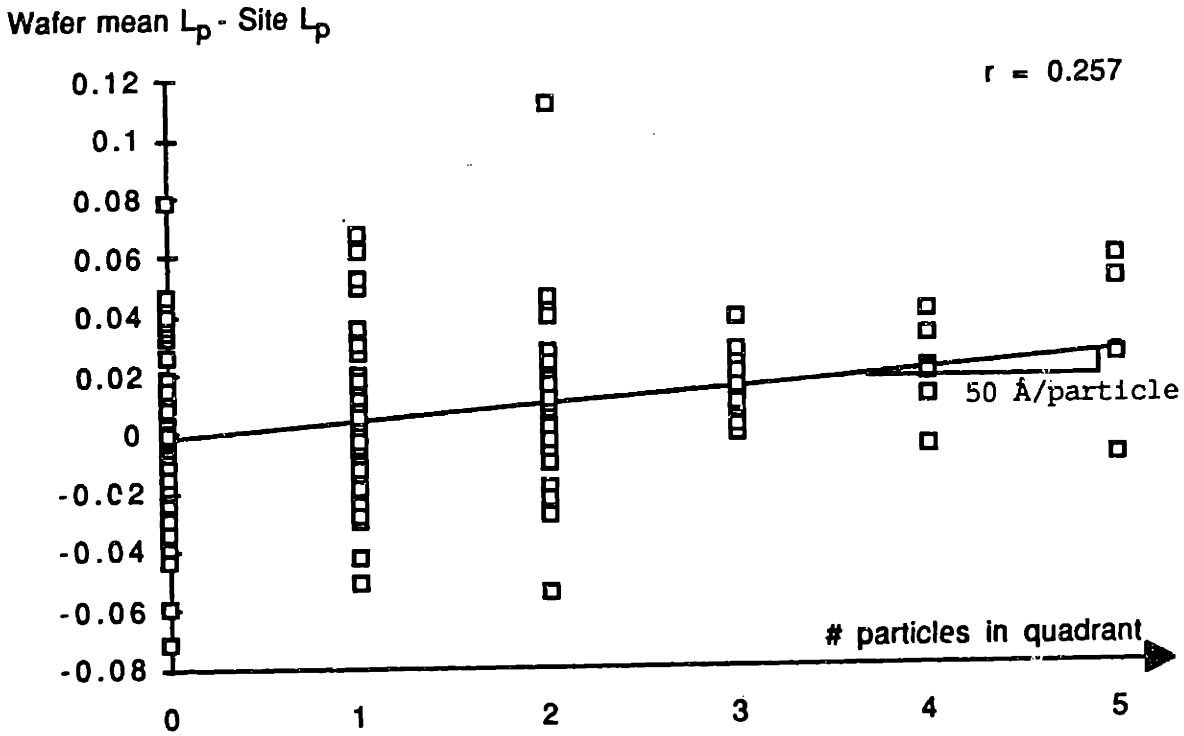


Figure 5.26: Correlation of region L_p minus wafer mean L_p to particles in that region of the wafer

Results are summarized in Figure 5.22. Most notably, poor contact contributes die-to-die variation in L_p of 100 Å. This input also contributes wafer-to-wafer variability in L_p of 50 Å, although this effect is less important than other causes of wafer-to-wafer variability. The contribution to batch-to-batch variability, while nominally 50 Å, is assumed to be tuned away through the develop time targeting method.

The presence of particles on the wafer prevents good contact between mask and wafer during exposure, thereby increasing diffraction effects. Because a clear field mask is used for gate lithography, the resulting PMMA line becomes smaller as diffracted light encroaches upon the normally unexposed PMMA directly beneath the gate pattern on the mask. In particular, light diffraction causes a PMMA line profile that is more undercut than usual (see Figure 5.24).

Because the correlation of figure 5.26 ignores some obvious factors, such as the size of the particle on the wafer, the accuracy of the effect observed, 50 Å smaller per particle in the vicinity, is suspect and should be used only as a figure of merit until physical relevance is demonstrated . The high error estimate associated with the effects of this input confirms that caution should be attached to the precise value of the effect.

Nonetheless, the value calculated suggests that particles on the wafer are a sizable contributor to die-to-die, and, to a smaller degree, wafer-to-wafer, variability in PMMA line length.

5.5.4 Variation in intra-mask CD (Mask CD)

Masks for the contact gate lithography step are fabricated by an outside vendor. In this fabrication process, the gate line pattern on the gate mask is replicated several times on a single plate because a single circuit contains several gates and each mask contains the gate images of several circuits. However, the length of this gate line pattern varies across the mask due to variability in the mask making process. When used to print a device, this intra-mask variability leads to intra-wafer, or "die-to-die", variability.

Note that this process 'input' considers only the repeatability of the mask line length, not the accuracy. Because mask accuracy has a much different effect on L_p variation and because mask accuracy is controlled in a different manner, it was considered to be a separate process input and is addressed in the following section, 5.5.5 .

Five sites were measured on each of seven recent MMIC product masks for the gate level using a transmitted visible light measurement system[49]. The range of the site values for each of the seven masks (0.02 µm, 0.02 µm, 0.04 µm, 0.06 µm, 0.04 µm, 0.06 µm, 0.08 µm) were averaged to provide a mean value of intra-mask range, 0.046. Assuming these mask CD values follow a normal probability distribution, the standard deviation for the process was estimated using equation 5.10.

$$\sigma = \frac{\overline{R}}{d_2} \qquad 5.10$$

A standard deviation of 170 Å was found.

Since the mask image is directly transfered to the PMMA lines on the wafer, any variation in the mask image will also be directly transfered and show up as variation in PMMA line length across the wafer; although, the expose and develop process will shrink the mask image in transfering it. The intra-mask CD variation should contribute a variation in PMMA line length proportional to the shrinkage ratio, obtained by

$$\frac{\text{image transfered}}{\text{mask image}} = \frac{\text{mean PMMA line length}}{\text{mean mask length}} = \frac{0.35\mu m}{0.48\mu m} = 0.723$$

This effect is noted in Figure 5.27.

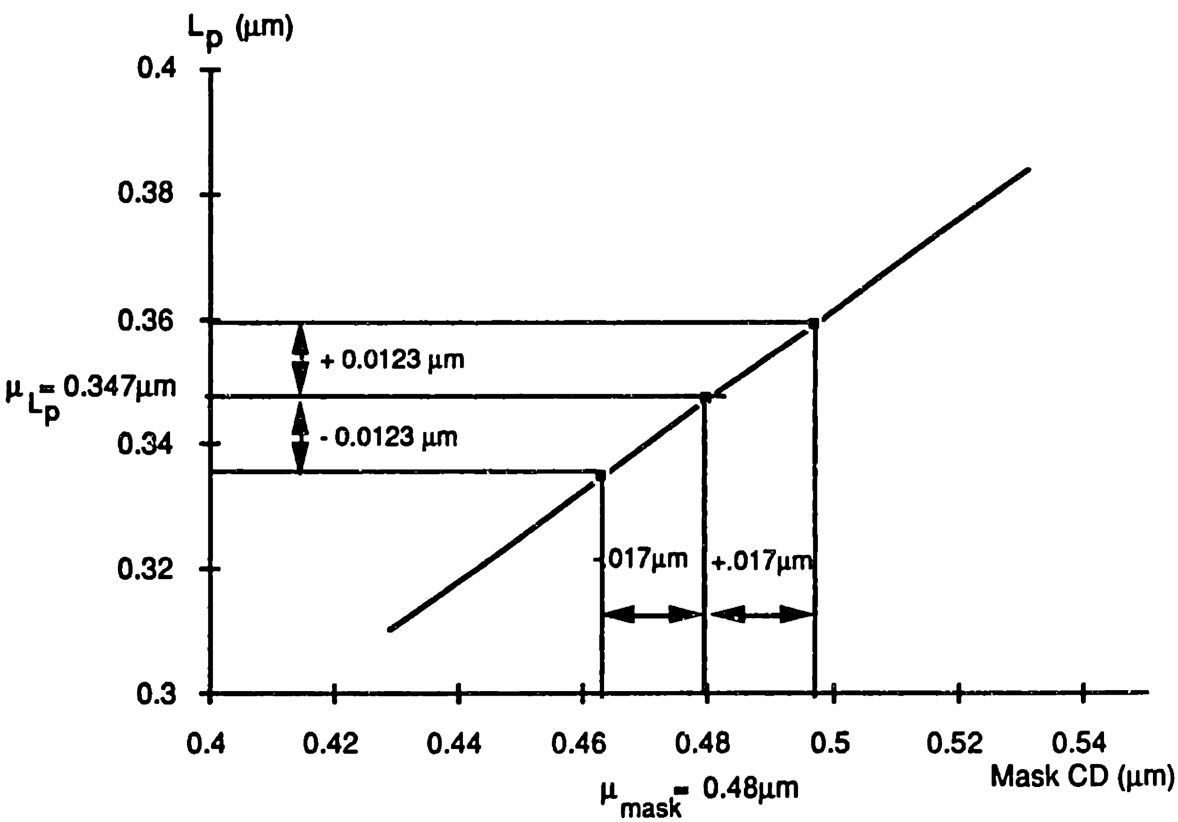


Figure 5.27: Transfer of intra-mask CD variation to intra-wafer L_p variation

The results summarized in Figure 5.22 come from multiplying the observed variability by the expected effect of this variability as calculated above. These results suggest that mask CD variation causes die-to-die variation in L_p of 123 Å, and that this input should not lead to wafer-to-wafer or batch-to-batch variation. This figure is in line with specifications on the mask which are based on what the vendor is capable of providing, 5000 Å ± 500 Å.

This variation results from limitations in the mask maker's ability to accurately replicate the critical gate linewidth across the entire mask. It is endemic to most production processes that the product is limited by the quality of the tools used. Different vendors may be able to supply higher repeatability masks or masks might be screened to choose only those with high intra-mask repeatability. Because the range of intra-mask CD variation (one standard deviation) observed on a given wafer varies from 200 Å to 800 Å, it seems clear that some masks can be made with low intra-mask variability. Correcting this problem is made more difficult by the difficulty of accurately and objectively measuring sub-micron masks. While MWTD is currently aware of the values that demonstrate intra-mask CD variability, they have focused more on mask to mask repeatability and accuracy.

5.5.5 Variation in Mean Mask CD (Mask Accuracy)

A second problem with the mask, beyond intra-mask CD variability, develops from the inability of the mask maker to repeatably produce an accurately sized mean gate line pattern. Hence, mask-to-mask variability can be significant. Since masks are not replaced often, one seemingly acceptable solution is to tune the exposure dose and develop time for each new mask. Indeed, this works well as a means for targeting the mean PMMA line length produced. However, this new operating point may be less tolerant to small variances in the process and lead to a different variance in L_p (higher batch-to-batch variability), even while mean L_p remains constant through targeting.

Variability in mask accuracy was estimated by looking at seven masks with the same nominal CD and was found to be 650 Å. The effect of variation in mask accuracy was estimated by examining batch-to-batch variability in wafers exposed using masks with different mean gate CDs. (see Figure 5.22)

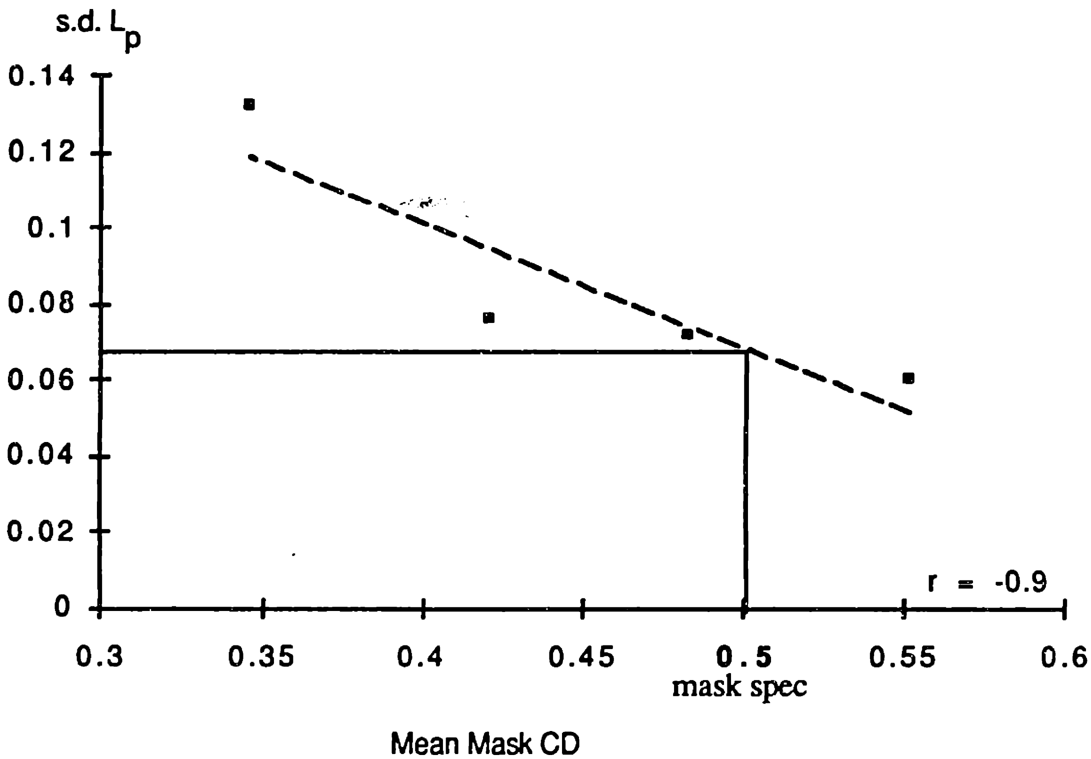


Figure 5.28: Batch-to-Batch variability as a function of mean mask CD

The results, shown in figure 5.28, suggest that masks with smaller gate images exhibit higher batch-to-batch variability in L_p. Each angstrom of variation in mean mask CD is estimated to lead to 0.328 Å variation in L_p. This is presumably because a certain amount of shrink from mask CD to PMMA line is a useful operating point for the process that is tolerant of process noise. Variation in mask accuracy was found to contribute batch-to-batch variability in L_p of 214 Å. The assumption that develop time targeting "corrects away" batch-to-batch variability does not help this, because the required over-develop time is part of what causes this effect.

5.5.6 Variation in developer bath temperature

The developer solution is mixed daily in the morning and put into a circulating water bath, temperature controlled to ±0.5°C. Nominally, this should control the temperature of the develop bath, reasonably well. Agitation during develop should further stabilize this temperature. However, any significant variation in bath temperature would change the rate of the develop reaction, thereby effecting L_p.

Process input variability of all types was estimated by measuring temperature in several zones in the developer solution at different times on different days and then separating out the respective effects. The total temperature range measured was only slightly greater than 1° C. The separation between die-to-die and wafer-to-wafer variation was not studied and the two were assumed to be equal because of the batch nature of the develop process. Interestingly, die-to-die and wafer-to-wafer variation were measured to be larger than batch-to-batch variation.

The effect of change in temperature on L_p was estimated using manufacturer's data. Since data was only available for MIBK:Iso (Methyl Iso-Butyl Ketone : Isopropyl Alcohol) concentrations different than the 50:50 ratio used in the MMIC-A process, interpolation was used to estimate the value of 5.26 (Å/s)/°C.

Develop bath temperature control is a well known source of process variation[50] and current practice calls for ±0.1 °C control of bath temperature, yet the current controller at MWTD is only specified to deliver ±0.5 °C control. The chemistry of the develop reaction confirms that such temperature variations will adversely impact the develop rate.

The results, see figure 5.23, suggest that when bath temperature variation is multiplied times the effect of temperature on develop rate times the nominal time for which this rate is in effect, 120 seconds, it contributes die-to-die and wafer-to-wafer variation in L_p of 183 Å. Batch-to-batch contribution, while nominally equal to 107 Å, is presumed to be "controlled away" through develop time targeting.

The current temperature control scheme is relatively new at MWTD. While it was probably a significant improvement over uncontrolled temperature variations, further control seems to still be needed.

5.5.7 Exhaustion of the developer due to loading effects

Exhaustion of the developer due to a lower concentration of developer reactants and a higher concentration of reacted compounds can reduce develop rate. Nominally, this effect is corrected for by using a sufficiently large batch of developer so that exhaustion is an inconsequential effect. At MWTD, the develop solution is 150 ml MIBK: 150 ml Iso and a single batch of developer can be used in several 5-wafer lots before being disposed of.

Variation in the number of wafers previously processed in a given batch of develop solution was calculated by observing the production line and the number of times a develop solution was used before disposal. Wafer-to-wafer and batch-to-batch variation were found to be 4.8 and 5.2 wafers, respectively. The effect of developer exhaustion on L_p was calculated by batch processing several silicon wafers to the develop step and then developing them throughout the day, with normal production wafers being developed between successive tests to increase the load effect on the develop solution. The results of the effect experiment are shown in figure 5.23 and suggest that additional wafers change the rate of develop by a factor of 0.37 (Å/s)/ wafer.

When this effect on rate is multiplied times the nominal time the rate is in effect, 120 seconds, and times the measured batch-to-batch and wafer-to-wafer variability in wafers already processed, the results suggest that developer exhaustion contributes to high wafer-to-wafer variability in L_p, 213 Å. Again die-to-die variation has little meaning for this input, and batch-to-batch variation, though nominally equal to 231 Å, is presumed to be "controlled away" by develop time targeting.

Developer exhaustion is avoided by providing a sufficient quantity of develop solution to dilute the exhaustion effect. Nominally, the current use of 300 mL of solution for no more than 20 wafers would seem adequate, although the results of this experiment suggest otherwise. While this is a problem that would benefit from more in depth research in the future, for now, the problem can be avoided by enlarging the develop volume or replacing the solution more frequently.

5.5.8 Develop time targeting method

The develop process uses a silicon dummy wafer to target the develop time for each batch. Nominally, this corrects for all batch-to-batch variation. However, the targeting process relies on a one wafer sample of the current batch mean. This implies taking a one data point sample from a population whose variance equals the wafer-to-wafer variation, a non-trivial number. Inherently, this control scheme biases the final output of the develop process, L_p, and induces a batch-to-batch variability in L_p equal to the error of the controller which in this case equals the total wafer-to-wafer variability. (see Figure 5.29)

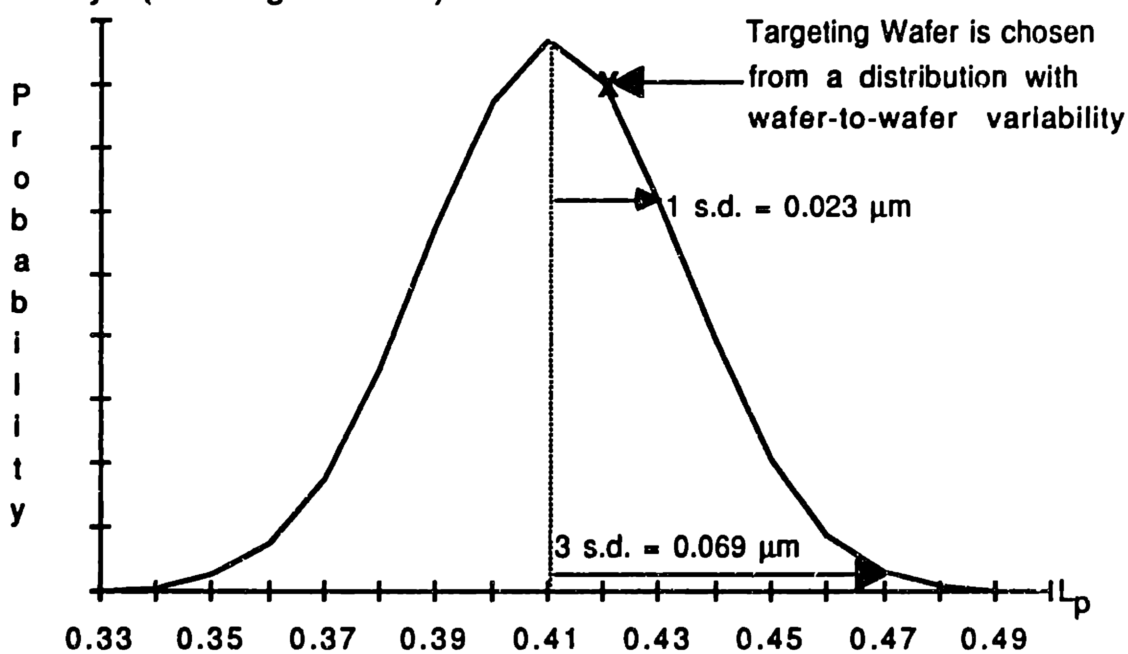


Figure 5.29: Biased Control in Develop Time Targeting

As expected, the targeting method contributes substantial variation to batch-to-batch variation in L_p, 230 Å. This makes the targeting method the single largest source of L_p variation.

Eliminating this variation is difficult in the sense that the very use of control does a lot to substantially reduce L_p variability.

Thus, MWTD has potentially reduced batch-to-batch variability from 602 Å to 314 Å, although this reduction was not apparent in these results. However, the control method adds considerable variability itself. Furthermore, better control methods are not clearly visible. Using the mean response of two dummy wafers might help eliminate some of this targeting bias, as would reducing the wafer-to-wafer variability in the process, through controlling developer exhaustion and developer bath temperature, among others, as discussed above.

5.5 Managerial Implications of Experimental Results

Having gone into considerable depth to identify the causes of process variability, I now consider whether the information provided can add value to the resource allocation decision process.

This variability study identified six process inputs as accounting for 90+ % of variability in L_p. Of these six, MWTD appears to be aware of four of the six, and actively pursuing only two of the six.

	MWTD aware of variability in:	Worked to reduce in past:	Working to reduce now:
Particles	Yes	Starting	Starting
Intra-Mask CD	No	No	No
Mask Accuracy	Yes	Yes	Yes
Loading Effects	No	No	No
Bath Temp.	Yes	Yes	No
Develop target	Yes	Yes	No

Figure 5.30: MWTD Attention to Prime Causes of Process Variability

MWTD has worked on three of the inputs in the past and appears to have greatly reduced the variability of two of them, develop bath temperature and develop time targeting method; however, these two inputs are still significant drivers of L_p variability. MWTD has only recently begun to consider the importance of particles on the wafer during contact lithography and is just beginning to invest explicitly in controlling this parameter.

The summary view is that MWTD is today not directly investing in the control of many of the primary causes of L_p variability. This would appear to leave the causes of most parametric yield loss, which consumes an estimated 7.5 % of total die started, largely unchallenged.

While explicit comparison between projects to control parametric yield loss and projects to control other yield loss mechanisms, such as wafer breakage or visual loss, are not possible without a consideration of the effort involved in controlling the inputs, it is clear that considerable savings can be realized by attacking known sources of variability.

An understanding of why HP appears to be ignoring primary causes of process variability can be gained, in part, from figure 5.31. which shows that the control of the three processes focused on in this work (PMMA spin, align/expose, and PMMA develop), while important in avoiding parametric gain slope yield loss, are but a few of a great many considerations that MWTD, like most manufacturing environments, must continually balance in order to meet the customer's needs. This figure shows how this research has focused on but one branch of what would be a very full tree if all other branches were considered in equal detail (in the diagram, other branches have not been drawn out, to avoid undue complexity; however, one might expect that avoiding, say, "visual yield loss" would require an equivalently heavy branch). Clearly, this environment is a complex one, that demands attention to a wide array of small details in order to meet the customer's needs.

In such a complex environment, correct and available information is often the key to progress; however, such information always comes at a price. In the case of this research, approximately four man-months of work was necessary to arrive at the results shown. While this information is quite relevant to the problem of efficient resource allocation, it does not represent a sufficient set of information to make an efficient choice. More information is needed, such as the investment required to improve control of one or more of the six primary causes of in-process variability. MWTD has previously invested in better controlling four of these six and is thus generally aware that better control of an input, even once that input is identified, is far from trivial. Furthermore, even the information provided in this research is not clear. While most of the variability observed appears to be explained in a seemingly reasonable manner, the error margins that go with these calculations are quite large. While future work could make better use of advanced experimentation techniques, this is a limited solution at best. The resolution of available measurement tools is insufficient in some cases to provide accurate measurement. The variation present in the environment, relative to the tight control levels required for high process yield, clouds experimental results. Finally, the use of improved experimentation techniques or the correction of these last two problems requires a considerable investment in and of themselves. Such investments in enhancing capability, while useful in the long term, are often difficult to make in the short term due to the lack of immediate return.

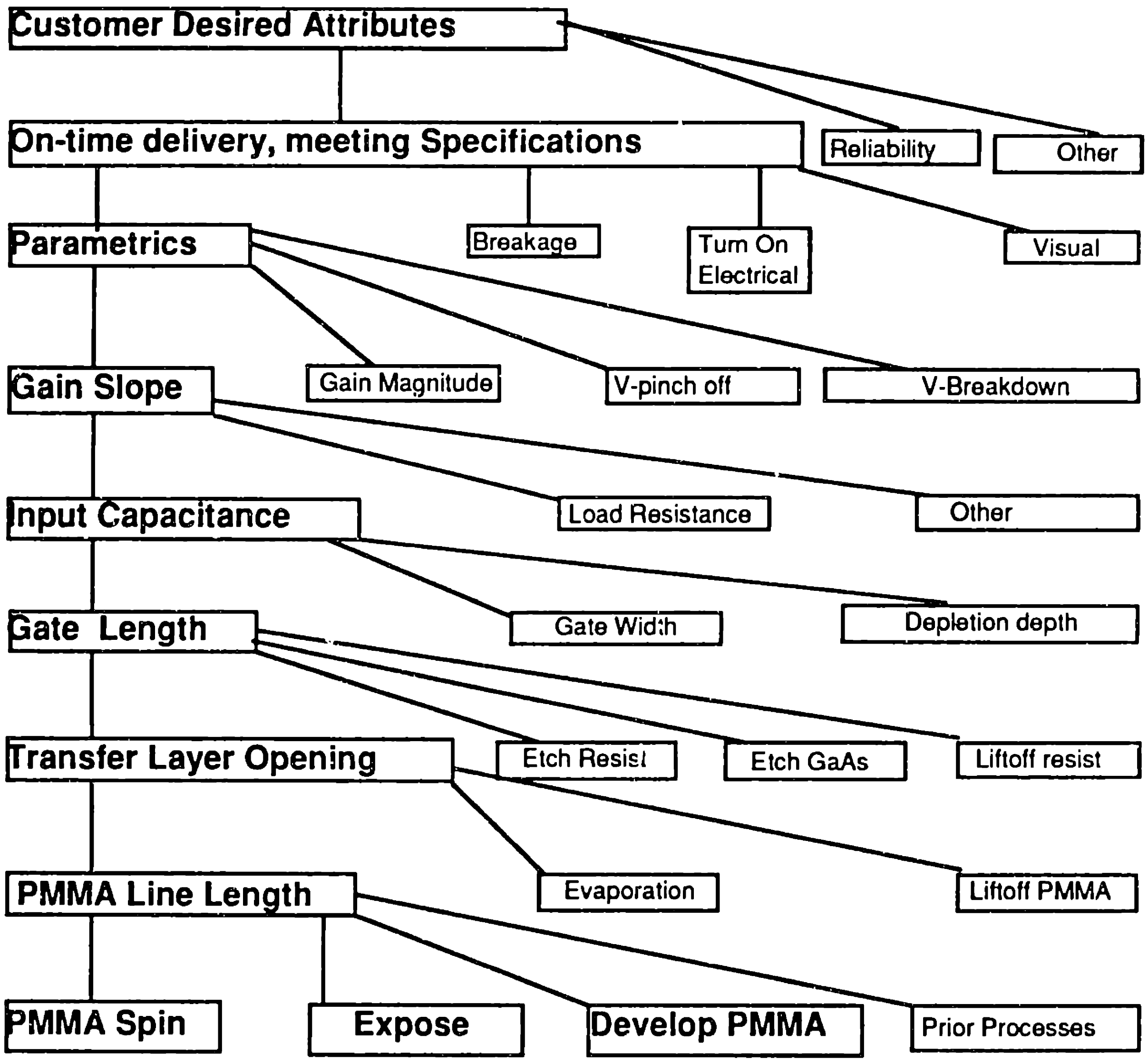


Figure 5.31: Breakdown of what information has been investigated in this research relative to delivering the customer's desired attributes in products made by MWTD.

Chapter 6 Summary and Conclusions

This thesis has shown that the pattern of allocations towards projects that improve process control at HP's Microwave Technology Division seems to be one factor that leads to two limitations in their ability to compete over the long term:

- higher risk of failing to achieve strategic objectives because current resource allocations to process control will not necessarily lead to needed changes in that area.
- possibly inefficient use of resources- due to reliance on a system that does not search for projects with the highest return and relies heavily on the perceptions of a single supervisor.

While it cannot be stated with certainty, it would appear that these limitations are the result of a sub-optimum pattern of investment that results, in part, from ignoring <u>some</u> critical variables regarding the benefits of process improvement.

6.1 Key Conclusions

This work has endeavored to explore in greater depth one finding of previous work as to why large corporations do not allocate resources efficiently: organizations ignore certain critical variables - especially after a large environmental shock to the organization - in allocating resources. This study has found that, in a particular industrial situation, the findings of the literature appear to hold true. In particular, MWTD fails to consider all of the benefits of process improvement, does not accurately estimate the costs involved with controlling this variability and does not accurately identify the sources and magnitudes of process variability. I have shown that process improvement offers significant benefits from many sources: fabricating more circuits from a fixed capital base, reducing line support cost, increasing potential market size by getting to market sooner, providing better customer support in terms of on-time delivery, maintaining flexibility to deal with changes in demand or new products, and increasing the speed of learning which ultimately reduces other costs. Furthermore, I have shown that one piece of this information - identification of the sources and magnitudes of process variability - can be made discreetly available to the resource allocator at a relatively nominal cost.

This increased information suggests a modified pattern of allocations to process improvement from the current pattern; however, given the other barriers to efficient resource allocation, especially the static allocation of manpower between functional groups, it is not clear that this additional information is sufficient to greatly improve the current allocation system.

This suggests that paying attention to critical information is but one of several changes needed to improve the current system for allocating resources. Thus, taken alone, this research is of little value in improving current methods. But by paying attention to all of the critical variables and making other modifications to the resource allocation system at MWTD, see section 6.3, one can produce potentially significant effects in this industrial environment:

- increased certainty of meeting strategic goals by necessarily allocating sufficient resources to process improvement,
- increased efficiency of allocating resources to those projects that offer the greatest return, and
- reduced risk of depending on the perceptions of a single supervisor.

6.2 Summary of Findings

In Chapter Three, it was shown that MWTD's resource allocation process is characterized by the following scenario:

Process improvement project options are originated by the process engineering supervisor on the basis of his perception of where the greatest process yield loss is occurring. Ultimate decision approval on process improvement projects is maintained at a very low level; for process improvement projects, much of this responsibility rests with the process engineering supervisor; However, the supervisor is constrained to allocating manpower from his current staffing level.

"Fire-fighting" on the production line remains a top priority for process engineers. After this obligation is met, the supervisor allocates engineering time between process improvement and new process introduction - as he perceives the respective need. The primary decision criteria between process improvement projects seems to be perceived benefit in terms of increased yield or removal of a production blockade; notably, a low financial return does not appear to affect project approval. Thus, the ultimate criteria for approval of process improvement projects is typically manpower availability and perceived yield loss.

In Chapter Four, it was shown that this resource allocation process appears to suffer from many of the same problems that have previously been described in the literature:

1) MWTD's current system does little to insure that all important options are considered and tends to emphasize improvement of areas that have recently caused highly observable trouble, rather than areas that offer the highest long term value.
2) MWTD's current allocation system only lightly assess the costs, benefits and risks associated with various project options. In particular, benefit measurement is largely limited to reducing the believed, although not measured, yield loss of a process step and thereby increasing the throughput without additional investment in capital. Other possible benefits, such as increased manufacturing predictability or faster organizational "learning", are not often considered.
3) The system is successful at pushing decision making power on allocations to a very low level, typically the first line supervisor. This allows decisions to be made by one who has experiential knowledge of the various project options that can, to a large degree, compensate for the very vague information that has been formally gathered on various options. However, such a system has two distinct limitations. This decision method relies heavily on the skills, knowledge, and biases of the particular supervisor. Second, a particular group supervisor does not have equal experiential knowledge of other

types of allocation options; hence, the supervisor cannot accurately tradeoff resources between differing types of allocation options.

4) Manpower does not appear to be dynamically allocated, rather it is largely static at current staffing levels. Hence, the organizations dynamic response to new allocation options or competing options is limited.

At this point I began to focus more narrowly on one key problem addressed above in item number two: not observing, measuring, and incorporating certain critical variables into the decision. It was shown that three variables must be known to estimate the return of a given process improvement project in a manner that is useful as a decision criteria: expected profit from a given change in yield, expected yield improvement per project and the respective resource required per project.

While the resource investment required per project is left to future research, exploration of the other two values was conducted. Again, it was shown that MWTD ignores certain substantial benefits of process improvement in their implicit calculation of increased benefits: increased manufacturing predictability, faster learning, and faster time to market. Furthermore, they cannot evaluate the expected yield improvement per project, and they tend to emphasize process targeting and reaction to sudden yield problems over proactive variability reduction to improve yield. It was shown that the yield improvement per project can be estimated from knowing where in the process circuit performance variability arises and in what magnitude at each point. Measurement of this variability required significant experimental work and an understanding of the device and process physics involved. The variability in the particular figure of merit studied, gain slope, was found to arise primarily from six process inputs:

- particles on the wafer during contact gate lithography
- variation in intra-wafer CD
- accuracy of mean mask CD
- developer exhaustion due to loading effects
- developer bath temperature
- the method used for targeting develop time

These six account for 80% of parametric electrical yield loss due to gain slope variation. It was noted that while control of three of these inputs had been recently or are already were being worked on at MWTD, three of the six are not currently under focus at MWTD. While the high error estimates associated with the effects of these inputs and the existence of other important yield loss mechanisms (breakage, visual, DC-electrical) obscures a definitive conclusion, it seems that the current allocation system, has failed to allocate resources to those projects that could best attain the substantial benefits of process improvement. Finally, this work shows that one key piece of the information necessary to accurately make tradeoffs between process control projects and other resource investments can be obtained and that the benefits of process improvement projects can be substantial.

6.3 Recommendations for Future Work

The findings of section 6.1 suggest a few possible directions for future research. In regards to solidifying the results presented thus far:

1) Repeat effect experiments to reduce high error estimates for those variables that appear to have the greatest effect on gain slope variability.
2) Apply analogous experimental techniques to the other major yield loss mechanisms: wafer breakage, visual discrepancies, and DC electrical test failures.

In furthering the incorporation of explicit information into the resource allocation decision at MWTD:

1) Explore the relative cost of reducing or eliminating each of the causes of process variability. While understanding the causes and magnitude of variability is one factor needed to identify those process improvement projects that offer the highest return from resources invested, it ignores the investment required to control the variability identified. Ultimately, the return per resource invested in a process improvement project is needed to allow objective comparison between process improvement projects and other types of allocation options.
2) Further define the value of a given increase in yield. While I demonstrated that yield increase provides several significant benefits, the link between these values and some defined increase in yield was not made explicit. As with the cost of reducing the causes of variability, this relationship is needed

to allow an objective statement of return from a given investment of resources for a process improvement project.

3) Identify methods that better deal with or change the current static allocation of manpower because an explicit assessment of return per resource invested in a process improvement project is a necessary, but not sufficient, requirement for making more efficient allocations in an environment such as this.

Finally, in extending these findings to other settings, ground level work is still needed to examine other industrial settings to find whether critical variables are ignored. As situations are observed in which they are ignored, one must identify the variables and why more attention is not paid to these variables. Finally, one must identify to what degree these variables are linked with the technology of the industry, as this link appears to be an invisible, yet unavoidable one in the MWTD environment.

Appendix 1: Summary of Secondary Process Inputs

The experimental approach for all 22 inputs studied is summarized in Figure A1.1. The entries in the table are categorized as follows:

- Input - name of one of the 22 process inputs under study, arranged by process (PMMA spin/bake, expose/align, develop)
- Variability - Procedure used to estimate variability of process input: observation of the process, experimentation, data from the manufacturer's literature, or analytical calculation.
- Effect - Method used to estimate the effect of variability in a process input on L_p: experimentation, correlation of existing process data, data from the manufacturer's literature, or analytical calculation.
- High, Medium, and Low - Settings for the effect experiment, if appropriate. Mean was typically used as center point for experiment. "NM" refers to mean not measured
- Units - of the input variable.

After the summary of experimental setup, each of the 16 process inputs found to have a relatively minor effect on L_p variability is briefly discussed.

Input		Variability	Effect	High	Mean	Low	Units
PMMA Spin							
Spin Speed	1	Expt.	Expt.	5200	4185	3200	rpm
Viscosity	2	Liter.	Liter.	----	3 e-4	----	ft^2/s
Spin Time	3	Expt.	Expt.	45	30.6	15	seconds
Off-Center	4	Process	Expt.	10	0.9	0	mm
Volume	5	Process	Expt.	10	2.7	1	cc
Temp.	6	Process	Liter.	----	21.9	----	°C
t-post ash	7	Process	Expt.	3600	600	0	seconds
Exposure							
Intensity	8	Expt.	Expt.	Same as "expose target"Experiment			mW/cm2
Contact	9	Process	Correl	4	0.7	0	# prtcls
Expose Target	10	Analyt.	Expt.	14	12.0	10	J/cm^2
Intra-mask CD	11	Process	Correl	800	460	200	Å
Mask Accur.	12	Process	Correl	5500	4800	3700	Å
Develop							
Mixture	13	Process	Expt.	175	150	125	mL
t-develop	14	Process	Expt.	150	120	90	seconds
Loading	15	Process	Expt.	20	N.M.	0	# wafrs
Develop Cycle	16	Process	Expt.	8	1.7	1	# Cycles
Devlpr. Aging	17	Process	Expt.	4	~2	0	weeks
T-bath	18	Process	Liter.	----	20	----	°C
T-wafer	19	Process	Analyt.	----	20	----	°C
Humidity	20	Process	Liter.	----	36	----	%RH
Agitation	21	Process	Expt.	All	vary	None	Yes/No
Target Meth.	22	Process	Analyt..	----	----	----	Å

Figure A1.1: Summary of Experimental Setup

A1.1 The Effect of PMMA Spin Process Inputs on Variation in L_p

None of the inputs to the PMMA process was found to have a significant impact on variability of L_p.

All PMMA process inputs contribute to producing the key topographical measure of this process step, PMMA thickness (X), which has a nominal value of 5000 Å; however, by exposing wafers with varying PMMA thicknesses, in the experimental region, it was found that variability in PMMA thickness has a muted effect on variability in L_p.

$$\frac{\partial L_p}{\partial X} \cong -0.93 \frac{Å(L_p)}{Å(X)} \qquad \text{A1.1}$$

This relationship is to be expected because a thicker layer of PMMA will allow more sideways encroachment of the exposure energy and, later, of the develop solution, resulting in smaller L_p. This effect largely matches that found in prior work.[51]

It has generally been assumed that variation in any reasonable PMMA spin/bake production process was of little relevance to overall process control. However, in looking towards the future, it is interesting to note that the effect of some PMMA Spin process inputs follow closely after the primary causes of wafer-to-wafer variation already discussed, developer exhaustion due to wafer loading and variation in develop bath temperature. In particular, a wafer being off-center on the chuck during the spin process was found to contribute 83 Å to wafer-to-wafer variation in L_p.

The experimental results for the seven PMMA inputs are summarized in Figure A1.2, after which each of the experiments and the results found are briefly discussed. In the row of this figure titled "Effect on X", the effect of variation in the given process input on the thickness of the PMMA layer is presented. This effect is multiplied by the 0.93 Å/Å value already calculated for the effect of variation in PMMA thickness on L_p.

		Unit	1	2	3	4	5	6	7
	Unit		rpm	ft^2/s.	sec.	mm	cc	°C	min.
Input Varia-tion	Die-to-Die	---	0	0	0	0	0	0	0
	Wafer-to-Wafer	---	34	0	.42	3.58	2.5	0.5	10
	Batch-to-Batch	---	26	0.01	0	0	1.1	0.23	25
Effect on X	$\partial X/\partial$ Input	Å/input units	.563	15220	.175	25	8.1	42.6	0.87
Effect of X on L_p	$\partial L_p/\partial X$	Å / Å	0.93	0.93	0.93	0.93	0.93	0.93	0.93
L_p varia-tion	Die-to-Die	Å	0	0	0	0	0	0	0
	Wafer-to-Wafer	Å	18	0	0.07	83	19	20	8
	Batch-to-Batch	Å	14	85	0	0	8	9	20

Figure A1.2: Summary of Results for Secondary Spin process inputs

1) Velocity: The rotational spin speed of the chuck sets the angular acceleration experienced by the PMMA on the wafer surface. After covering with PMMA, wafer is spun for 30 seconds at a mean speed of 4000 rpm. This should be, together with PMMA viscosity, the primary determinant of PMMA thickness.

Wafer-to-wafer variability is, presumably, induced by small fluctuations in the drive mechanism and circuitry because no settings are changed between wafers within a batch. Batch-to-batch variability is induced by hysterisis and positioning error when spin speed is set before each batch.

Wafer-to-wafer variability was estimated by measuring spin speed using a stroboscope during several spins at a constant speed setting. Batch-to-batch variability was estimated by measuring spin speed after each resetting of the spin speed. The effect of spin speed on PMMA thickness was estimated by spinning at different speeds from 3200 to 5200 rpm.

The results suggest that spin speed contributes only minimal variability to L_p, less than 20 Å.

2) Viscosity: Viscosity of PMMA may change due to age, exposure to air or light and bottle/batch of PMMA. These effects should lead to batch-to-batch variability in L_p. Viscosity, together with velocity and spin time will set final PMMA thickness.

Manufacturer's literature was used to estimate variability in viscosity, ± 2% or 1 s.d. of 0.006 ft^2/s, and the effect of viscosity change on PMMA thickness, 15220 Å/ (ft^2/s).

This data suggests that viscosity variation causes only batch-to-batch variability, which should be "controlled away" through develop time targeting. Even then, the effect on L_p variability is small (85 Å) compared to several other inputs.

3) Spin Time: An auto-shutoff circuit is used on the spinner to end the spin cycle after a mean spin time of 30.6 seconds. Due to variations in the timing circuit and spinner machinery, spin time varied from wafer to wafer.

Wafer-to-wafer variability was estimated by timing several spin cycles. The effect of spin time on PMMA thickness was estimated by spinning for different times from 20 to 40 seconds.

The results suggest that spin speed contributes only minimal variability to L_p, less than 1 Å.

4) Off-Center Spin: Placement of wafers on the vacuum chuck is done by hand. Wafers are often placed and then spun significantly off-axis, up to 1 cm off on a 2 inch (5.08 cm) wafer.

Wafer-to-wafer variability is induced by operator placement of wafers relative to the center of the spin chuck.

Wafer-to-wafer variability was estimated by measuring placement relative to center immediately before spinning. The effect of on-axis wafer spinning on PMMA thickness was estimated by spinning at different settings relative to center, from on center to 10 mm off center.

The results suggest that on-axis wafer spinning contributes only minimal wafer-to-wafer variability to L_p, 83 Å. Notably, however, this ranks as the largest cause of wafer-to-wafer variability after the two primary causes previously identified.

5) PMMA Volume: Wafers are covered with a thin layer of PMMA which is allowed to flow across the entire surface of the wafer before beginning the spin cycle. The method used for

covering the wafer and, hence, the volume of PMMA used to cover the wafer varies with operator, leading to wafer-to-wafer variation.

Wafer-to-wafer and batch-to-batch variability were estimated by measuring PMMA used, immediately before spinning, on each wafer within several batches, then separating out the respective variabilities. The effect of PMMA volume on PMMA thickness was estimated by coating with different volumes of PMMA, from 1 cc to 10 cc.

The results suggest that PMMA volume contributes to only minimal wafer-to-wafer variability in L_p, less than 19 Å, and even less batch-to-batch variability, 8 Å.

6) PMMA Temperature: Temperature variations in PMMA lead to changes in PMMA viscosity not accounted for under input #2 above, and, hence, changes in PMMA thickness. The temperature of the PMMA should be approximately equal to room temperature as the PMMA is stored within the work environment, thus room temperature measurements were used to approximate PMMA temperature. However, because of the large thermal mass of the PMMA bottle, variations in room temperature should overestimate variations in PMMA temperature.

Wafer-to-wafer variability was estimated using fab environmental data taken at regular intervals in the same bay where PMMA is stored and used. The effect of PMMA temperature on PMMA thickness was estimated from manufacturer's data for the effect of temperature change on viscosity times the effect of viscosity change on PMMA thickness, estimated in input #2.

The results suggest that PMMA temperature contributes to only minimal wafer-to-wafer variability in L_p, 20 Å, and even less batch-to-batch variability, 9 Å. The seemingly larger value for wafer-to-wafer variability is probably an artifact of the poor choice of an estimator for PMMA temperature, surrounding air temperature. Thermal lag suggests that batch-to-batch variability should be somewhat larger than wafer-to-wafer variability.

7) Time-post ash: The wafer is run through an asher to remove volatile organics immediately prior to spinning on PMMA. The process heats the wafer significantly, so the wafer is allowed to cool for a nominal time of ten minutes; However, due to the operator's need to process several batches at one time, this 'cool time' is not rigidly adhered to resulting in batch-to-batch variability. Furthermore, once a batch is started, the operator must sequentially spin each wafer and may be interupted between wafers leading to wafer-to-wafer variability.

Wafer-to-wafer and batch-to-batch variability were estimated by noting the time between completion of ashing and commencement of spinning for all wafers within each of several batches, then separating out the respective variabilities. The effect of waiting after the ash process was estimated by waiting different times before spinning, from 0 time to 1 hour.

The results suggest that waiting after the ash process contributes only minimal wafer-to-wafer and batch-to-batch variability in L_p, 8 Å and 20 Å, respectively.

A1.2 The Effect of Align/Expose Process Inputs on Variation in L_p

The experimental results for the two minor process inputs of the align/expose process are summarized in Figure A1.3, after which both of the experiments and the results found are briefly discussed.

		Unit	8	10
	Unit		mW/cm²	J/cm²
Input Variation	Die-to-Die Wafer-to-Wafer Batch-to-Batch	--- --- ---	0.07 0.24 0	0 0 2.34
Effect on L_p	$\partial L_p/\partial$ Input	Å/input units	125	100
L_p variation	Die-to-Die Wafer-to-Wafer Batch-to-Batch	Å Å Å	9 30 0	0 0 234

Figure A1.3: Summary of Results for Secondary Align/ Expose process inputs

8) Exposure Intensity: Exposure intensity is monitored using an exposure intensity analyzer. This process input considered only the variations observed using this exposure analyzer system, however imperfect it might be. (Inaccuracies in exposure targeting, due in part to the exposure intensity analyzer, were considered as part of the next process input, exposure targeting method.)

Exposure intensity is measured at the center of the exposure field immediately before processing each batch of wafers. The measured value of intensity is then used to set total exposure time to deliver a fixed dose of light energy. Intra-field variation in intensity leads to intra-wafer (die-to-die) variation in L_p. Time variation in intensity leads to wafer-to-wafer variation in L_p, although the long exposure time should lead to some averaging.

Die-to-die and wafer-to-wafer variability were estimated by measuring the intensity at five points across the exposure field 5 times during each of 3 days, then separating out the respective variabilities. The effect of variation in exposure intensity was estimated by exposing wafers for different lengths of time to change the cumulative dose delivered, from 10 to 14 J. This last experiment served the dual purpose of measuring the effect of the exposure targeting method.

The results suggest that variation in exposure intensity -as measured by the exposure intensity analyzer currently used - contributes only minimal die-to-die variability in L_p, 9 Å. Furthermore, over the comparatively short time periods required to expose a wafer (≈ 20 minutes) or expose a batch of 5 wafers (≈ 150 minutes), wafer-to-wafer variability was also found to be small, 30 Å.

10) <u>Exposure targeting method:</u> As above, an exposure analyzer is used to measure the intensity of the illumination at the wafer surface immediately prior to running each batch. The sensor used for this measurement is nominally centered in its

responsiveness around 240 nm, which limits the sensor to observing only a small fraction of the exposure dose emitted by the bulb in the contact lithography system. Furthermore, the PMMA that is being exposed is responsive over yet another spectrum of frequencies. Additionally, bulb emission changes in a non-linear fashion with respect to frequency over time. This biased information is then used to correct for exposure time and achieve a constant dose. Hence, one expects that the useful energy dose received by the PMMA varies over time because of the use of biased information not accounted for in input #8 and despite the targeting method used in input #8 to keep the nominal dose setting constant, resulting in batch-to-batch variation in L_p.

This variation was estimated by examining the frequency response curves of the analyzer sensor and the PMMA relative to manufacturer's data on frequency based bulb degradation. (see Figure A1.4)

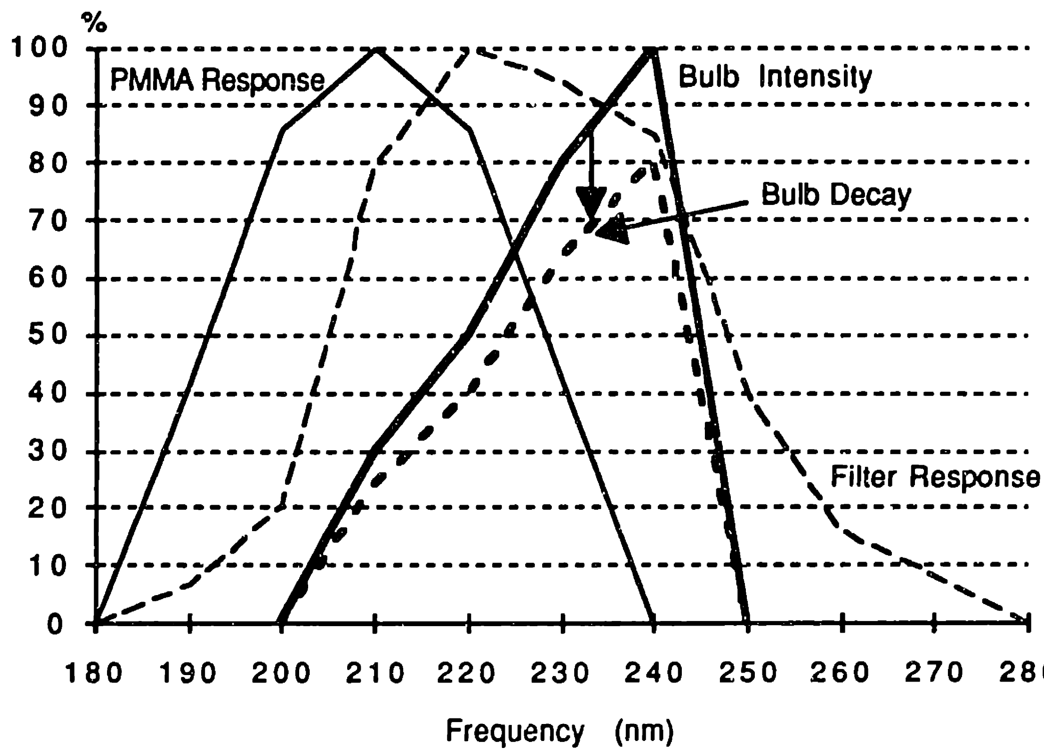


Figure A1.4: Contrast of Bulb intensity, PMMA response and Sensor response over the UV spectrum.

As the bulb intensity varies over time, the sensor observes too small of a change in intensity, leading to under-exposure of future batches. Straight line approximations to the above curves allow estimation of a nominal batch-to-batch variability by observing the change in intensity transmitted by the filter relative to the change in intensity experienced by the PMMA. This results in L_p variation of 234 Å which would be a highly important cause of L_p variability were it not for the assumption that develop time targeting "corrects away" this variability.

A1.3 The Effect of Develop Process Inputs on Variation in L_p

The experimental results for the seven minor process inputs of the PMMA develop process are summarized in Figure A1.4, after which each of the experiments and the results found are briefly discussed. The second row, "Effect on time" presents the effect of that input on total time experienced by a wafer, if applicable. This figure is multiplied times the nominal develop rate of 8.6 Å/s (found in the "Effects of t(ime)/r(ate) on L_p" row), calculated at the nominal develop time of 120 seconds to calculate an overall rate. For those inputs that effect develop rate rather than develop time, the effect on develop rate is calculated in the "effect on rate" row. As before, this figure is multiplied times the nominal develop time, 120 seconds, found in the "Effects of t(ime)/r(ate) on L_p" row.

		Unit	13	14	16	17	19	20	21
	Unit		cc	sec.	# cycl	weeks	°C	%R.H.	Ys/No
Input Varia-tion	Die-to-Die	- - -	0	1	0	0	0	0	No
	Wafer-to-Wafer	- - -	0	0	0.58	0	1	.05	No
	Batch-to-Batch	- - -	15	34.1	0.58	1.5	0.42	2	Yes
Effect on time	∂ time/ ∂ Input	sec/input units	- - -	1	3	- - -	- - -	1	- - -
Effect on rate	∂ rate/ ∂ Input	Å/s /input units	0.07	- - -	- - -	1.1	0.1	- - -	1.773
Effect of t/r on L_p	$\partial L_p/\partial$ time or $\partial L_p/\partial$ rate	Å/sec or Å/ (Å/s)	120	8.6	8.6	120	120	8.6	120
L_p varia-tion	Die-to-Die	Å	0	9	0	0	0	0	0
	Wafer-to-Wafer	Å	0	0	15	0	12	4	0
	Batch-to-Batch	Å	126	293	15	198	5	17	213

Figure A1.4: Summary of Results for Secondary Develop process inputs

13) Develop Mixture: The develop solution is a mixture of MIBK and Isopropyl alcohol (Iso). Nominally, the ratio is 1:1 for a total volume of solution of 300 mL; however, this solution is mixed

by operator's pouring the two components from large bottles directly into the solution beaker. This tends to be an inexact process and can result in large variations in the composition of the develop solution.

Because MIBK has a considerably greater develop rate than Iso, a small change in the mixture can greatly effect the develop rate leading to batch-to-batch variability in L_p.

Batch-to-batch variability in mixture was estimated by observing the amount of MIBK used in each of several batches. The effect of develop mixture on L_p was estimated by developing wafers from the same batch in develop mixtures that contained different concentrations of MIBK:Iso, from 175:125 to 125:175.

The results suggest that develop mixture variation contributes a nominal batch-to-batch variation in L_p of 126 Å, which would be an important cause of L_p variability were it not for the assumption that develop time targeting "corrects away" this variability.

14) Develop Time: Develop time is nominally set to 120 seconds. However, the actual time of exposure to the develop solution can vary somewhat by die, by wafer, and by batch. Die-to-die variation occurs due to the wafer being slowly lowered in and out of the develop solution such that the bottom of the wafer spends slightly more time in the solution than the top. Wafer-to-wafer variation occurs because wafers are dried sequentially after being quenched in an Iso bath to slow the

reaction. This drying process leaves some wafers exposed to Iso slightly longer than others. However, the develop rate in Iso alone is an order of magnitude less than that of an MIBK:Iso 50:50 mixture; therefore, wafer-to-wafer variability was assumed equal to 0. Batch-to-batch variation occurs because batches are deliberately processed for different lengths of time, where the develop time is targeted to account for and "control away" other causes of batch-to-batch variation.

Die-to-Die variability was estimated by observing the entrance to and removal from the develop solution of five batches of wafers. When wafers underwent multiple dips in MIBK:Iso, the times for each develop cycle were added together. Batch-to-batch variability was calculated by noting the total develop time of each of these five batches. The effect of develop time on L_p was estimated by developing wafers from the same batch for each of two batches for different lengths of time, from 90 to 150 seconds. This effect was also used to calculate the effect of variation in the number of develop cycles and of variation in humidity.

The results suggest that develop time contributes only minimal die-to-die variability to L_p, 9 Å. Batch-to-batch variability nominally appears to be high; however, this variability is precisely what allows the targeting method to compensate for other sources of batch-to-batch variability and hence this value has already been implicitly accounted for in the consideration of the develop time targeting method presented in section 5.5.6.

16) Number of development cycles: The targeting process sometimes requires wafers to undergo multiple develop cycles in iterating towards an acceptable L_p. This induces batch-to-batch variability if a lag exists in starting or stopping the chemical develop reaction upon repeated entry to and removal of the wafer from the develop solution.

Batch-to-batch variability was estimated by observing the number of develop cycles of several batches of wafers. The effect of develop cycles on L_p was estimated by splitting a single batch of wafers and developing them for the same total length of time, but over a varying number of develop cycles, from 1 to 4.

The results suggest that variation in develop cycles contributes only minimal batch-to-batch variability to L_p, 15 Å.

17) Developer aging: Many chemicals change their effectiveness over time. For some chemicals there is a "shelf life" effect, while for others, chemical effectiveness changes only after the bottle has been opened in production. Only the latter of these was measured. This changing effectiveness, if it effects the develop rate of the solution, leads to batch-to-batch variability in L_p.

Batch-to-batch variability was estimated by observing the length of time that a bottle of MIBK is used in the fab from the time it is first opened. The effect of developer aging on L_p

was estimated by developing wafers from the same batch for each of two batches in identical developer concentration solutions, using MIBK from bottles that had been opened for different lengths of time, just opened to 4 weeks in the fab.

The results suggest that developer aging contributes nominal batch-to-batch variability to L_p of 198 Å, which would be a highly important cause of L_p variability were it not for the assumption that develop time targeting "corrects away" this variability.

19) Wafer temperature: While it is well established that variations in developer temperature greatly effect develop rate, the effect of wafer temperature upon entry into the developer has not typically been considered in the past. During the expose process, the wafer is heated substantially by the exposure energy. it is assumed that heat transfer from the wafer to the developer can cause localized heating of the develop solution, thereby increasing the develop rate and causing wafer-to-wafer or batch-to-batch variability in L_p, depending on how long a wafer and/or a batch waits between expose and develop.

Wafer-to-wafer and batch-to-batch variability in wafer temperature were estimated by observing wafer temperature immediately before develop for each wafer in each of five batches of wafers and then decomposing the respective variabilities. The effect of wafer temperature on L_p was estimated by analytically calculating the wafer to developer heat transfer rate and then multiplying times the already known figure for the effect of changes in developer bath temperature.

The results suggest that wafer temperature contributes only minimal wafer-to-wafer and batch-to-batch variability to L_p, 12 Å and 5 Å respectively.

20) Humidity: Humidity in the cleanroom varies over time due to a variety of environmental factors; however, the cleanroom air conditioning system attempts to control this variability. Any variability that does occur can affect the development process by changing the rate at which the develop solution volatizes and/or changing the rate at which the reaction stops after the wafer is removed from Iso, thereby causing wafer-to-wafer and batch-to-batch variability.

Wafer-to-wafer and batch-to-batch variability were estimated using fab environmental data taken at regular intervals in the fab where the develop process is carried out. Since humidity is not currently measured in the exact aisle where the develop step is performed, data from a nearby aisle was used. The effect of humidity on develop rate was estimated from manufacturer's data for the effect. This figure was then multiplied by the nominal develop time to estimate the effect of humidity change on L_p.

The results suggest that humidity contributes only minimal wafer-to-wafer and batch-to-batch variability to L_p, 4 Å and 17 Å respectively. Furthermore, it is assumed that this batch-to-batch variability is "controlled away" through develop time targeting.

21) Agitation: Agitation is a known cause of variability in the develop process due to its effect on develop rate. While a specific, subjective method of agitation is prescribed in MWTD processing procedures, it is not uniformly adhered to by operators, although each operator is relatively consistent unto himself or herself in the agitation method used from batch-to-batch. This leads to batch-to-batch variability in develop rate and thus L_p.

Batch-to-batch variability in the agitation method used was estimated by observing all operators, each on two or more develop batches. The range observed ran from mild back and forth agitation to vigorous back-and-forth and up-and-down agitation. Since agitation affects develop rate, by insuring a continuous supply of active developer to the undeveloped surface through mixing and motion of the developer, an experimental range for agitation began with "none" on the low side up to the greatest develop enhancer, "vigorous back and forth, up and down agitation", on the high side.

The results suggest that agitation method contributes a nominal batch-to-batch variability to L_p of 213 Å, which would be a highly important cause of L_p variability were it not for the assumption that develop time targeting "corrects away" this variability.

Appendix 2: Experimental Error Analysis

The goal of this experimental work was to identify the contributions of various process inputs to variability in L_p. The critical measured values are summarized in a graph of variability induced in L_p broken down by process input (see Figures 5.18 and 5.21). Necessarily, the experimental methods used to calculate these figures only approximate the true values of induced variability, because of environmental noise, measurement resolution error, unaccounted for uncertainty, etc. Accordingly, an error analysis was performed using the method of Kline and McKlintock[52]. This appendix assesses the error bars around the values presented in figures 5.18 and 5.21 within which the true values of induced variability per process input should lie.

Brief Summary of Kline and McKlintock

When

a result $R = R(m_1, m_2, \ldots m_n;$ time, instrument calibrations),

and

if each m_i is

- independent
- comes from a Gaussian distribution, and
- can be described by $m_i = \overline{m}_i \pm \delta m_i$ at (ODDS)

  where "ODDS" describes the probability of enclosing the true value of m_i or R within the error range expressed

then

$R = \overline{R} \pm \delta R$ at (ODDS) where

$$\delta R = \left[\sum_{m=1}^{N} \left(\frac{\partial R}{\partial m_i} \delta m_i \right)^2 \right]^{\frac{1}{2}} \qquad \text{A2.1}$$

Kline and McKlintock found that equation A2.1 most nearly preserves the true statistical probability of combining the individual effects of each δm_i.

Some applicable simplifications used in this work:

for

$$R = \sum_{i=1}^{N} m_i \qquad \delta R = \left[\sum_{i=1}^{N} (\delta m_i)^2 \right]^{\frac{1}{2}} \qquad \text{A2.2}$$

$$R = m_1^{a_1} m_2^{a_2} m_3^{a_3} \ldots \qquad \delta R = \left[\sum_{i=1}^{N} \left(\frac{a_i R}{m_i} \delta m_i \right)^2 \right]^{\frac{1}{2}} \qquad \text{A2.3}$$

<u>Applications of Kline &McKlintock to this experimental study</u>

In estimating error, some simplifying assumptions were made:

- All measurements are taken from gaussian distributions.
- All errors are reported for one standard deviation, or a 68.3% chance of enclosing the true value. By tripling these errors, the probability of encompassing the true value increases to 99.7%.
- All of the process inputs are presumed to be independent. This assumption has been shown to be a poor one for certain of the process inputs, e.g., develop time and exposure dose are known to interact; nevertheless, the assumption is maintained to

keep the math simple and because the assumption typically does hold to first order.

Using this assumption, equation A2.2 is applicable for estimating the overall estimated variability in L_p due to the process inputs considered and reduces to

$$\delta R = \left[\sum_{i=1}^{N} (\delta \text{ induced variability}_i)^2 \right]^{\frac{1}{2}}$$

where δ induced variabil_i is calculated as outlined below.

- In Chapter 5, the induced variability was typically calculated by

variability of process input * effect of input on L_p

or, when an intermediate effect is used,

variability of process input * effect of input on X
* effect of X on L_p,

where X represents an intermediate variable such as PMMA thickness

Therefore equation A2.3 is applicable for estimating the error in the induced variability for a specific process input and reduces to

$$\begin{aligned} \partial R_i &= \partial \text{ induced variability }_i \\ &= [(\text{effect }_i * \partial \text{ variability }_i)^2 \\ &\qquad + (\text{variability } * \partial \text{ effect }_i)^2]^{1/2} \end{aligned}$$

or, when an intermediate effect is used,

$$\partial R_i = [(\partial \text{ effect }_{i1} * \text{effect }_{i2} * \text{variability})^2 + (\text{effect }_{i1} * \partial \text{ effect }_{i2} * \text{variability})^2 + (\text{effect }_{i1} * \text{effect }_{i2} * \partial \text{ variability }_i)^2]^{1/2}$$

- Error was calculated only for those types of variability induced by a given process input

Input data

Input data to the error model come from three sources:

1) Error for measurements of input variability was assumed equal to the repeatability of the measurement equipment, or

$$\partial \text{ variability }_i = \text{repeatability of the measurement tool}$$

2) Error for measurements of input effects that were determined experimentally, was measured by replicating the mid-level experimental setting (center point) as part of experimental design, or

$$\partial \text{ effect }_i = \text{standard deviation (results at center point)}$$

3) All error for measurements of input variability or input effects from manufacturer's data, when not reported by the manufacturer, were arbitrarily estimated to equal 25% of the value.

The input error data used and the resulting values of estimated error in induced variability in L_p are presented in Figure A2.2. This estimated error is expressed as a percentage of induced variability in Figure A2.3. The error figures reported are quite large, with the one standard deviation error equal to 271 - 334 % of the effect calculated as shown in Figure A2.1.

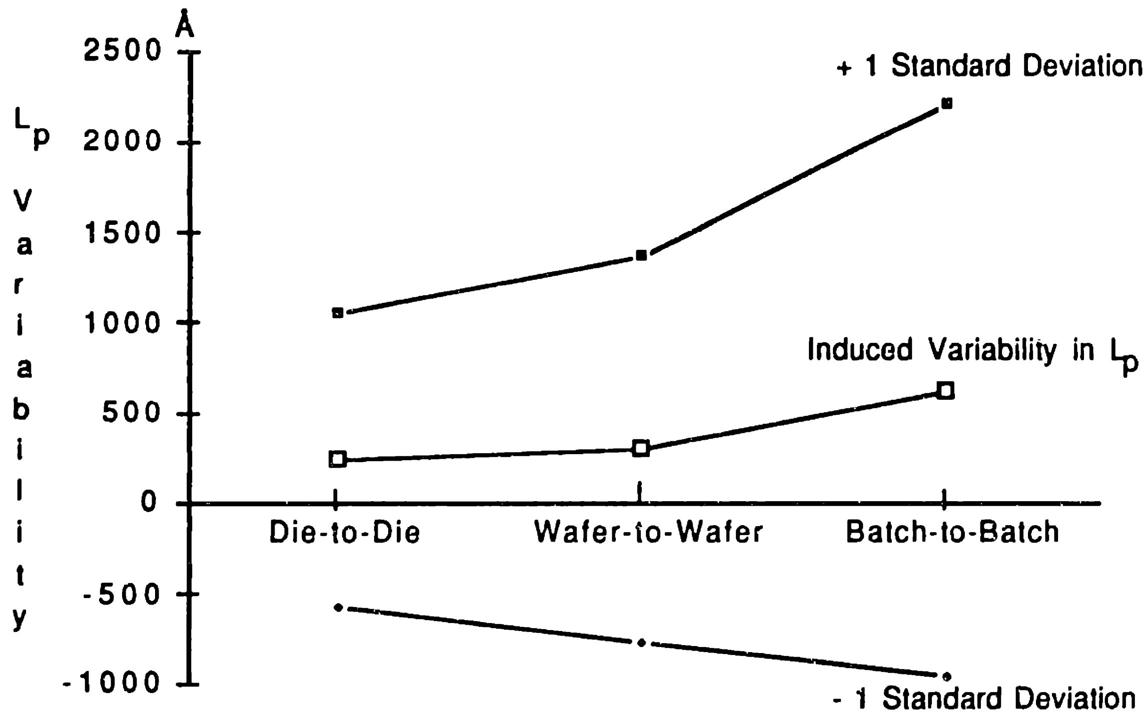


Figure A2.1: Induced variability and one sigma error range for each of three types of variability

Summary of Error	#	Input Error			Estimated Error ĉ induced variabilityᵢ		
		∂ variabiltᵢ	∂ effectᵢ #1	∂ effectᵢ #2	Die -Die	Wafer -Wafer	Batch -Batch
Input		1 s.d.	1 s.d.	1 s.d.	Å	Å	Å
PMMA Spin							
Spin Speed	1	16.7	0.02	0.92	0	20	16
Viscosity	2	0.01	3805	0.92	0	0	90
Spin Time	3	0.02	0.44	0.92	0	0.1	0
Off-Center	4	0.17	2.84	0.92	0	83	0
Volume	5	0.17	0.91	0.92	0	19	8
Temp.	6	0.5	10.65	0.92	0	20	9
t-post ash	7	0.17	0.3	0.92	0	8	21
Exposure							
Intensity	8	0.01	258.2	0	18	62	0
Contact	9	0	399.96	0	800	400	400
Expose Target	10	2.34	206.65	0	0	0	537
Intra-mask CD	11	70	0.09	0	52	0	0
Mask Accur.	12	70	0.49	0	0	0	322
Develop							
Mixture	13	4.17	0.07	0	0	0	125
t-develop	14	0.17	6.67	0	7	13	227
Loading	15	0	0.33	0	0	192	208
Develop Cycle	16	0	9.29	0	0	46	46
Devlpr. Aging	17	0	1.71	0	0	0	308
T-bath	18	0.17	1.33	0	117	117	111
T-wafer	19	0.17	8	0	0	960	960
Humidity	20	0.02	25	0	0	17	18
Agitation	21	0	6.84	0	0	0	821
Target Meth.	22	0	0	0	0	0	375
Total		- -	- -	- -	810	1071	1587

Figure A2.2: Results of Error Analysis, Input Error and Estimated Error in Induced Variability in L_p.

		induced variability (Å)			δ i.v.$_j$/ induced variability		
Input		D-to-D	W-to-W	B-to-B	D-to-D	W-to-W	B-to-B
PMMA Spin							
Spin Speed	1	0	18	14	0	111%	118%
Viscosity	2	0	0	85	0	0	105%
Spin Time	3	0	0.07	0	0	273%	0
Off-Center	4	0	83	0	0	100%	0
Volume	5	0	19	8	0	100%	101%
Temp.	6	0	20	9	0	102%	102%
t-post ash	7	0	8	20	0	105%	105%
Exposure							
Intensity	8	9	30	0	207%	207%	0
Contact	9	100	50	50	800%	800%	800%
Expose Target	10	0	0	234	0	0	230%
Intra-mask CD	11	123	0	0	42%	0	0
Mask Accur.	12	0	0	273	0	0	118%
Develop							
Mixture	13	0	0	126	0	0	99%
t-develop	14	9	17	293	79%	78%	78%
Loading	15	0	213	231	0	90%	90%
Develop Cycle	16	0	15	15	0	310%	310%
Devlpr. Aging	17	0	0	198	0	0	156%
T-bath	18	183	183	107	64%	64%	103%
T-wafer	19	0	12	12	0	7999%	7999%
Humidity	20	0	4	17	0	401%	103%
Agitation	21	0	0	213	0	0	386%
Target Meth.	22	0	0	375	0	0	100%
Total		- -	- -	- -	334%	355%	254%

Figure A2.3: Results of Error Analysis, Induced Variability and Estimated Error as a percentage of Induced Variability in L_p.

Error was found to be significant for all six primary causes of L_p variability, with one standard deviation error ranging from 42% to 800% of the estimated induced L_p variation. The major sources of significant error were:

- instrument resolution relative to signal being observed, e.g., intra-wafer mask CD and developer bath temperature.
- experimental setups that did not consider all sources of variability, e.g., the effect of particles on the mask during exposure.
- reliance on a noisy output, specifically L_p, e.g., developer exhaustion and mask accuracy.

Error for each of the six primary causes of L_p variability is briefly discussed below.

1) Contact - the experimental supposition failed to account for particle size or particle location relative to L_p measurement location (except within large quadrants of the wafer). Consequently, the one standard deviation error estimate of 800% is not unexpected.

2) Intra-Mask CD - the repeatability of the optical CD measurement tool was the limiting factor. Such tools are inherently limited in the sub-micron range. The repeatability of these instruments, 70 Å, is too large compared to the very small mask CD differences being observed, 123 Å, and results in a one standard deviation error estimate of 42%.

3) Mask Accuracy - Reliance on experimental data (L_p) for wafers which already have a high variance in L_p clouds the mask accuracy effect results considerably. This directly supports Bohn's[53] argument of the need for a high signal to noise ratio in experimental results in order to increase the speed of learning. The high noise in L_p in the environment, one sigma wafer-to-wafer $\cong$ 0.03 µm, relative to the experimental output signal, $\Delta L_p \cong 0.026$ µm, led to a one standard deviation error estimate of 118% of the induced variability.

4) Developer Exhaustion - Reliance on experimental data (L_p) for wafers which already have a high variance in L_p also clouds the developer exhaustion effect results. The high noise in L_p in the environment, one sigma wafer-to-wafer $\cong$ 0.03 µm, relative to the experimental output signal, $\Delta L_p \cong 0.033$ µm, led tc a one standard deviation error estimate of 90% of the induced variability.

5) Developer Bath Temperature - The resolution of the measuring tool, a mercury thermometer, was large relative to the temperature difference being observed. This propagated through to a one standard deviation error estimate of 64 - 103% of the induced variability signal.

6) Develop Time Targeting - Because this effect was calculated analytically there is no convenient means for estimating error. Future work could experimentally determine the value of

induced variability and estimate the associated variability. For the purposes of providing more realistic "Total" error estimates, an arbitrary value of error equal to 100% of the induced variability was used in Figures A1.1, A1.2, and A1.3.

References:

1 Azzam, W.J. and J.A. del Alamo. *An All-Electrical Floating Gate Transmission-Line Model Technique for Measuring Source Resistance in Heterostructure Field-Effect Transistors.* to be published in IEEE Transactions on Electron Devices. 1990.

2 Baker, Norman R. *R&D Project Selection Methods: An Assessment.* IEEE Transactions on Engineering Management. Vol. EM-21, No. 4, November 1974, pp. 165-171.

3 Baker, N. and J. Freeland *Recent Advances in R&D Benefit Measurement and Project Selection Methods.* Management Science. Vol. 21, No. 10, pp. 1164-75, 1975.

4 Baker, N.R. and W.H. Pound. *R&D project selection: where we stand.* IEEE Transactions on Engineering Management. Vol. EM-11, No. 4, p. 124, 1964.

5 Bohn, Roger E. *Learning By Experimentation in Manufacturing.* Working Paper, Harvard Business School 1987.

6 Bower, Joseph L. *Managing the Resource Allocation Process.* Boston: Harvard Business School Press, 1986.

7 Breely, Richard A. and Stewart C. Meyers, Principles of Corporate Finance. McGraw-Hill Book Co. 1988.

8 D'Avanzo, D., D. Estreich, V. Peterson. *A 26 GHz GaAs MMIC Technology: Process Devices, Circuits, and Applications.* 1989 RF and Microwave Measurement Symposium and Exhibition. June 1989.

9 D'Avanzo, D., D.B. Estreich, T.W. Taylor, H. Kondoh, J. Orr, A. Quach, E.H. Wong, R. Yeats, V.K. Makker, and R.A. Fisher, *A Manufacturable, 26 GHz GaAs MMIC Technology.* Proceedings of the GaAs IC Symposium. November 6-9, 1988, pp.317-320.

10 Danila, N. *Strategic Evaluation and Selection of R&D Projects.* R&D Management. Vol. 19, No. 1, January 1989.

11 Dertouzos, Michale L, Richard K. Lester, and Robert M. Solow. Made In America. Cambridge, MA: MIT Press, 1989.

12 Fine, C.H. and A.C. Hax. *Manufacturing Strategy: A Methodology and an Illustration.* Interfaces. 15(6), November-December 1985, pp. 28-46. 1983.

13 Goldratt, Eliyahu M., and Jeff Cox. The Goal: A Process of Ongoing Improvement. Croton-on-Hudson, NY: North River Press, 1986.

14 Hauser, John R. and Don Clausing. *The House of Quality.* Harvard Business Review. May-June 1988. pp 63-73.

15 Hayes, Robert H., Steven C. Wheelwright, and Kim B. Clark. Dynamic Manufacturing: Creating the Learning Organization. New York: The Free Press, 1988.

16 Hannan, Michael T. and John Freeman. *Structural Inertia and Organizational Change.* American Sociological Review. Vol. 49, April 1984, pp. 149-64.

17 Higgins, J. and K. Watts, *Some Perspectives on the Use of Management Science Techniques in R&D Management.* R&D Management. Vol. 16, No. 4, 1986. pp. 291-6.

18 Kline, S.J. and F. A. McClintock, *Describing Uncertainties in Single-Sample Experiments,* Mechanical Engineering, January 1953.

19 Ladbrooke, P.H. *Reverse Modeling of GaAs MESFETs and HEMTs.* GEC Journal of Research. Vol. 6., No. 1, 1988, pp. 1-9.

20 Ladbrooke, P.H. MMIC Design: GaAs FETs and HEMTs. Boston: Artech House, 1989.

21 Lawton, James W. Wafer Release in a GaAs Fab Environment. S.M.E.E. Thesis, Massachusetts Institute of Technology, Cambridge, MA, May 1990, Chapter 7.

22 Liberatore, M. and G. Titus. *The Practice of Management Science in R&D Project Management.* Management Science. Vol. 29, No. 8, 1983, pp.962-74.

23 Mansfield 71 and 77, R&D in Modern Corporation and Production and Application of New Industrial Techniques.

24 Moreau, W. C. Semiconductor Lithography: Principles, Practices, and Methods. New York: Plenum Press, 1988.

25 Orr, Jerry. MMIC Parameter Correlations and Process Specifications. HP/MWTD Internal Memorandum. August 12, 1988.

26 Phadke, M.S. Quality Engineering Using Robust Design. MIT Draft Copy, June 22,1988, Chapter 2.

27 Quinn, J. *Strategic Change: Logical Incrementalism.* Sloan Management Review. Fall 1978. pp 7-21.

28 Scott, Bob. Final Report: Cost Projections for a XXX-wafer GaAs Fab. HP/MWTD Internal Memorandum, August 31, 1988.

29 Stivers, Paul. 0.5 µm CD Measure. HP/MWTD Internal Memorandum. November, 17, 1989.

30 Taylor, B.W, L.J. Moore, and E.R. Clayton. *R&D Project Selection and Manpower Allocation with Integer Non-Linear Goal Programming.* Management Science. Vol. 28, No. 10, 1982, pp.1149-85.

31 Vlasak, Dean A. Process Verification Test Cell for Low Volume ASIC-like MMIC Production. S.M.E.E. Thesis, Massachusetts Institute of Technology, Cambridge, MA, August 1990.

Endnotes

---

[1]Dertouzos, et al.
[2]Hayes, et al.
[3]Dertouzos, et al.
[4]Bower
[5]Breely and Meyers
[6]Fine and Hax
[7]Higgens and Watts
[8]Baker and Pound
[9]Baker and Freeland
[10]Taylor, et al
[11]Liberatore and Higgins
[12]Danila
[13]Higgins and Watts
[14]Bower
[15]Quinn
[16]Hannan and Freeman
[17]Bower
[18]Baker and Freeland
[19]Mansfield
[20]Baker and Pound
[21]Mansfield
[22]Bower
[23]Baker
[24]Higgins and Watts
[25]Hayes, et al.
[26]Bohn
[27]Goldratt and Cox
[28]Phadke
[29]Hauser and Clausing
[30]Hayes, et al
[31]MWTD Internal Memorandum
[32]Lawton
[33]Lawton
[34]Lawton
[35]Hauser and Clausing, p. 64-65.
[36]Hauser and Clausing, p. 64-73.
[37]D'Avanzo, Don, Don Estreich, and Val Peterson, pp. 12-14.
[38]Orr, p. 1.
[39]Ladbrooke, "Reverse Modelling of GaAs MESFETs and HEMTs", pp. 1-9.
[40]Ladbrooke, "MMIC Design: GaAs FETs and HEMTs", pp. 91-146
[41]Orr, pp. 2-3.
[42]Orr, pp. 2-3.

[43]Orr, Appendix E, Figure 1.
[44]Orr, Appendix C, Figure 1.
[45]Orr, pp. 1-7.
[46]Azzam, W.J., and J.A. del Alamo
[47]D'Avanzo, D., D.B. Estreich, T.W.Taylor, et al.
[48]Moreau, pp.813-855.
[49]Stivers
[50]Moreau
[51]Moreau
[52]Kline and McKlintock
[53]Bohn