

966

MEASURING MACHINE INTERFERENCE TO EVALUATE AN OPERATOR CROSS-TRAINING PROGRAM

by

Benjamin G. Goss

Sc.B. Mechanical Engineering, Brown University, 1992

Submitted to the Sloan School of Management and the Department of Mechanical Engineering in partial fulfillment of the requirements for the degrees of

Master of Business Administration
and
Master of Science in Mechanical Engineering

In Conjunction with the Leaders for Manufacturing program at the Massachusetts Institute of Technology

May 2000
[June 2000]

© Massachusetts Institute of Technology, 2000. All rights reserved.

Signature of Author _____

Sloan School of Management
Department of Mechanical Engineering
May 5, 2000

Certified by _____

Professor Sara Beckman, Thesis Advisor
Haas School of Business, University of California, Berkeley

Certified by _____

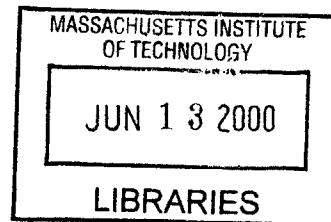
Dr. Stanley B. Gershwin, Thesis Advisor
Department of Mechanical Engineering

Accepted by _____

Margaret Andrews, Director of Master's Program
Sloan School of Management

Accepted by _____

Aih Sonin, Chairman, Departmental Committee on Graduate Studies
Department of Mechanical Engineering



ARCHIVES

MEASURING MACHINE INTERFERENCE TO EVALUATE AN OPERATOR CROSS-TRAINING PROGRAM

by

Benjamin G. Goss

Submitted to the Sloan School of Management and the Department of Mechanical
Engineering in partial fulfillment of the requirements for the degrees of

Master of Business Administration
and
Master of Science in Mechanical Engineering

Abstract

Factory capacity can only be utilized when three resources are simultaneously available: equipment, material, and operator. Over time, the variation in availability of these three resources interacts to produce an uneven distribution of inventory in the factory. This work examines the impact of cross-training on responding to factory disruptions that cause spikes in factory workload. Strategic cross-training creates a flexible workforce that can provide short term burst capacity to mitigate these disruptions. A simulation model is presented to predict where cross-trained operators offer the greatest impact on factory output. A pilot study implements the recommendations of the simulation model, and the Machine Interference metric is introduced as a metric to isolate and assess the benefits of cross-training. Pilot results indicate that cross-training appears to contribute to factory performance in staffing areas characterized by both high workload and high variability of workload. Cross-training also provides several second order benefits that are not easily linked to financial performance. An implementation plan is introduced to speed future expansion of cross-training programs.

Thesis Advisors:

Dr. Sara L. Beckman, Haas School of Business, University of California, Berkeley

Dr. Stanley B. Gershwin, Department of Mechanical Engineering

Acknowledgements

The author gratefully acknowledges the support and resources made available to him through the MIT Leaders for Manufacturing (LFM) program, a partnership between the Massachusetts Institute of Technology and major manufacturing companies, including Intel Corporation.

I would like to thank Dave Myron for championing the LFM program at Intel, setting up my internship, and creating such a positive environment for the Flexible Workforce Team; my company supervisor, Joanna Shear, who proved to be an ideal player-coach by demonstrating the ability to find an informed, practical solution to any obstacle, and helping me understand important nuances in the field of industrial engineering; our group leader, Sean Cunningham, who mobilized resources, provided critical insights, and gave helpful guidance in keeping the project focused; and my fellow intern, Brian Wolkenberg, who helped me polish my rough ideas.

Members of the Flexible Workforce Team helped me appreciate the complexity of the wafer fab and the importance of teamwork in pulling off a pilot study. Team members included: Shelly Keel, Sandra Raposa, Darren Schiele, Anne Rizzo, Rudy Gallegos, and Marlo Beltrano.

Special thanks goes out to my thesis advisors, Dr. Stan Gershwin, MIT, Department of Engineering, and Dr. Sara Beckman, UC-Berkeley, Haas Business School, who helped me bridge the sometimes daunting gap between industry and academia.

Finally, I want to celebrate my wife and classmate, Cindy, as every step in our lifelong adventure reminds me of what a perfect match I have found.

Table of Contents

Table of Contents	5
Table of Figures.....	7
Table of Tables.....	7
1 Introduction	8
Motivation for Study.....	8
<i>Understanding Capacity</i>	8
<i>Understanding Flexibility</i>	9
<i>Tactical Horizon vs. Strategic Horizon</i>	10
Literature Review.....	12
<i>Analysis of Dual Resource Constrained Production Systems</i>	12
<i>Cross-Training</i>	14
<i>New Concepts</i>	16
Factory Context	17
<i>Factory Environment</i>	18
<i>Tactical Performance Measures Already In Use</i>	20
2 Measuring Cross-Training Impact	22
Decision Making Context.....	22
<i>Effective Capacity and Expected Output</i>	23
<i>Expected Workload</i>	25
<i>Capacity Tradeoff</i>	25
Performance of a Single Tool	27
<i>Machine Interference Definition</i>	27
<i>Machine Interference Calculation</i>	28
<i>Availability-Utilization Gap</i>	29
Performance of Multiple Tools	30
<i>Staffing Areas</i>	30
<i>Extension of Single Tool Formulation</i>	32
<i>Data Collection</i>	33
3 Identifying and Simulating Pilot Areas.....	35
Selecting Pilot Areas	35
<i>Complementary Workload</i>	35
<i>Operator Availability</i>	37
<i>Operator Effectiveness</i>	38
<i>Pilot Staffing Areas</i>	39
Modeling Effective Capacity of Pilot Staffing Areas	40
<i>Modeling Approach</i>	41
<i>Model Output</i>	42
Tactical Decision Support.....	44

<i>Workload Report</i>	44
4 Results and Conclusions	46
Analysis	46
<i>Cross-trained vs. Control Shifts Across Pilot Areas</i>	47
<i>Statistical Significance</i>	50
<i>Frequency of Transfers</i>	52
Implementation	53
<i>Implementation Steps</i>	53
<i>Data Resolution</i>	54
<i>Keys to Success</i>	55
Conclusions	57
<i>Cross-training Benefits</i>	58
<i>Data Collection and Performance Evaluation</i>	59
Future Direction	60
Appendix A	61
<i>Model Inputs</i>	61
<i>Model Outputs</i>	61
Bibliography	62

Table of Figures

Figure 2-1 Impact of Staffing Flexibility.....	23
Figure 2-2 Expected Output vs. Number of Operators.....	24
Figure 2-3 Capacity Tradeoff.....	26
Figure 2-4 Role of Workload in Reducing Gap.....	30
Figure 2-5 Staffing Areas A and B.....	31
Figure 3-1 Queue Size Variability for Areas A and B.....	36
Figure 3-2 Simulation Results.....	43
Figure 4-1 Results for Each Pilot Area.....	48

Table of Tables

Table 1-1 Decision Making Scope and Frequency.....	11
Table 1-2 Typical Factory Performance Measures.....	20
Table 2-1 Distinct Tool States.....	28
Table 3-1 Pilot Area Selection Criteria.....	37
Table 3-2 Cross-Training Pilot.....	39
Table 3-3 Source of Simulation Inputs.....	40
Table 3-4 Characterization of Pilot Areas.....	41
Table 4-1 Results of Statistical Analysis.....	51
Table 4-2 Operator Transfers.....	52

1 Introduction

Motivation for Study

Semiconductor wafer fabrication is a capital intensive process due to the extreme technological requirements and rapid obsolescence of subsequent generations of processing equipment. Facility and equipment depreciation can contribute more than half of overall manufacturing costs. Since staffing decisions have a second or third order effect on manufacturing costs, issues of labor productivity do not receive the same priority or analytical attention that is typically given to equipment productivity. Cross-training usually happens by default rather than by design. Most decisions are executed casually or automatically without explicit recognition that a problem even exists, much less that a solution is obtained. In some cases, the lack of attention to operator resources can undermine the gains in equipment performance. This research focuses on understanding the impact of a pilot cross-training program to better utilize factory capacity.

Understanding Capacity

Factory capacity is typically viewed as a simple function of the quantity and capacity of equipment on the factory floor. From this perspective, equipment is the resource that limits factory output. Even though factory output cannot exceed equipment capacity, it is also important to recognize that three resources must be simultaneously available to turn capacity into output: equipment, material, and operator. The implication is that capacity can be further constrained by material release policies and operator staffing policies so that actual factory output in practice may fall well below demonstrated equipment capacity. Staffing policies (as well as material release policies) can be simple but effective ways to insure that expensive equipment capacity is not wasted. Operator cross-training is an example of a staffing policy that provides the flexibility to distribute operators in response to workload spikes throughout the factory.

At the time of the pilot study, the market is willing to absorb all of the production that Intel can flow through its factories. During these times of strong demand, plant capacity can be

labor constrained, where understaffing results in inefficient use of capital equipment. During periods of low demand, the factory may find itself overstaffed for the existing market conditions. This demonstrates the fundamental cost tradeoff between idle labor resources and idle capital equipment (Stafford, 1988). Additional operators increase payroll cost, while fewer operators cause lost capacity due to the inability to keep equipment running. This dilemma falls under the academic categorization of **dual resource constrained** systems (Treleven, 1989), where both equipment and operator resources constrain capacity. This occurs when there are not enough operators available to keep all of the equipment continuously running at capacity. In a steady-state system, the inability of the workforce to keep up with the production equipment will result in idle equipment. Idle equipment hurts overall system performance either by reducing capacity at the bottleneck or by starving the bottleneck operation. By understanding the interaction of staffing decisions with equipment capacity in this dual resource constrained environment, overall factory performance can be improved.

Understanding Flexibility

In the complex and dynamic environment of semiconductor wafer fabrication, there is a continuing drive to be organizationally flexible to keep pace with both technological advances and global competition. Shorter development lifecycles and increasing competitive pressure require a more agile manufacturing environment that can meet the needs of a growing number of product segments. As a result, flexibility is gaining acceptance as one of the competitive dimensions of manufacturing strategy (Malhotra, 1990). From the shop floor perspective, short term factory performance is strongly influenced by the variability of each processing step. As Gershwin (1994) points out, there are only two possible solutions to improving performance in this environment: eliminate the source of variability or limit the disruptions caused by variability. Though both avenues should be pursued in the course of factory improvements, cross-training focuses on limiting disruptions from existing variability.

One difficulty in understanding the role of flexibility in a manufacturing operation is the choice of performance measures. Manufacturing performance measures focus on cost,

quality, resource availability, customer lead time, and level of product customization. Every day, managers balance these competing and sometimes conflicting objectives to provide both short term and long term value to key stakeholders, including stockholders, customers, and employees (Hopp, 1996). By balancing short term (tactical) objectives with long term (strategic) growth, the factory is positioned to continue delivering value to stakeholders over time. In general, factory performance measures are used to help highlight the tradeoffs involved with each management decision. Different weights will be given to different performance measures based on how the firm delivers value to customers. It is important to isolate the contribution of cross-training on each of these performance measures to quantify the impact of the cross-training program.

Tactical Horizon vs. Strategic Horizon

Labor assignment decisions cover two different time horizons. The long-term **strategic** horizon is used to prepare aggregate production plans at the plant management level, while the short-term **tactical** horizon is used to control day-to-day operations at the level of first line supervisors (Holstein, 1970 and Uzsoy, 1992). Tactical decisions are typically made within the boundaries of the strategic constraints. For example, once a staffing level for each shift is established, adding staff is no longer a tactical option. The tactical decision consists of where to best use the operator headcount that is allocated to each shift. This research focuses on quantifying and improving the short term performance of tactical decision-making. Current approaches such as simulation and queueing models emphasize steady-state solutions. However, the transient problems are of immediate consequence in shop floor control. Investigations of these transient problems often do not take into account the impact on long-term performance. As Johri (1993) notes, tactical decisions are "difficult enough to handle without having to worry about their (long-term) implications on performance measures. There are few, if any, indications of how these local decisions affect long-term goals." It is important to clearly define the scope of the tactical decisions that can be affected once strategic decisions have been locked in place.

In order to relate staffing decisions to factory results, performance metrics must match the scope and frequency of the decision-making environment. Decision-making scope can be

arbitrarily divided into strategic, policy, and tactical decisions. Decisions must be made for labor, equipment, and material resources at the appropriate decision making frequency. Examples of this categorization scheme at Intel are shown in Table 1-1. To support strategic decisions, information needs to be aggregated over long-term horizons to help support decisions on the size of the workforce (usually updated each quarter), the quantity of equipment purchased, and the number of product families offered. All of these are beyond the control of the tactical decision maker. At a more intermediate scope, policies are set to provide a rule-of-thumb for allocating resources among staffing areas, equipment, or lots based on their priority in the factory. Policies are usually the direct results of industrial engineering or other analytical studies. These policies can be overridden on a case-by-case basis, but the majority of the time the policy decision will represent the default behavior that is followed on the shop floor.

Decision Making Scope	Decision Making Frequency	Manufacturing Resource	Tactical Control	Strategic Control
Strategic	Annually, As needed	Labor: Size of factory workforce	No	Yes
		Equipment: Quantity of equipment purchased for each process step		
		Material: Number of product families		
Policy	Every quarter	Labor: Priority for staffing assignments	Policy Override	Analytically Based
		Equipment: Priority for equipment maintenance/loading		
		Material: Priority for lots to run		
Tactical	Every shift change, Every lot loaded	Labor: Beginning of shift staffing assignment	Yes	No
		Equipment: Equipment taken down for PM		
		Material: Lot loaded in tool		

Table 1-1 Decision Making Scope and Frequency

Tactical decisions are made by both supervisors on a daily basis and individual operators on a lot-by-lot basis in direct response to the factory environment. Supervisors assign operators

to cover specific equipment, and there is usually some flexibility around preventive maintenance schedules and lot loading sequences to minimize disruptions where possible. This leaves a great deal of discretion in the hands of the tactical decision maker, and very few tools (other than policies) to analytically guide decisions. For real performance improvement on a tactical level, it is important to bring some decision support information to the tactical level. The important distinction between policy decisions and tactical decisions is that policy decisions represent average or expected behavior over time, while tactical decisions represent the day-to-day and hour-to-hour decisions that will fluctuate with the dynamics of the factory environment.

Literature Review

Analysis of Dual Resource Constrained Production Systems

A dual resource constrained system has operators who are responsible for running more than one piece of semiautomatic equipment. Since the equipment is not permanently staffed, operators move between equipment as needed. This is a useful model for dealing with semiconductor equipment where operators cover several machines. Operators are expected to set up recipes, load wafers, unload wafers, and perform some preventive maintenance. Equipment that is processing does not require an operator to be present. In wafer fabs, processing times are very strictly controlled by recipes, so the non-availability of operators becomes a source of variation affecting otherwise relatively deterministic processing times (Johri, 1993). Operators are typically cross-trained to run any equipment technology in their home area. When capacity is constrained by both equipment and operators, moving operators can increase capacity in an area that is falling behind and causing a temporary bottleneck. Models for adding resources such as operators are found to reduce cycle time (increase throughput) with diminishing returns as operators are continually added (Uzsoy, 1992). These results are consistent with the results of the simulation study for operators working in the pilot areas at Intel's development factory.

In a machine-limited job shop (where operators are assumed to always be available) research focuses on material dispatching rules. However, when there are multiple machines that

require an operator at the same time, at least one of the machines must wait and is interfered with by the machine getting serviced. This is called **machine interference**. Stecke (1992) describes how industrial engineers approach the machine interference problem. The typical machine interference formulation focuses on the optimal number of machines to assign to an operator to maximize production through the system. An alternative formulation is to decide on the number of operators needed to run a particular number of machines to optimize production or some other performance measure. This research follows the latter formulation of the machine interference problem. Stecke points out that "an assessment of the amount of machine interference is required to determine machine efficiencies, the amounts of production, and hence profits which are obtained by varying different factors, such as...changing the number of machines assigned to an operator." Typical performance measures affected by machine interference include production rate, cost, and idle resource time. Because of the difficulty in capturing data about the interaction of multiple operators with multiple machines, many of the analytic methods used to quantify machine interference are based on probability, queueing theory, or simulation. In this research machine interference is calculated based on historical data captured directly from the equipment status which is used to infer labor performance.

Treleven (1989) points out that the modeling research on dual resource constrained systems has pointed to the conclusion that the frequency of reassigning operators to equipment was a significant driver of factory performance. This result leads directly to the importance of cross-training to provide the labor flexibility to make frequent reassignments possible. The initial motivation for this research was based on some of the key insights collected by Treleven, starting with the idea that cross-training has proven to be very beneficial for factory performance. Cross-trained workers can be reassigned to match variation in production requirements due to changes in product mix or general factory disruptions such as material shortages. Fryer (1974) suggests that interdivisional labor flexibility has a greater impact on shop performance than intradivisional labor flexibility. Even when long run workload is equally distributed across divisions, having workers quickly respond to short term changes in the workload conditions can improve factory performance more than moving workers between work centers within the divisions. A caveat acknowledges the

understanding that it is easier to implement intradivisional cross-training because individual work centers share similarities, whereas divisions can have unique work environments. Each Intel factory is organized as a collection of departments that have a high level of intradepartmental staffing flexibility but very little interdepartmental staffing flexibility. It is expected that cross-training between interdepartmental skill sets will provide benefits to the overall factory.

Cross-training benefits are affected by the ratio of operators to machines, where high machine staffing levels restrict flexibility and performance, and low staffing levels limit the efficient use of capital resources. Dual resource constrained systems operate most efficiently with staffing levels from 50%-75% of the number of machines. Intel staffing areas operate over a broad range of staffing levels with pilot areas falling in the 15-55% range. When focusing on small dual resource constrained systems, results tend to be more sensitive to changing initial conditions than larger systems. However, the ranking order for both small and large systems tends to be stable, suggesting that results from studies involving smaller systems can be applied to larger systems. Since the pilot study is done at one of Intel's smaller development fabs, it is expected that any beneficial pilot results will scale to larger factory environments.

Cross-Training

Hopp (1996) points out that cross-training adds flexibility to an inherently inflexible system by extending the ability to cope with changes in product mix and fluctuations in demand. Multifunctional workers can move where needed to maintain a smooth flow of production. In U-shaped work areas (as are common at Intel) one operator can see and attend all of the machines with a minimum of walking, and additional operators can be brought in to respond to changing production requirements. Having a flexible workforce is one more option available to production managers to achieve smooth and orderly movement of material (Holstein, 1970). According to Holstein's research, the benefits of a flexible workforce can be offset by unnecessarily large control costs to execute assignment decisions. Holstein recommends putting the highest priority on moving people to tools with long queues, on main flow paths, with short downstream queues, and with processing times greater than

downstream processing times. This research follows Holstein's insight that system performance will be improved by local decisions that move operators to target areas with very long material queues thus providing short term burst capacity until the target area returns to a normal queue size. This leads to the expectation that cross-training benefits will be greatest when flexible resources are used to respond to workload spikes.

Cross-training receives special attention from Treleven (1989) in his survey of dual resource constrained system research, as a means of achieving more efficient use of labor resources. An increase in operator flexibility (measured in terms of number of operator certifications or frequency of transfers) is shown to positively affect system performance. A similar performance benefit can also result from purchasing general-purpose flexible equipment. However, the combination of both flexible workforce and flexible equipment only marginally improves performance over either isolated change. In the case of highly specialized wafer fab equipment, general purpose machines are not an economically feasible approach, so cross-training proves to be a more realistic means of introducing flexibility to the manufacturing floor. Malhotra (1990) employs similar logic to recommend increasing resource flexibility to buffer manufacturing systems against variability in contrast to the less attractive alternatives of growing inventory buffers or purchasing excess equipment capacity. With the high value of work-in-process inventory and capital equipment at Intel, it makes sense to design flexibility into the system wherever possible to provide a cushion against variability.

Most of the benefits of cross-training are achieved without going to the extreme of total flexibility, where all operators are trained to run all machines (Treleven, 1989). A goal of this research, then, is to help understand the conditions under which a cross-trained workforce provides the most impact. The simulation modeling for the cross-training pilot follows the general approach of O'Ferrell (1995), by assuming that there is always an adequate inventory of materials needed for the process. The models of O'Ferrell and this research both build off of deterministic operator performance numbers from industrial engineering time studies with processing times captured in the historical factory data collection system. Simulation models use the same inputs and general data sources to

generate the pilot area simulations. O'Ferrell carries out experiments to determine the optimum staffing levels to maximize throughput while minimizing operator idle time. Concave graphs show the results in terms of actual throughput through the area. The simulation results for the pilot areas are found to mimic this qualitative behavior. Additional results from O'Ferrell show that cross-training prevents decreased throughput when absence rates are increased. The conclusion is that cross-training improves operator utilization while also providing a larger pool of potential operators to cover for unexpected absences. O'Ferrell points out that this modeling process proved very useful in extending understanding of the manufacturing process and illuminating the benefits of cross-training. It is expected that this research will also provide insights and intuition to the participants in Intel's complex factory environment.

New Concepts

This research extends the current dual resource constrained system and cross-training literature along three important dimensions. The first contribution is the application of solutions to a real world factory environment. Several authors (Conway, 1967 and Treleven, 1989 and Gershwin, 1994) have pointed out a disappointing chasm between theory and practice. Uzsoy (1994) sums up the general consensus in the statement: "Thirty years of research on deterministic scheduling has had very little evidence of industrial application, and is viewed as solving theoretical research without a sense of actual factory practice." By simulating actual factory performance, implementing a cross-training program, and then comparing actual factory performance both with and without cross-training, it is possible to observe whether the theoretical benefits of cross-training are strong enough to impact a dynamic factory environment.

This research also provides a more accurate model of manufacturing capacity for labor-limited production systems. A better model of manufacturing capacity is needed to link the objectives of production planning and shop floor control decisions in a wafer factory (Uzsoy, 1994). This research uses simulation to demonstrate that capacity is a non-linear function of the number of operators available. Factory data are then collected to break down the separate impacts of the availability of equipment, material, and operators to quantify the disruptions to

processing time. Machine interference (as defined in Chapter 2) is used as a factory metric to measure the impact of cross-training, suggesting that machine interference can be used as a feedback mechanism to address both the needs of production planning and shop floor control.

The final area for extending existing research, mentioned by Treleven (1989), is the use of look-ahead information about the anticipated workload at work centers when making labor assignment decisions. Semiconductor wafers flow through many processing steps with re-entrant flow, where a wafer may return to the same machine for multiple processing steps. Because of the large number of total processing steps, queues at individual steps are kept relatively short to ensure low cycle times through each operation. To assess the workload at a given work center, it is necessary to not only look at all of the wafers currently in these queues, but also the expected arrivals of upstream wafers within the time horizon of the labor assignment decision. This is a non-trivial calculation because of multiple product types, re-entrant flow, and variable resource availability at each processing step.

Factory Context

This research study was carried out at one of Intel's development factories. The factory is a dual charter facility responsible for both developing manufacturing processes for new products and also maintaining production performance on current technologies. This research focuses on those operators who are primarily responsible for maintaining production performance with standardized activities. Like most semiconductor manufacturers, Intel operations are dominated by the concerns of high capital equipment costs and a make-to-stock mentality where production lots are rarely associated with a specific customer order. This results in a clear management emphasis on maintaining high wafer throughput and high equipment utilization (Uzsoy, 1992). Even though an adequate labor force is required to sustain factory output, Intel management does not traditionally give rigorous analytical attention to labor performance, instead focusing on the performance of capital equipment. Because of the high equipment costs, improvements in capital equipment performance have a first order impact on Intel's bottom line. For the purposes of this analysis, the number of machines, capacity per machine, and factory layout are treated as fixed constraints, with the

understanding that significant time and effort is already being applied to these high-leverage challenges.

Long term factory staffing is acknowledged to have a second order effect on factory productivity, typically receiving management attention through negotiating the optimal staffing levels. There is a continual positive tension to meet increasing production targets as factory floor personnel want higher headcount and upper level managers want higher productivity out of the same headcount to meet the production goals. Operators are currently assigned to a home area where they are cross-trained to run most or all of the equipment. This is necessary to cover for absences such as vacations, sick time, and training. However, it is not easy for current staffing to respond to workload spikes in one area without calling in operators from other shifts to put in overtime. Since workload spikes cannot always be anticipated, and have an uncertain duration, it is not always possible to quickly staff up an area by using overtime. Cross-training across functional areas is expected to help factory floor personnel deliver higher productivity by being better equipped to respond to workload fluctuations.

Factory Environment

Every physical system has a finite capacity that changes with time and events. For the wafer factory, each critical processing step has a dynamic capacity that changes with product mix, equipment availability, factory congestion, setup requirements, rework, quality (yield), and operator availability. In contrast to other types of manufacturing, wafer fabs spend a large part of clock time performing scheduled and unscheduled maintenance (Johri, 1993). Uzsoy (1992) makes the statement that "the main cause of uncertainty in semiconductor manufacturing operations is due to unpredictable equipment downtime." Interactions between separate equipment failures can cause conditions to quickly propagate through the factory as a result of re-entrant flow where multiple steps can take place on the same tool. Another source of variation is the diverse equipment characteristics, where processing times can range from several minutes to several hours and the quantity of wafers simultaneously processed on a single machine can range from 1 to 125. This causes very different arrival patterns for downstream tools. Because of re-entrant flow, a single tool can be fed by a mix

of steady and lumpy processing volumes. The factory environment is also characterized by a large volume of data collection (Uzsoy, 1992). The challenge is to identify the key pieces of data that can be used to assess factory conditions and recommend appropriate actions that increase the dynamic capacity as well as the utilization of that capacity.

Currently, Intel's factories are designed around constraint tools, near constraint tools and non-constraint tools. These categorizations are based on the availability-utilization gap that represents an allowable amount of idle time when equipment is available but not processing. This idle time represents how much of a capacity buffer is built in to each tool. Constraint tools (usually the most expensive) are expected to run with the least amount of idle time, and require the highest material buffers to achieve the required utilization. Near constraint tools are scheduled to have a moderate amount of idle time as well as a moderate queue size. Non-constraint tools (usually the least expensive) have the highest idle time, providing the largest capacity buffer, and are characterized by the shortest queues. From a scheduling perspective, lots should quickly travel through non-constraint processing steps with individual cycle times kept close to processing times. Longer cycle times (still on the same order of magnitude as processing times) are expected for constraint tools.

The factory is **labor constrained**, meaning that there are not enough operators to keep all of the tools in the factory running at the same time. This makes the **effective capacity** of each equipment technology depend on whether there is an operator available to load it. Effective capacity is defined by Hopp (1996) to be the natural capacity (based on processing time alone) adjusted to reflect the expected availability of resources. Factory capacity can only be utilized when three resources are simultaneously available: equipment, material, operator. Since multiple operators cover multiple tools, the number of operators in a factory staffing area is one factor in determining the usable capacity of the area. Cross-training can increase workforce agility to ensure that certified operators are available to provide the excess burst capacity in response to inventory build up. Operators who have been with Intel for a long time become cross-trained as changes in technology and market conditions change the demands on the different functional areas within the factory. However, as long as a technology does not become obsolete, specialization is usually encouraged. It is typical for

senior operators to be promoted to maintenance technicians as part of career progression. Specialization keeps maintenance technicians in their home area, where they can also fill in for operator vacations, training, and breaks as long as there are not equipment failures or other maintenance activities required in the area. It is a novel concept at Intel to value the skills of cross-trained generalist operators who provide manufacturing flexibility.

Tactical Performance Measures Already In Use

Manufacturing performance is currently tracked along a variety of dimensions. At Intel's development Factory, typical daily performance measures are listed in Table 1-2. Performance measures are aggregated at the levels of individual process steps, equipment families, departments, and across the entire factory depending on the audience of the report. Typically, if reported metrics fall significantly below goals, factory conditions are investigated to look for staffing issues, excessive equipment down time, inventory build-up, quality problems, or other unplanned occurrences. Labor performance is only measured

Performance Dimension	Metric	Performance Goal
Production Volume	Value-added processing steps completed during shift (irreversible)	Shift goal for value-added processing steps (set by Industrial Engineering)
	Total processing steps completed during shift (including inspection, washing, etc.)	Shift goal for total processing steps (set by Industrial Engineering)
Manufacturing Lead Time	WIP turns (shift output divided by work in process)	Shift goal for WIP turns (set by Industrial Engineering)
	Cycle time (weekly average)	Average shift cycle time goal (set by Industrial Engineering)
	Idle lots (in queue for given length of time)	Zero idle lots
Quality	Line Yield	Shift goal for line yield (set by Industrial Engineering)
Work In Process Inventory	Beginning of shift inventory End of shift inventory	Inventory goal (set by Industrial Engineering)
Equipment	Availability Utilization	Availability-Utilization gap goal (set by Industrial Engineering)

Table 1-2 Typical Factory Performance Measures

implicitly as a component of the shift-end goal formulas which assume a static labor efficiency. Currently, Intel does not provide regular feedback to assess the impact operator coverage has on factory performance.

Labor performance is a difficult manufacturing variable to quantify in a meaningful and consistent way. Variables such as operator training, experience, vacation schedules, break schedules, efficiency, and skill sets interact so that any given factory event may result in a range of responses. For example, when a visual inspection uncovers an irregularity in the surface of a wafer, an expert operator may be able to troubleshoot the problem on her own, a typical operator may need to involve quality specialists or engineers, and an operator in training may need to pull a senior operator off of another job to get a second opinion. Usually, variables such as equipment status and lot status are tracked in detail to summarize the aggregate impact on equipment capacity and inventory build up. However, for each factory event, it is not clear how long it takes an operator to notice the event, how many additional operators are involved in responding to the event, and how efficient the response measures are. Factory data is currently tracked for equipment and material, but not for operators. This research uses the detailed equipment and material performance data, and infers operator performance based on assumptions about how factory data is segregated into distinct tool states (listed in Table 2-1). This inferred operator performance fulfills a need for better feedback to the tactical decision makers on the factory floor as well as to the long range production planners who set targets for capacity and productivity that are used to forecast future production capability.

2 Measuring Cross-Training Impact

Decision Making Context

Tactical decisions to assign cross-trained workers are primarily based on identifying areas that need an immediate boost in output or effective capacity to handle workload spikes. Typically, capacity is calculated based on tool processing speed without incorporating the availability of operators. Effective capacity, as defined by Hopp (1996), represents the maximum amount of output for a given set of tools and operators and a given product mix. The decision to assign a cross-trained worker to a target area is based on three factors: 1) the effective capacity benefit of having surplus operators in a target area, 2) whether the workload exceeds the current effective capacity (before a worker is added), and 3) the relative effective capacity benefit from adding operators to the target area rather than an alternative area. Because people represent the most flexible manufacturing resource, it is difficult to track and quantify the behavior of individuals interacting with the manufacturing system. At the same time the factory has multiple objectives that are constantly balanced to meet changes in markets and technology. In the complex, dynamic factory environment it is challenging to quantify these benefits and tradeoffs.

Interdepartmental cross-training enhances tactical options by increasing staffing flexibility. Increasing flexibility through interdepartmental cross-training is the scope of this pilot study, shown as a pyramid (Figure 2-1). At the bottom of the pyramid, the first step (with the highest impact) is to establish a plant layout and to decide how to group different equipment technology. Intel's departmental grouping of equipment results in a similar grouping of operator specialists. Flexibility is enhanced by pooling similar equipment. At the second level of the pyramid, intradepartmental cross-training helps to increase flexibility by pooling similar operators who can all run the pooled equipment. The third level, where the scope of the pilot study starts, represents interdepartmental cross-training which improves flexibility by allowing operators to move among their home areas. This extends the pool of operators who may be available depending on the workload in their own home area. When two separate areas have similar skill sets, cross-training has a very low cost, and provides a pooling benefit to handle workload spikes in either area. The pyramid is capped by an

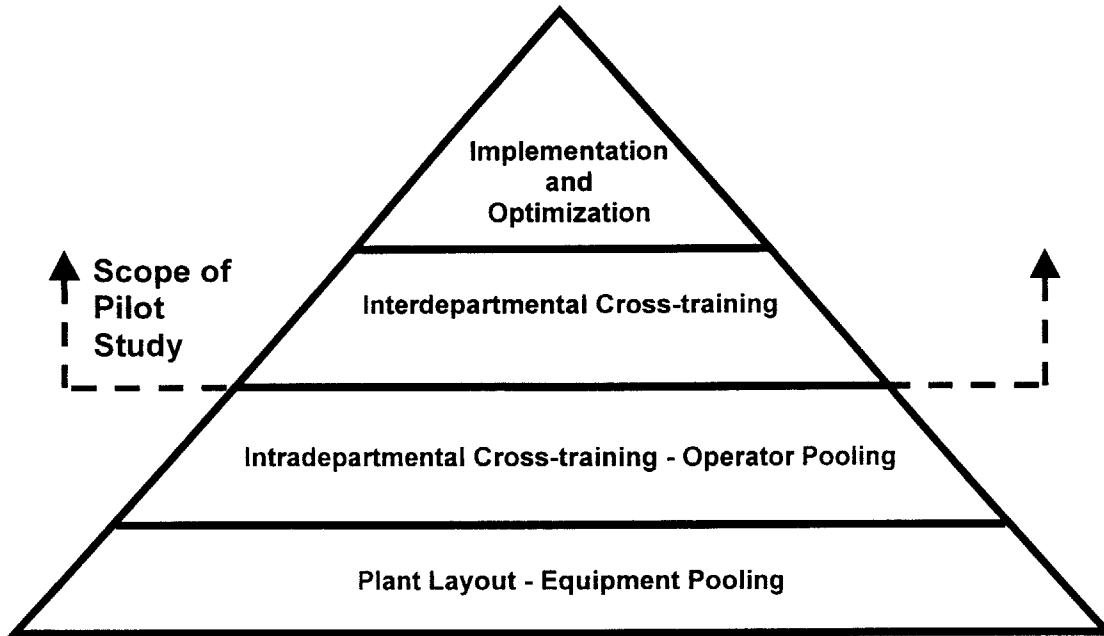


Figure 2-1 Impact of Staffing Flexibility

iterative fine-tuning of implementation and optimization, where refinements on all three levels are made based on performance results.

Effective Capacity and Expected Output

The effective capacity of a department or other staffing area in the factory is determined by the equipment run rate, equipment availability, and operator availability as illustrated in Figure 2-2. Effective capacity, expected material, and expected output are all measured in units of wafer volume per unit of time. The thick dashed-line curve in the middle represents the baseline effective capacity for a given set of equipment. Effective capacity increases and then levels off as the number of operators increase. When there are zero operators, no material can be loaded, resulting in an effective capacity of zero. As operators are added, equipment can be loaded to the extent that there are operators available. The effective capacity approaches the limit set by equipment run rate and equipment availability (assuming there is always an operator available for each setup, loading, and unloading operation). The entire curve shifts upward if additional equipment capacity is added, as represented by the top curve.

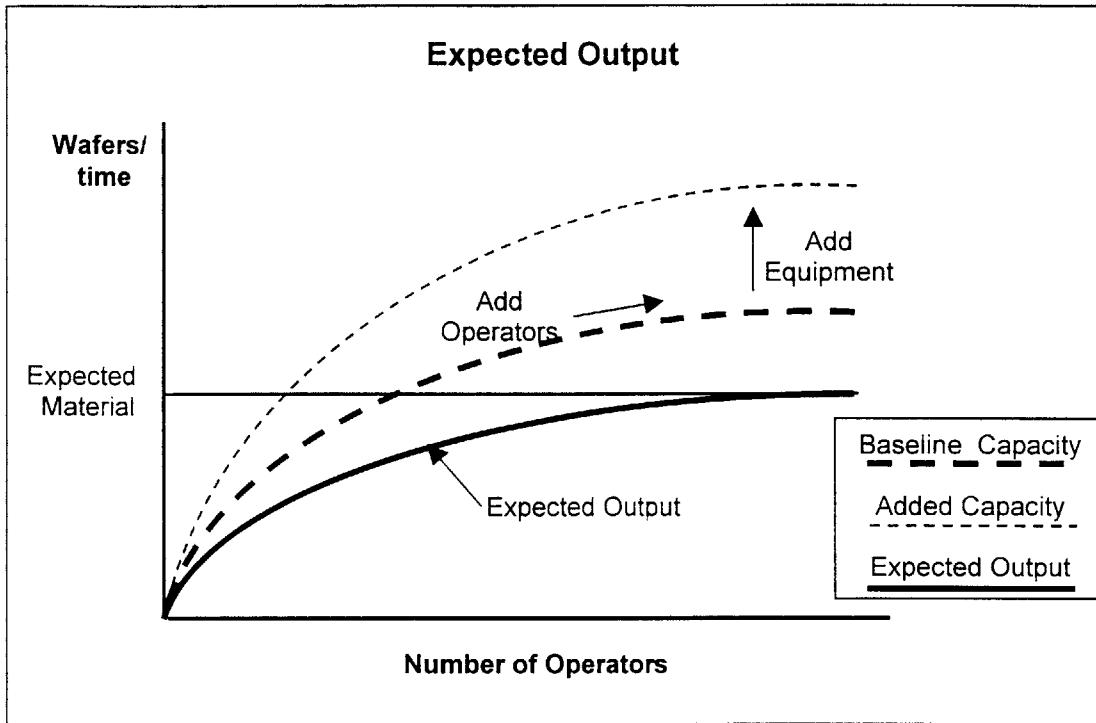


Figure 2-2 Expected Output vs. Number of Operators

The expected material per unit time is independent of the effective capacity. If the expected material is less than the effective capacity, as shown by the horizontal line in Figure 2-2, then the expected output will be limited by both effective capacity and expected material, as shown in the bottom curve. On the left end of the Expected Output curve, the area is understaffed and capacity limits the output. As the number of operators increases, the effective capacity increases, but the Expected Output can never exceed the Expected Material. For a sufficiently high expected material, the expected output is the same as the effective capacity since it is no longer constrained by material. Adding operators increases effective capacity, but the additional capacity will only be used to the extent that there is material available to run.

In general, the effective capacity is a function of the processing speed of each tool (where processing speed is determined by product mix) as well as equipment availability. A product mix that can be processed faster or a higher equipment availability shifts the effective

capacity curve up. A product mix that must be processed slower or a lower equipment availability shifts the effective capacity curve down. Each new factory event that causes a significant change in the product mix becomes a new decision point for adjusting staffing assignments. From the perspective of the tactical decision maker, product mix is a constant quantity over the time horizon of the decision (by definition). Since product mix is set exogenously, the only tactical decision is the number of operators to allocate to each area. At the next event that substantially shifts the curve, another staffing decision will be made.

Expected Workload

Equipment downtime or personnel absence in certain key pinch points in a wafer factory causes an immediate queue explosion where material builds up in one location (O'Ferrell, 1995). This can cause a sudden increase in cycle time through the factory until the long queue is processed through the area. Expected output will only approach effective capacity when the expected workload exceeds the effective capacity. **Expected workload** is the expected amount of material to be processed during a given time period including both material in the queue and material expected to arrive from upstream processes. Low workload relative to effective capacity will result in idle resources (tools or operators). Adding resources when the queue is zero increases effective capacity, but does not increase the expected production output of the area, because the additional resources are expected to be idle. High workload relative to effective capacity will result in high utilization of resources and high expected output. Holstein (1970) warns against assigning operators to areas with short queues because they will quickly reduce the material queue to a very low level and therefore soon cause the potential for another transfer to another area. Moving operators from areas of low workload to areas of high workload provide the best performance when choosing between staffing locations for a cross-trained operator.

Capacity Tradeoff

The marginal benefit of adding an operator is the product of complex interactions between sequencing and the simultaneous availability of three resources: material, equipment, and operator. Long term strategic plans are based on expectations of average workload and

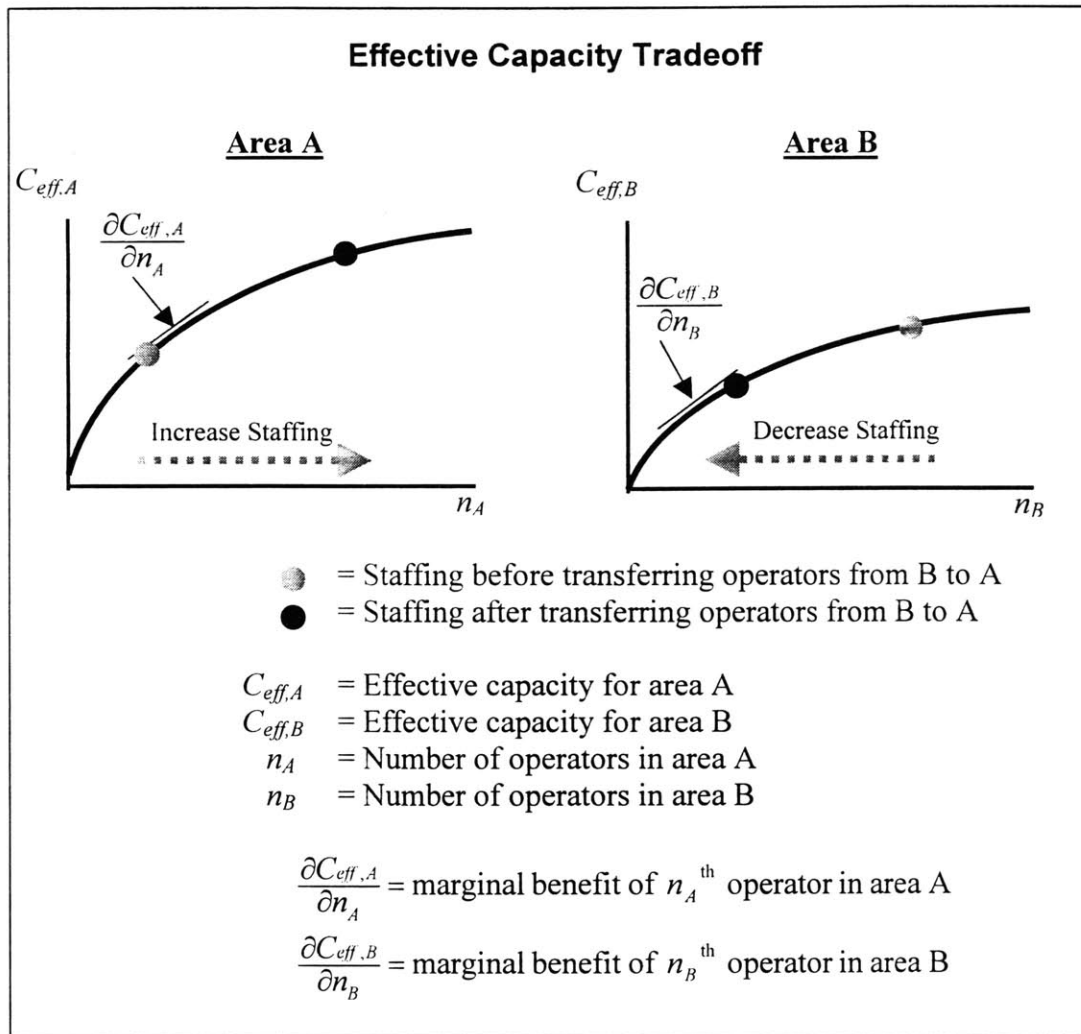


Figure 2-3 Capacity Tradeoff

effective capacity to set overall factory headcount. Short term tactical decisions require knowledge of the concave relationship between effective capacity and the number of operators. Priority for staffing is given to the area with the higher workload relative to effective capacity. Since there is a finite supply of operators, every cross-trained operator added to Area A becomes unavailable to work in Area B. Figure 2-3 illustrates the tradeoff of moving operators from Area B to Area A. The gray dot on each curve represents the initial distribution of operators, where Area A has a relatively low number of operators and Area B has a relatively high number of operators. The effective capacity of Area A is on the left end of the curve, while the effective capacity of Area B is well into the right end of the curve.

After operators are transferred from Area B to Area A, the distribution of operators shifts to the points identified by the black dots on each curve. Because of the concave shape of the curves, additional operators provide a diminishing marginal benefit in. This corresponds to an increasingly negative marginal impact on Area B. If the objective of the decision maker is simply to maximize the aggregate output of Area A and Area B, she will set staffing levels so that the marginal benefit of additional manhours in Area A is equal to the marginal benefit of additional manhours in Area B. However, tactical decisions are primarily based on identifying areas that need an immediate boost in effective capacity to handle workload spikes. The tradeoff focuses on how many operators Area B can afford to give up to provide short-term burst capacity to Area A. **Burst capacity** is an increase in effective capacity beyond normal capacity levels usually in response to the buildup of a large queue of material. The key assumption is that Area A has a high workload where output will be close to effective capacity, while Area B has a low workload (a small queue of material) so that decreasing capacity will not significantly change output. Operator transfers should move operators from areas of low workload to areas of high workload to provide the best performance improvement.

Performance of a Single Tool

Machine Interference Definition

Machine interference refers to the time that equipment is available to run but is waiting for an operator to finish servicing other equipment (Stecke, 1985). By definition, equipment must be in one of the following mutually exclusive states: 1. **Processing**, 2. **Down** (unscheduled repair or scheduled maintenance), 3. **Starved** (equipment available but no material), 4. **Machine interference** (equipment and material available but no operator) as summarized in Table 2-1. Machine interference only counts the time when both equipment and material are simultaneously available but the machine is not processing. If equipment is down at the same time an operator is unavailable, the state is defined as down. This hierarchy reflects the fact that equipment and material availability set the maximum effective capacity within the horizon of tactical decisions.

Equipment State	Conditions Included in State
Processing	<ul style="list-style-type: none"> - Processing material - Loading/unloading material (usually staged while tool is processing) - Processing at partial capacity (not all chambers processing)
Down	<ul style="list-style-type: none"> - Tool failure - Scheduled maintenance - Quality excursion (may require operator time) - Other tool disposition (may require operator time)
Starved	<ul style="list-style-type: none"> - No lots to run on tool - Batch delay waiting to accumulate multiple lots (for batching tool)
Machine Interference	<ul style="list-style-type: none"> - Tool and material available, tool not running (waiting for operator)

Table 2-1 Distinct Tool States

Machine Interference Calculation

Machine interference is on a percentage scale from zero to 100%. Zero machine interference indicates that an operator is available every time both equipment and material available to run, while 100% machine interference indicates that operators are never available forcing available equipment to remain idle. Stecke (1992) points out that it has long been recognized that machine interference times are inconvenient to obtain directly because it is difficult to directly time simultaneous events for several machines. As a result, most studies define the amount of machine interference by developing mathematical relationships that either treat events deterministically, or use queuing theory or binomial expansions.

For Intel's factory, machine interference for a given tool is based on observations of the tool state recorded every 15 minutes. Individual observations are counted across a shift and then divided by the total number of observations to come up with the measures of processing time, down time, and starved time that are used to indirectly calculate machine interference. This formulation refines the formulations suggested by Stecke (1992) by also separating out starved equipment.

- n = Total number of state observations (collected every 15 minutes)
- p_i = 1 when equipment is processing at observation i
= 0 otherwise
- d_i = 1 when equipment is down at observation i
= 0 otherwise
- s_i = 1 when equipment is starved at observation i
= 0 otherwise
- \bar{m} = Average equipment machine interference across n observations

$$\bar{m} = 1 - \frac{\sum_i p_i}{n} - \frac{\sum_i d_i}{n} - \frac{\sum_i s_i}{n} \quad (2.1)$$

Equation (2.1) represents the machine interference performance of a single tool. It is indirectly measured as the percent of observations where the tool is not processing, not down, and not starved (based on the categorizations given in Table 2-1).

Availability-Utilization Gap

Intel typically focuses on the **availability-utilization gap** as a measure of how efficiently capacity is used. The gap measures the amount of time that equipment is available to run, but is not processing. The equipment is idle either because of lack of material or lack of operators. The availability-utilization gap is used to evaluate actual performance versus planned performance. As shown in Figure 2-4, when there is low workload (low queue and infrequent arrivals), additional material is needed before the amount of processing time will increase. The distribution of low workload states (Down, Starved, Processing) does not change when operators are added. However, when material is added to produce a high workload (without any additional operators), the equipment has material to run and is no longer starved. The gap does not change, since the available equipment and material both sit idle waiting for an operator (machine interference). When there is high workload and operators are added to cover the additional production volume, the elimination of machine interference can significantly increase output. Idle equipment no longer has to wait for an operator to load material and the availability-utilization gap is reduced. Adding operators to increase effective capacity provides more of a benefit when there is high workload.

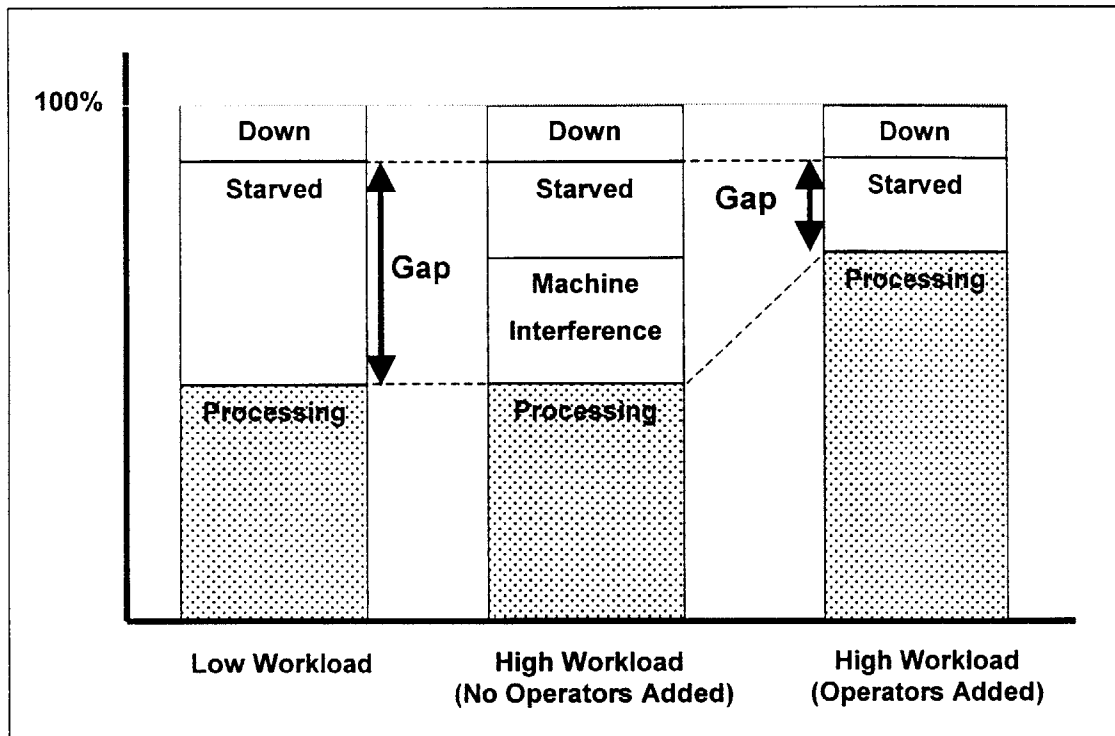


Figure 2-4 Role of Workload in Reducing Gap

Performance of Multiple Tools

Staffing Areas

Factory bays, grouped roughly by processing technology, serve as the natural units for staffing decisions. A given **staffing area** may consist of one or more bays covering anywhere from 5 to 25 tools. Each staffing area is characterized by the following traits:

- Tools are roughly grouped by process technology
- Each area can have multiple types of tools
- Operators are 100% cross-trained to run all tools in their home area
- All area operators report to the same supervisor

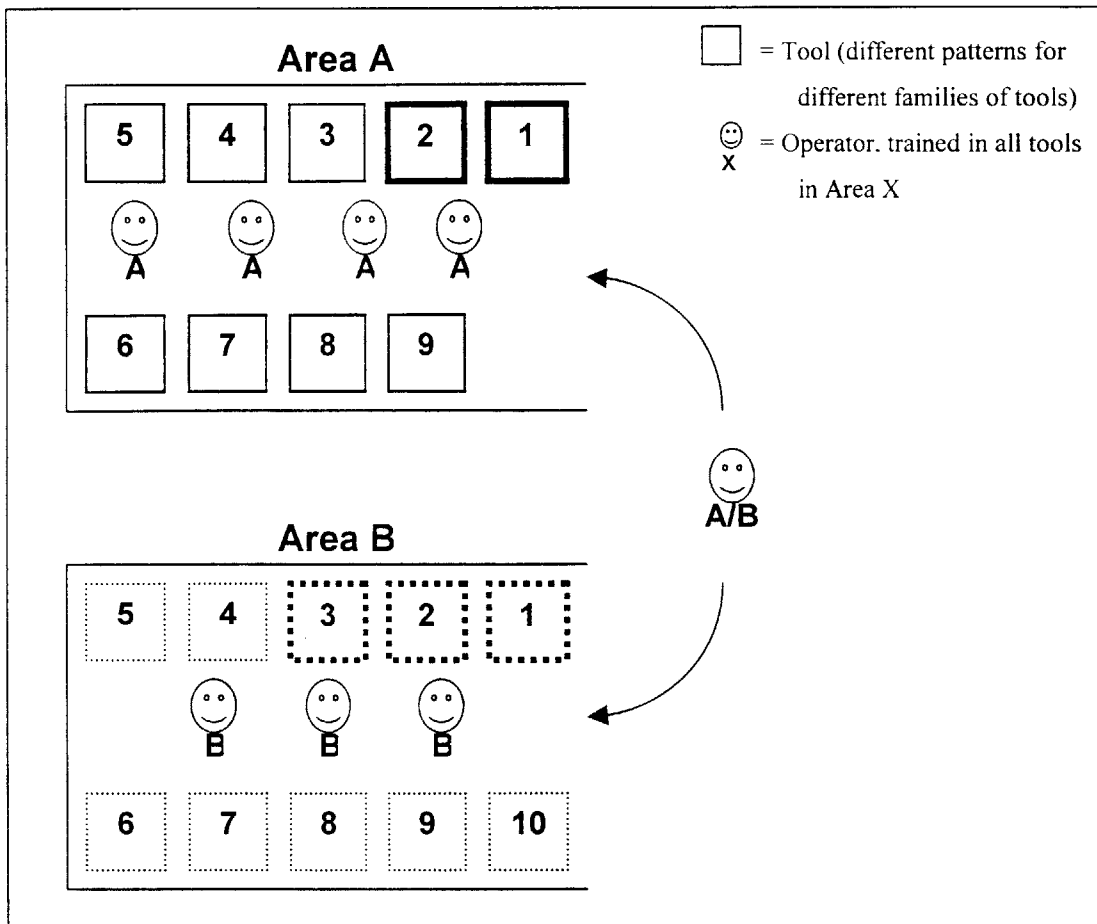


Figure 2-5 Staffing Areas A and B

An example of two single bay staffing areas is shown in Figure 2-5. Four Area A operators are all trained to run the two types of tools found in Area A, and three Area B operators are trained to run the two types of tools found in Area B. One additional cross-trained operator is trained to run all four types of tools. Strategic planning decisions set the staffing levels in each area based on the expected maximum effective capacity (which is calculated from long run averages or future expectations of product mix and tool availability). For an operator who does not have a home staffing area, tactical decisions can simply assign the cross-trained operator to either Area A or Area B based on the area with the higher workload. A more realistic tactical environment would assign the cross-trained operator to home Area B, and then shift the cross-trained operator whenever Area A experiences a significant spike in workload at the same time that Area B has low workload.

Extension of Single Tool Formulation

The benefits of assigning cross-trained workers are based on boosting the effective capacity to handle workload spikes across an entire staffing area. The machine interference formula given in Equation (2.1) for a given tool is easily extended in Equation (2.2) to give a single machine interference value that represents the performance of the entire staffing Area A, consisting of k tools.

- n = Total number of state observations (collected every 15 minutes)
- p_{ki} = 1 when tool k is processing at observation i
= 0 otherwise
- d_{ki} = 1 when tool k is down at observation i
= 0 otherwise
- s_i = 1 when Tool k is starved at observation i
= 0 otherwise
- \overline{m}_A = Average machine interference for all tools in Area A

$$\overline{m}_A = 1 - \frac{\sum_k \sum_i p_{ki}}{n} - \frac{\sum_k \sum_i d_{ki}}{n} - \frac{\sum_k \sum_i s_{ki}}{n} \quad (2.2)$$

Equation (2.2) is used to evaluate the how well a group of operators assigned to a single staffing area are able to sustain a high level of effective capacity across a set of equipment. The average machine interference for Area A is indirectly measured as the percent of all Area A tool observations where the observed tool is not processing, not down, and not starved. High effective capacity is only important when there is a high workload (and low starvation) so that the availability-utilization gap closes. If there is a specific bottleneck tool that should receive priority over other tools in the staffing area, then Equation (2.1) is more appropriate to assess the performance of the bottleneck tool in isolation. The appropriate formulation of machine interference measures the performance of the staffing area. Because the staffing area represents the tactical decision-making unit for assigning operators, reporting on machine interference is a useful way of both providing feedback and evaluating the performance of the tactical decision-maker who assigns cross-trained operators.

Data Collection

Performance measures need to reflect the scope and frequency of the decisions they are supporting. Strategic planning performance measures can be aggregated across families of equipment and products and across long time horizons. Tactical performance measures need to feed back more detailed and current information. For Intel's cross-training pilot, the natural decision-making cycle for labor assignments repeats at the beginning of each shift based on the status of equipment, work in process inventory distribution throughout the factory, and the labor pool available for staffing. This environment is monitored throughout the shift and adjusted as needed.

One of the obstacles in establishing an analytical treatment of labor has been that current factory data collection systems do not collect and report data from the perspective of staffing areas. A set of equivalent tools can be spread out across multiple staffing areas, and in some cases can be paired with different process technologies. Factory reports tend to aggregate across the entire set of equivalent tools, regardless of where they are physically installed or how they are staffed. The resolution of factory data collection and archiving determines whether the performance of individual tools, individual lots, and individual operators can be separately monitored. Intel's historical data collection system archives availability data about individual tools, but aggregates the other information across tool types to simplify the calculation of various performance measures. If a tool fails, it is possible to find out specifically which tool went down. However, if four tools are running and a fifth tool is starved, for the purposes of performance metric computation, it is not possible to identify which tool was starved. This presents a problem for tools installed across multiple staffing areas, because the data resolution does not provide enough information to correctly assign the starved tool to only one staffing area. Since it is difficult to track the specific interaction of each operator with each tool, the machine interference calculation given in Equation (2.2) is not always feasible.

As a practical measure, machine interference metrics are calculated only for those types of tools that are either entirely within one staffing area, or that are dominated by the performance of one staffing area. For example, the first two tools of Area A shown in Figure

2-5 are the same type of tool. If there are also three additional tools of that same type scattered in different areas, the Intel data collection system will automatically sum the performance across all five tools before archiving the data. This prevents the machine interference calculation from isolating the performance of the two tools in Area A. The convention used in this research is to leave ambiguous tools out of the calculation, instead using the performance of the remaining tools as a proxy for the actual performance of the entire staffing area. As staffing areas get more analytical attention, data will be archived at a finer resolution to allow more flexible analysis and reporting, so future information systems will eliminate the need for using a proxy . However, actual simulation models and performance metrics in this research used the proxy method in order to isolate cause and effect within the constraints of Intel's current factory collection system.

3 Identifying and Simulating Pilot Areas

Selecting Pilot Areas

With the theoretical benefits of cross-training understood, a pilot cross-training program is established to empirically capture results in the actual factory environment. A cross-functional team representing first-level supervisors, second-level supervisors, the training department, and the industrial engineering department is responsible for defining the pilot program. Operators are typically trained to run all of the tools in their home staffing area. The **cross-training** in this research refers to further training of operators to become proficient running all the tools in some complementary staffing area where they can be moved in response to factory needs. Since the key benefit of cross-training operators between staffing areas is to provide additional burst capacity to a given area, a good pilot should focus on an area with high workload as well as high variation in workload. The first staffing area in the pilot, Area A, is a high-traffic area for material in the factory, so it exhibits both high workload as well as high variation. For the pilot study, a second staffing area, Area B, must be selected so that cross-trained operators will be able to move between Area A and Area B as needed.

Complementary Workload

Cross-trained operators will only be available to supply burst capacity at times when their home area can afford to give them up (usually when workload is below capacity). Analysis of the correlation between workload in the home area and workload in other areas in the factory can be used to filter out staffing areas that are unlikely to have surplus operators available when needed. If two areas are perfectly correlated, then a workload spike in the home area corresponds to a workload spike in the complementary area, so both areas will be capacity-constrained regardless of where the cross-trained operator is assigned. However, if the two areas have perfect negative correlation, then a spike in workload in the home area corresponds to low workload in the complementary area, so the complementary area can afford to run short-handed. Similarly, a lull in workload in the home area corresponds to a spike in the complementary area, so the home area can afford to give up the cross-trained

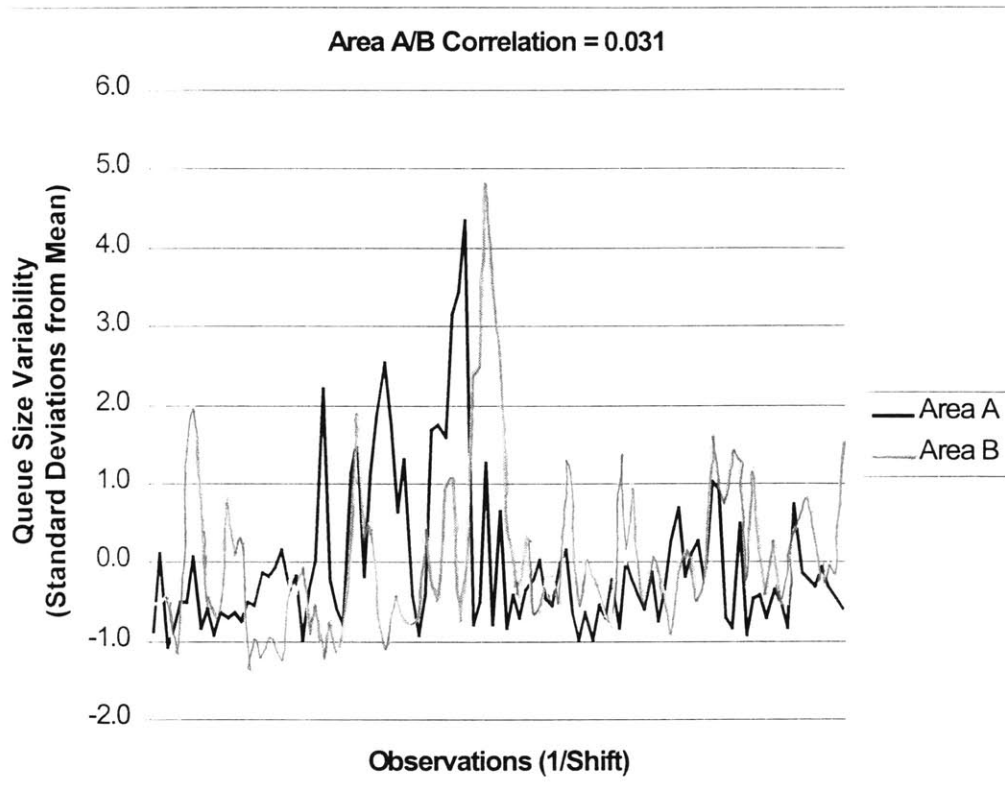


Figure 3-1 Queue Size Variability for Areas A and B

operator to provide additional burst capacity where it is needed. The variation of queue size at the beginning of each shift is used as a measure of workload as shown in Figure 3-1. The vertical axis is scaled in terms of the number of standard deviations from the mean. Each data point represents the queue length at the beginning of a shift. Opportunities for cross-training are indicated in the high spikes for either area by itself. Since baseline staffing levels are set based on static estimates of workload using aggregate demand forecasts, sudden spikes can overwhelm the baseline effective capacity and create a need for additional short term burst capacity. Similarly, queues falling significantly below the mean indicate that operators can be made available to an area in need. The correlation between Area A and Area B is 0.031, so that the two areas are essentially uncorrelated. This makes the two areas reasonable candidates for cross-training because the timing of workload spikes are spread over different shifts. This allows operators to move to the area that is experiencing a workload spike. A cross-trained workforce does not add value when demand for burst

capacity is synchronized between the two areas. When multiple areas are experiencing workload spikes, having cross-trained workers will not provide any additional benefit beyond the existing intradepartmental cross-training.

Operator Availability

Cross-training pays off when an operator is available to immediately leave her home area in response to a workload spike in another area. The decision on where to cross-train operators (for the pilot as well as for normal operations) is largely determined by expectations for both operator availability and operator effectiveness as listed in Table 3-1. **Operator availability** is viewed from the perspective of the area experiencing a workload spike. A cross-trained operator is available to immediately provide burst capacity only when she has the necessary certifications and there is not a spike in workload in her home area. The amount of flexibility in responding to factory workload spikes is driven by the size of the cross-training pool. This pool of operators can be quickly expanded by aggressively pursuing cross-training in those areas that require a minimum of training time and then keeping those certifications current

Goal	Selection Criteria	Description
Operator Availability	Complementary workload	- Negative or weak workload correlation
	Minimize training time	- Operators who already have certifications outside their home area - Tools with similar technology and operator procedures
	Stable process	- Operator procedures do not change frequently - Operator certification stays current without additional training requirements - Tools with low complexity that do not require special expertise
Operator Effectiveness	Operator fit	- Ability to deal with ambiguity - Works well with new people - Positive attitude
	Low risk of misprocessing	- Avoid tools that can damage many wafers simultaneously - Avoid tools that require setup expertise

Table 3-1 Pilot Area Selection Criteria

with the minimum amount of additional effort. Operators spend at least six weeks training for each new equipment certification. Leveraging existing certifications and looking at equipment with similar operator procedures can minimize the amount of time operators spend in training. For equipment that does not change dramatically across product generations, operator certifications are less likely to incur future re-training costs. If operators are cross-trained on rapidly changing technology, then their certifications are likely to be out of date when a workload spike triggers demand for those skills. Increased operator availability drives the process for selecting which areas to add to the cross-training pool.

Operator Effectiveness

Operator availability contributes to the effective capacity, but delivering on this gain also requires the operator to be effective outside of his home area. Individuals who are expected to join the pool of cross-trained workers need to be comfortable dealing with ambiguity, working with new people, and promoting a positive attitude. Because the reporting relationships at Intel are based on the home area, when an operator moves, he is working outside of the responsibility of his direct supervisor. This will likely result in exposure to a different management style and different expectations. The teams that run each area also have their own ways of handling work and dividing responsibilities among team members. The cross-trained operators helping out in the area need to be able to seamlessly integrate into the flow of the team and sense how best to help alleviate the spike in workload. People who do not possess these skills and sensitivities will not likely provide a great deal of value even when they are physically available to work. It is also important to recognize and control the risk of misprocessing for new operators inserted into a dynamic team situation. To minimize the exposure to misprocessing costs, temporary relief operators should not be responsible for batch tools, where a mistake can simultaneously damage many wafers, or for tools that require a great deal of setup expertise. Instead, home area operators who run those high-risk tools every day should be responsible for continuing to run them in periods of high workload. The cross-functional team responsible for defining the pilot program screened the selection of pilot areas and operators for this research through the criteria for operator availability and operator effectiveness.

Pilot Staffing Areas

Two pilot areas are selected to empirically investigate the benefits of cross-training. Since there are four different shifts at Intel, two shifts were involved in the pilot and two shifts maintained their pre-existing workforce as an experimental control (no newly cross-trained operators within scope of pilot program) as shown in Table 3-2. Area A includes a single room with operators that reports to one supervisor while Area B includes two rooms with operators that report to another supervisor. All operators with a home in Area A are certified to run all of the A equipment and all of the operators in Area B are certified to run all of the B equipment. Typically five operators are on the payroll for each shift for Area A but only four actually staff Area A during each shift. The extra operator slot covers vacation, sick, training, or other reasons operators are not working in the factory. Area B has six operators on the payroll for each shift but only five actually staff Area B during each shift. Flexible staffing for Areas A and B comes from two newly cross-trained operators for the pilot program for each of the pilot shifts. These pilot operators start in other areas in the factory, but they already have certifications on most of the equipment in Areas A or Area B. As they are cross-trained, they rotate in to Areas A and B in such a way that the total number of operators on the payroll stays fixed. All shifts (including the control shifts) have the option of calling in Area A and B operators from each other's shift for overtime work. In general, the flexibility within each area is used to cover for vacation, training, and sick leave. This is a common dynamic in many of Intel's staffing areas that already provides a first order benefit of pooled operators and pooled machines. The benefit of cross-training between these existing pools is additional burst capacity to handle increases in workload spikes.

Areas	Payroll Operators	Typical Staffing	Source of Staffing Flexibility
A	5	4	- Newly cross-trained operators for pilot (2/shift)
B	6	5	- Overtime operators from other shifts

Table 3-2 Cross-Training Pilot

Modeling Effective Capacity of Pilot Staffing Areas

To help understand the marginal tradeoffs for staffing, a simulation model was created for each pilot area using ASAP™ software from AutoSimulations Inc. Intel has a simulation modeling team that has created a factory model called NexSIM. It primarily models factory performance without taking labor resources into account. The pilot area models started with the maintenance schedules and failure rate information already loaded into the NexSIM model and added specific tool performance and operator performance variables for the pilot areas from the sources shown in Table 3-3. Most machine and operator times are assumed to be deterministic based on the high level of automation covering both processing and the operator interface. Operator maintenance, preventive maintenance, and tool failures are defined to follow statistical distributions as listed in Appendix A. All values assume a constant product mix, representing the planned factory loading for the duration of the pilot. Operators are assumed to work 80% of their shift, with 20% of the time off for breaks. Break times are staggered to avoid having more operators than are necessary take a break simultaneously.

Variable	Source
Tool name Number of tools in each pilot area	Factory tool matrix (maintained by the equipment planning group)
Tool processing time	Actual cycle time report (from factory computer system)
Operator class Operator get (from queue) time Operator put (to next queue) time Operator load time Operator unload time Lot identification time Mean time between operator maintenance events Mean time to complete operator maintenance	Labor script files (based on industrial engineering studies)
Operator break schedule Mean time between preventive maintenance events Mean time to complete preventive maintenance Mean time between tool failure Mean time to repair	Existing NexSIM files (based on Intel's simulation modeling team)

Table 3-3 Source of Simulation Inputs

Modeling Approach

The two pilot areas are characterized in Table 3-4. Area A has 11 total tools grouped into three different process technologies. Each process technology requires a separate operator certification. There are only two independent material queues because two of the technologies are always processed in series before sending material to the next downstream process. Of the 11 total tools in Area A, only 9 proxy tools are used to report simulation output. The number of proxy tools represents those process technologies that can be easily separated out in factory data collection systems (as discussed in the Data Collection Considerations section in Chapter 2). The average processing time per lot is 41 minutes which drives the frequency of demand for an operator. The average amount of time required for an operator to get, load, and unload a lot is 6 minutes. Area B has 23 tools representing nine different process technologies, fed by 7 independent material queues. Only 13 proxy tools in Area B can be easily separated out in factory data collection systems. The average processing time is 100 minutes per lot with 6.5 minutes of operator time required to get, load, and unload each lot.

Pilot Area	Rooms	Proxy Tools/ Total Tools	Process Technologies (Certifications)	Material Queues	Average Process Time	Average Operator Time
Area A	1	9/11	3	2	41 min/lot	6 min/lot
Area B	2	13/23	9	7	100 min/lot	6.5 min/lot

Table 3-4 Characterization of Pilot Areas

Each simulation run starts with empty tools, and an infinite queue of wafers in each independent queue. Simulation runs lasted for one 12-hour shift, with twenty repetitions to average out the occurrence of random events. In order to capture the marginal benefit of an additional operator, the first simulation run starts with just one operator for each pilot area. An operator is added on each round of simulation runs until ten operators are staffing each pilot area. Simulation output provides the number of lots complete, final tool utilization and lot cycle time. The simulated data does not provide an exact prediction of the output expected for each staffing level, since there are factors unrelated to queue length such as downstream queue conditions, tools out of service for extensive upgrades, or other dynamics

that would be known by a supervisor at the time of a staffing decision but are not captured in the simulation model. Because the simulation model takes into account the first order effects, the relative marginal value of adding an operator is still valid and considered to be useful decision support information by the supervisors.

Model Output

The simulation runs produced two sets of curves for each of the pilot areas (Figure 3-2). Production rates are given as a percent of maximum throughput, tool utilization is the percent of time the tools are processing, and cycle time is given in number of minutes to process each lot, including any waiting that occurs after leaving the initial queue. The horizontal axis represents the number of operators working for a full shift in the area. An operator working for half of a shift counts as half of an operator. The simulation curves represent the effective capacity which is the maximum possible output for each area. The actual pilot production and staffing levels are shown as individual data points below the simulation curves. Operators are needed to load and unload each tool before and after each processing step. Since the simulation assumes an infinite queue of material, the total number of tools and the average process time dominates the performance of each pilot area. Area A has 11 tools with a relatively short average processing time of 41 minutes, resulting in high demand for operator time. The first graph in Figure 3-2 shows that throughput and tool utilization both increase dramatically until a staffing level of about five operators. The marginal benefit of additional operators in Area A drops as the throughput approaches the limit determined by tool processing speed. With only one operator, processed material waits to be unloaded from tools, causing high cycle times. As operators are added, they become available to load and unload tools, resulting in a drop in cycle time, increased tool utilization and increased throughput. The actual results indicate that output stays relatively close to effective capacity for Area A, and staffing levels are typically kept near the steep slope at the left end of the curve.

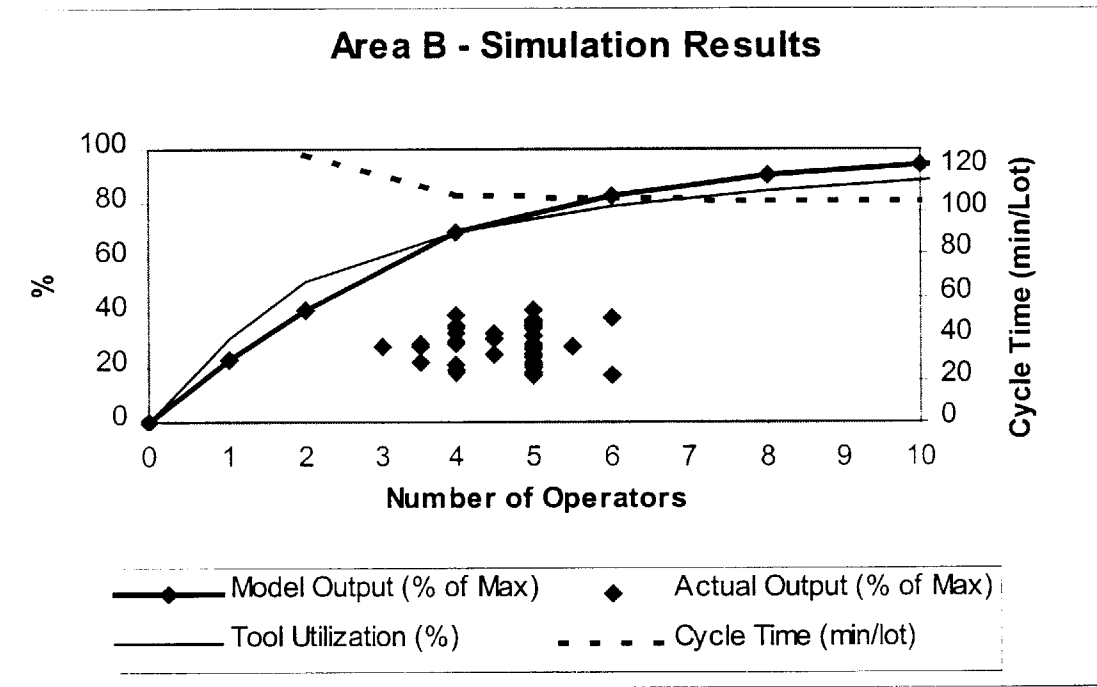
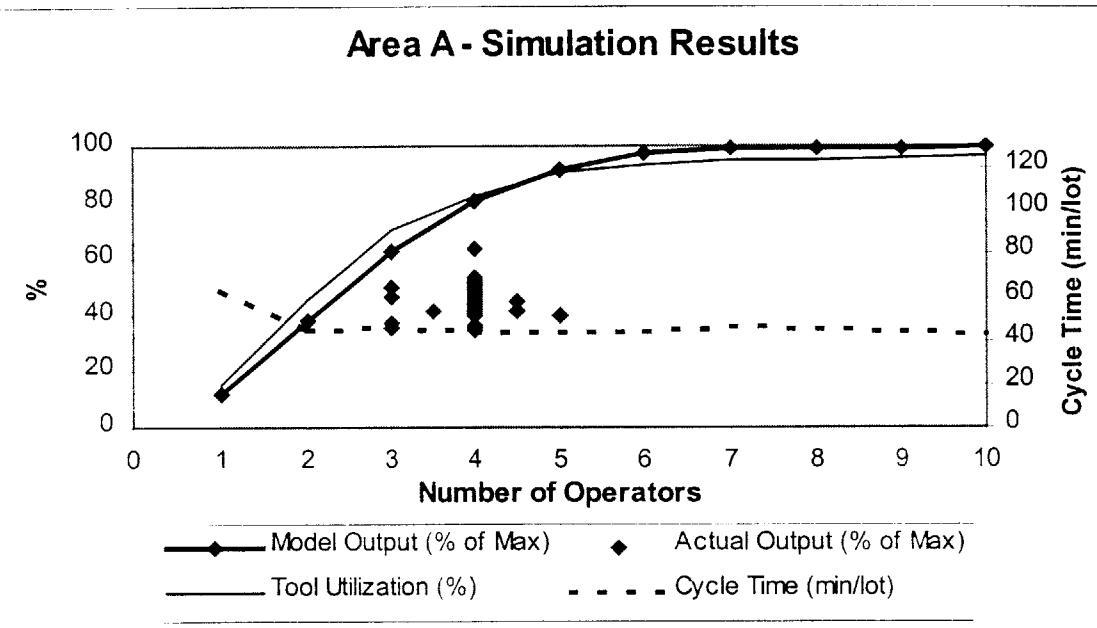


Figure 3-2 Simulation Results

Area B has 23 tools with a long average processing time of 100 minutes. The curves for Area B (Figure 3-2) more slowly approach the limiting case because there are so many tools to load and unload. Additional operators fill in gaps caused by break schedules and better respond to simultaneous demands for operator time. Because of the long processing time and

the large number of tools, the marginal increase in output for each additional operator is less in Area B than in Area A. The actual results indicate that output stays well below effective capacity for Area B, and staffing is kept near the middle of the curve where it has already started to flatten out. This indicates that Area B is more likely to be overstaffed than Area A.

Tactical Decision Support

The simulation output is used to assist in the allocation of operators across the pilot areas, rather than to specify an optimal solution. The role of a decision support tool is to get the user into the neighborhood of a good solution that can then be modified manually to accommodate constraints not captured in the model. Any model simplifies the problem in order to be generally applicable to many situations. As a result, model output should be combined with judgment and data from other aspects of the real problem. The decision maker should have an idea of the consequences of simple alternatives that might not dominate the decision, but should at least be considered in making it (Conway, 1967). By understanding the expected availability of equipment, material, and operators for each pilot area, supervisors can better distribute operators to areas of high workload to provide additional burst capacity without significantly lowering output in low workload areas.

Workload Report

Current factory reports do not present data separated and summarized by staffing units. In order to predict the incoming WIP for a shift, supervisors must spend time every shift reorganizing and recalculating the standard reports. To support tactical decision-making for the supervisors in the pilot areas, a separate report was created to present workload information in the proper grouping, together with the results of the simulation. This report captures both current queue levels and expected arrivals based on historical cycle times and current queue levels for upstream process steps. An additional piece of data in the report shows ergonomic impact in terms of number of lot lifts and lowers required by each operator to meet a given level of output for each staffing level. The estimate is based on predicted material flow and the ergonomic model for each pilot area. This reminds the supervisor that operators should not be exposed to repetitive motions beyond a factory-wide threshold. The supervisors compare current workload, the marginal output tradeoffs of transferring

operators, and ergonomic impact for different staffing levels to help decide where to best utilize cross-trained operators during the shift. By knowing the actual workload as a percent of equipment capacity, supervisors can use the curves in Figure 3-2 to assess the relative costs and benefits of overstaffing or understaffing each area.

4 Results and Conclusions

Analysis

The role of labor in factory performance is not clearly understood and labor policies are not consistently deployed, making it difficult to precisely measure the costs and benefits of implementation. Staffing movements and equipment assignments can be highly variable and are usually not tracked on a very detailed level. Current staffing models take a static approach by modeling labor based on average behavior. Because capacity for a specific group of equipment over a specific shift is the product of complex interactions between sequencing and the simultaneous availability of material, equipment, and operators, short-term factory performance is not well predicted by static analysis.

This creates an obstacle to understanding and demonstrating good or bad policies, and creates a blind spot for potential improvements in factory performance.

For the pilot, Intel's four shifts are divided into two that undertook addition cross training (the Cross-trained shifts) and two that did not (the Control shifts). The goal of the pilot is to increase factory output. Since output is a function of equipment availability, material availability, product mix, and operator availability, it is difficult to isolate the contribution due to cross-training by looking at output alone. Instead, the machine interference metric discussed in Chapter 2 is used to isolate the effect of operator availability on tool processing time. Since labor practices and machine interference performance are similar across all four shifts before cross-training, it is assumed that any relative change in machine interference between the Cross-trained and Control shifts is attributable to deploying cross-trained operators on the Cross-trained shifts. By increasing the flexibility of the workforce, cross-trained operators can move away from areas that are running significantly below capacity without much of a loss in productivity. These cross-trained operators then become available to respond to workload spikes in other areas that are capacity constrained.

Because Intel uses automatic equipment with more tools than operators, additional operators can increase the amount of time the tools spend processing rather than waiting for an operator to load or unload material. The simulation results of Chapter 3 model the relative

benefits of adding operators to either pilot area. For normal operations, some baseline staffing of each pilot area is usually maintained. Cross-trained operators should be deployed so as not to disrupt this balance when workload is evenly distributed between pilot areas. When a flood of material causes a workload spike in one pilot area, that area becomes capacity constrained. As long as the other pilot area is not similarly capacity constrained, cross-trained operators can respond to the workload spike to provide burst capacity. The performance gain will be reflected in lower machine interference and a corresponding higher processing time.

Cross-trained vs. Control Shifts Across Pilot Areas

Any performance difference between the Cross-trained and Control shifts is captured in the machine interference metric. Machine interference is calculated for Baseline values (before pilot), Control shifts (after pilot starts), and Cross-trained shifts (after pilot starts) as illustrated in Figure 4-1. During the course of the cross-training, there was a shift in product mix with less material processed through both Area A and Area B while a new product is ramped to full production. The charts demonstrate how actual tool capacity was distributed among the four possible equipment states: Processing, Down, Starved, and Machine Interference (defined in Table 2-1). Processing time represents throughput, while the other three states represent non-value-added burdens on equipment capacity. Since equipment availability is not a by-product of manufacturing policy decisions, it is viewed as an external constraint. Any effort to improve the non-value-added states is only successful if it improves the availability-utilization gap, either by reducing starvation or reducing machine interference.

For Pilot Area A, machine interference is 30.1% during Baseline observations and stays steady at 29.9% for the Control shifts during the pilot study. Area A has only 25.8% machine interference on the Cross-trained shifts, however, representing a 4.1% improvement on the Cross-trained shifts relative to the Control shifts. There is a 3.4% improvement in processing time as well as a 1.1% decrease in starvation time, balanced by a 1.8% increase in downtime. It is important to note that the improvement in Machine Interference did not

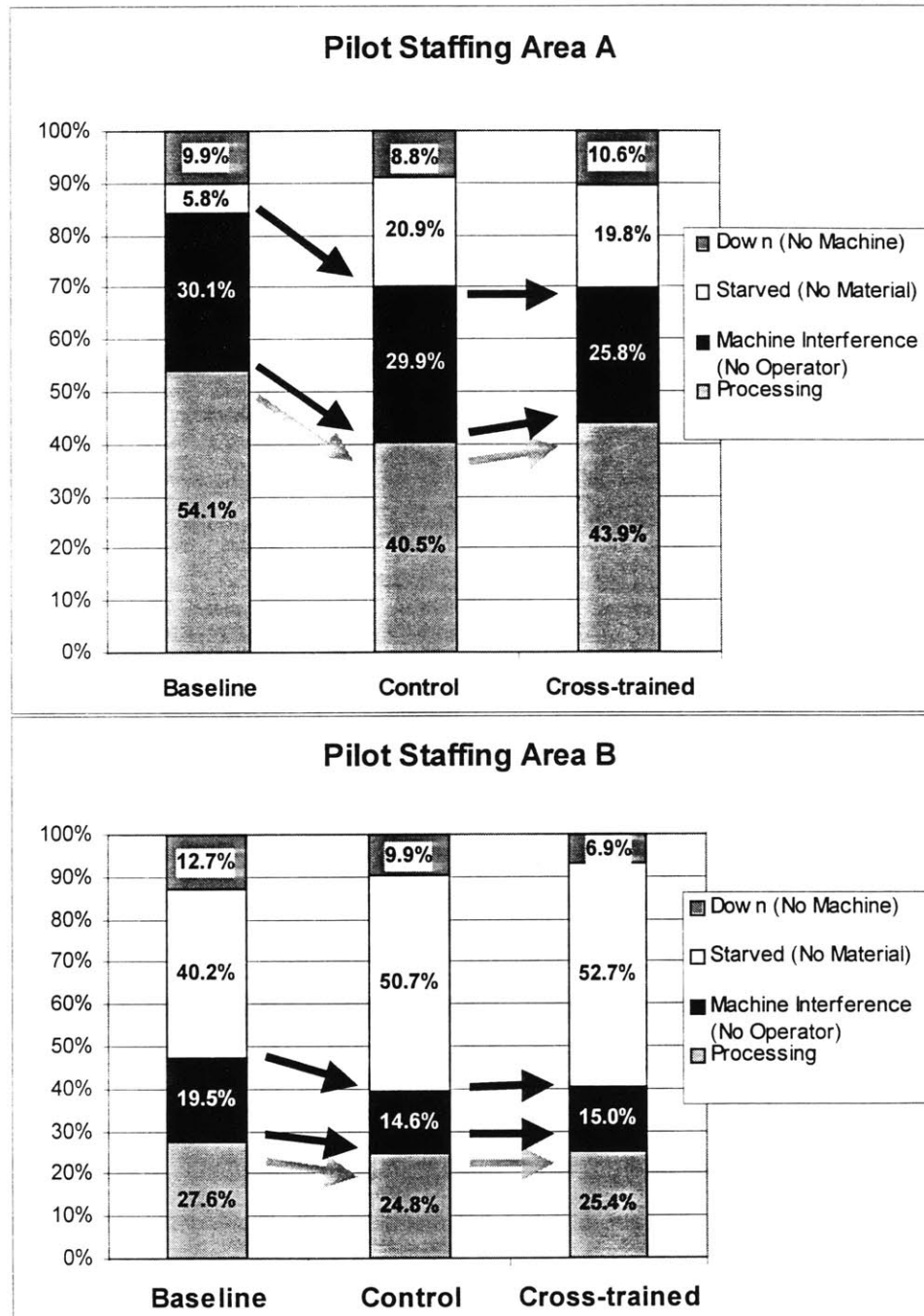


Figure 4-1 Results for Each Pilot Area

translate into an increase in Starved time. Any increase in Starved time would occur when an operator moves to an area with low demand relative to capacity, where she quickly processes the entire queue of material, resulting in a starved state (and an idle operator). On the other hand, if the operator moves to an area with high demand relative to capacity (a workload

spike), all of that operator's additional work can keep equipment processing without running short of material.

After the Baseline period, material queues decreased in Areas A and B to phase in a new product line. This caused a significant jump in the frequency of equipment starvation for both the Control and Cross-trained shifts for both pilot areas. As a result, the processing times during the Baseline period are higher than the processing times after cross-training was completed. The equipment availability is not systematically affected by the shift in product mix, and any difference in down time between the Baseline, Control, and Cross-trained observations for either Area A or B are assumed to be random fluctuations.

Area B is characterized by a high availability-utilization gap, composed of both starvation and machine interference. With such high starvation (low workload), the benefit of additional operators is expected to be negligible for Area B. Both the Control shift and the Cross-trained shifts demonstrate improved machine interference performance than the Baseline period, with a 4.9% improvement in the Control shifts and a 4.5% improvement on Cross-trained shifts. However, for areas with low workload, machine interference does not have a clear link to performance improvement. As operators are added to an area with low workload, machine interference will quickly be eliminated, as material queues become very short. This indicates that instead of having idle equipment with short queues, adding operators runs the queues down to zero, causing idle equipment with zero queues. In areas of low workload there is idle equipment, regardless of whether or not there is machine interference. The performance of Area B across all observations is dominated by equipment starvation, which occurs greater than 50% of the time during both the Control and Cross-trained observations. With such low material queues, improving machine interference performance would most likely show up as increased starvation, without impacting the availability-utilization gap. As expected, Area B does not seem to benefit from the flexibility of a cross-trained workforce. Instead the system works as planned; operators can be moved from an area (B) of low demand to and area (A) of high demand and improve the performance of the high workload area without negatively affecting the performance of the low workload area.

In summary, significant improvement occurred in Area A, where there is a moderate workload. Comparing the Control and Cross-trained shifts for the overall pilot study shows a significant reduction in availability-utilization gap for the shifts using cross-training in Area A. There was no relative performance difference between the Control and Cross-trained shifts for Area B, where there is a light workload. Cross-training does not show a significant reduction in the availability-utilization gap for the shifts using cross-training in Area B. Both Area A and Area B have high workload variability. The results indicate that cross-training is effective in areas with both high workload and high workload variability, and doesn't detrimentally affect those areas with low workload that are called upon to share their workers.

Statistical Significance

For the purposes of the pilot experiment, Intel's four shifts are divided into two Cross-trained shifts and two Control shifts. The two Cross-trained shifts have operators cross-trained between Area A and Area B. The two Control shifts do not have any operators cross-trained between these areas. The main performance variable is the machine interference metric discussed in Chapter 2. The statistical significance of the cross-training pilot is based on the difference in machine interference performance between the Cross-trained and Control shifts both before and after operators have been cross-trained. Since there are no systematic differences between the operator practices of any shifts before cross-training, it is expected that the performance of the Cross-trained and Control shifts will be the same in the period before cross-training was completed. After cross-training it is expected that any performance difference can be attributed to the cross-trained workforce that was only available for the Cross-trained shifts.

Analysis was performed using the Paired t-Test for means with a hypothesized difference of zero as described by Vining (1998). Observations were collected for 19 days before cross-training and 19 additional days after cross-training, resulting in 19 data points for each pilot area both before and after cross-training was completed. This results in 18 degrees of freedom for determining the critical region for Area A or Area B alone, and 37 degrees of freedom for determining the critical region for Area A and B combined. The hypothesis of

the test is that there is no difference between the Cross-trained and Control shifts for either of the time periods investigated. A two-tailed distribution is used to define the critical region because any departure from zero difference in machine interference represents a performance difference between the Cross-trained and Control shifts. Tests use a critical region bounded by $|t| = 2.1009$ for 18 degrees of freedom and a critical region bounded by $|t| = 2.0262$ for 37 degrees of freedom. Results are summarized in Table 4-1. In the period before the completion of cross-training, Cross-trained and Control shifts had similar machine interference performance for Area A and Area B. The conclusion is that there is no statistical difference in performance between the Cross-trained and Control shifts for before cross-training. This demonstrates that there is no pre-existing bias between the shifts selected for the Cross-trained and Control groups before the cross-training actually started.

Pilot Area	Machine Interference Comparison	t-Statistic (t Critical)	Conclusion
Area A	Two Cross-trained shifts vs. Two Control shifts before cross-training	-0.5580 ($ t < 2.1009$)	Fail to reject hypothesis. No statistical difference between Area A Cross-trained and Control shifts before cross-training.
	Two Cross-trained shifts vs. Two Control shifts after cross-training	2.2458 ($ t > 2.1009$)	Reject hypothesis. Cross-trained Area A performance is statistically different than Control performance after cross-training.
Area B	Two Cross-trained shifts vs. Two Control shifts before cross-training	0.6481 ($ t < 2.1009$)	Fail to reject hypothesis. No statistical difference between Area B Cross-trained and Control shifts before cross-training.
	Two Cross-trained shifts vs. Two Control shifts after cross-training	-0.2278 ($ t < 2.1009$)	Fail to reject hypothesis. No statistical difference between Area B Cross-trained and Control shifts after cross-training.

Table 4-1 Results of Statistical Analysis

With cross-training completed, machine interference performance is improved in Area A for Cross-trained shifts relative to Control shifts to a statistically significant degree. Area B did not display a statistically significant change in performance between the Cross-trained and Control shifts. The interpretation of these results is that Area A has a high workload and high workload variability and can benefit from having cross-trained operators to reduce machine interference and reduce the availability-utilization gap. Area B has a low workload and high workload variability and was not harmed by the additional cross-training.

Frequency of Transfers

Operators are transferred between Areas A and B as part of the targeted cross-training efforts, but operators can also be pulled in from other areas (using pre-existing cross-training). The effect of both of these actions is to add an additional operator to the pilot area, though in the case of transfers from outside areas, it is not clear how the staffing is covered in the outlying home area. Because of the difficulty of accurately tracking operator assignments, operator transfers are tracked over a period that does not exactly overlap with the pilot data for machine interference, but is representative of the frequency of transfers throughout the pilot study. Observations are made for 14 Control shifts and 36 Cross-trained shifts as summarized in Table 4-2. For the control shifts, only one operator is transferred to area A (from an outside area) in 14 shifts worth of observations. For the Cross-trained shifts, 17 transfers are made over 36 shifts worth of observations. This includes four transfers from Area B to Area A and three transfers from Area A to Area B using the newly cross-trained operators for the pilot. This also includes ten transfers in from other areas outside of the pilot areas. The Cross-trained shifts are much more aggressive about changing staffing levels while the control shifts are content with static staffing assignments.

Transfer	Control	Cross-trained
Add operator to Area A	From Area B: 0 From other areas: 1	From Area B: 4 From other: 6
Add operator to Area B	From Area A: 0 From other areas: 0	From Area A: 3 From other: 4
Total transfers	1 transfer/14 shifts	17 transfers/36 shifts

Table 4-2 Operator Transfers

Implementation

Intel has a well-developed training program, so it is not difficult to get operators cross-trained. Industrial engineering literature and the simulation results from this research indicate that there is a benefit from having cross-trained workers. The pilot study demonstrates improved performance for the high workload and high workload variability of pilot Area A with no measurable impact for the low workload and high workload variability of Area B. However, theoretical models and controlled pilot studies do not insure long-term success in the real world. Even if a proposed program demonstrates unambiguous improvement, threatened constituencies can still undermine implementation. Implementation requires buy-in from key decision-makers, clear communication to all affected constituencies, and project priority backed up by management actions.

Implementation Steps

Implementation of the pilot program follows a series of steps that provide a framework for future expansion of the cross-training program to other areas. The starting point is identifying the decision-makers and constituencies who will eventually be responsible for execution of the cross-training program or who will be directly affected by the cross-training program. It is important to strike a productive balance between challenging conventional thinking and acknowledging what is actually feasible. The specific sequence of steps recommended for future expansion is:

1. Identify decision-makers and affected constituencies
2. Involve decision makers
3. Map current operator certifications to staffing areas
4. Identify staffing areas with complementary workloads
5. Select target staffing areas for cross-training
6. Establish baseline performance metrics for target staffing area
7. Identify operators who match operator selection criteria
8. Begin operator cross-training in target areas
9. Monitor performance after cross-training is complete

10. Analyze results

11. Make system self-sustaining through intelligent automation

These represent the steps successfully employed during this study to implement the pilot cross-training program. It was not unusual for several iterations between steps as new information surfaced. Accurate operator certification records help identify candidate equipment/operator pools that can benefit from existing training and can also help identify prospective flexible workforce team members. Workload correlation between candidate staffing areas is analyzed to identify areas where cross-trained operators will have the greatest impact. Performance tracking is most constructive if baseline data is tracked before deploying the flexible workforce and then continued through implementation. This helps assess the relative impact of the flexible workforce separate of other factory variables and provides feedback for further improvements. Once the system is running smoothly, it is important to make the system self-sustaining, where the participants in the system have responsibility and authority to make changes to the system. This role can often be simplified by the intelligent use of information technology.

Data Resolution

Intel's wafer fabs have no shortage of factory data generation. Transactions are constantly generated at every processing step to monitor both material flow through the factory and process conditions at each step. Because of the massive volume of data generated, there needs to be some reasonable filtering of data into meaningful feedback for the different manufacturing constituencies. Operators tend to be more concerned with real time information about equipment status and lot status. Engineers focus on statistical excursions from stable behavior in process conditions as well as quality inspection results. Supervisors want up-to-date information on how their specific department is performing on the current shift. The management team wants summarized factory performance at the end of a shift, and detailed information about any deviations from expected performance. Individual reports as well as entire computer systems have emerged to address these diverse information needs. It is important to store factory data at a fine enough resolution so that data can be re-aggregated to meet different informational requirements. Otherwise, information needs are left

unanswered. For example, pilot data are required for individual factory tools, but historical data are already summarized for all tools with the same process technology. This prevents supervisors from getting feedback on the effectiveness of their staffing policies.

Current data collection and reporting at Intel do not allow easy analysis of labor performance. Without improvements in this area, analysis of the labor impact of staffing modifications will continue to be a highly labor-intensive activity itself. Feedback systems live and die by the quality and reliability of the data. Quality means that observed events match reported events, with all exceptions clearly understood. This is extremely difficult to do for new pilot reporting systems, which typically overlook many of the complexities involved in the mature data collection systems. Tactical planning and reporting tools need to be accessible from computers on the factory floor in the hands of the people running the equipment. This includes 24-hour x 7-day technical support for the systems that report the data. An improvement in information delivery will drive more effective deployment of cross-trained operators.

Keys to Success

Successful implementation requires the resources and support to overcome obstacles and maintain a clear priority within the factory. A strong mandate from management is required to ensure successful training and utilization of cross-trained operators. The foundation of this project involves the initial training of operators in new staffing areas. This causes an immediate negative effect in the new area as a senior operator is drawn away from his or her primary responsibilities to serve as a trainer. Even after the initial training is complete, there is still a period where the newly trained operator is honing skills in the new area. It is often difficult to convince a staffing area to absorb the short-term burden of training in return for the long-term benefit of increased flexibility. By clearly communicating the priority of cross-training, even over short-term performance, the management team sets the tone and establishes the credibility of cross-training as a lasting factory initiative.

Implementation can be simplified by working to align the path to success with the path of least resistance (recognizing that the latter is more likely to occur naturally). An example of

this type of thinking is to track the cross-training that happens as a result of employee turnover and movement from area to area within the factory, rather than mandating independent cross-training efforts. Typically, as operators move off their previous jobs, they eventually lose touch with their previous skill set. With better visibility of pooled operator certifications across the factory, existing options for flexibility can be maintained with minimal training disruption. A planned rotation back into the former staffing area protects the knowledge base as well as the potential to provide burst capacity without the short-term pain of additional training. One pilot shift identified an operator and an equipment technician who recently had moved to new positions. As a result, one day every two weeks they rotate back into their former staffing area for a shift. This provides a simple way of leveraging existing skills that would otherwise go to waste.

Demonstrating performance has beneficial results for both target staffing areas and prospective staffing areas for implementation. Areas that are successful in deploying a flexible workforce get immediate performance feedback on their efforts and motivation to continue expanding as long as performance continues improving. Prospective staffing areas that have not yet implemented a cross-training program are presented with a quantifiable performance benefit that they can expect from participating in the program. This provides the incentive for prospective staffing areas to accelerate the implementation schedule in order to capture the performance improvements as soon as possible. Without clear performance feedback, it becomes extremely difficult to get consensus on decisions between the multiple constituencies with a stake in staffing policies. Demonstrating performance is a key driver for making any initiative weave its way into the fabric of daily factory operations.

Set deployment rules are necessary to avoid multiple interpretations of the same data. At some level of expected workload or with some number of tools down an additional operator must be moved to better allocate factory resources in response to changing conditions. Providing resources to one area means making a capacity tradeoff with another area. This tradeoff must be analyzed to predict the best total allocation of operators. Ideally, the areas involved are chosen due to a negative correlation of workload requirements through time. Nevertheless, rules must be in place to ensure that the required resources are in fact moved to

the area that needs them the most. Clearly stated rules and objectives will help insure that the benefits of flexible labor pools are consistently delivered.

It is important to anticipate training and ramping complexities during the implementation phase. In selecting pilot areas for each shift, data was used for the product mix initially running in the factory. As the training proceeded, the development factory moved into preparation for the next generation product ramp. This caused a shift in management priorities where cross-training fell behind the higher priority installation and qualification of new process technologies. Conflicts in training and installation priorities contributed to difficulties in completing the pilot for some of the shifts. Because of the new technology ramp and a shift in product mix, there was a shift in the distribution of workload between staffing areas causing a reevaluation of the selected areas. The training and ramping complexities were not well anticipated in this study, and it is recommended that future work should incorporate full consideration of planned training and ramp activities along with predictions of future factory product profile.

Conclusions

The flow of material through Intel's wafer manufacturing environment is affected by a large number of interrelated events. Individual lots go through hundreds of processing steps on equipment characterized by a high frequency of preventive maintenance, equipment failures, and process changeovers. This presents a formidable challenge for scheduling the factory and then executing the schedule. Because of the high costs of capital equipment, the amount of variation in the availability of resources, and the make-to-stock environment, schedulers must strike a balance between overloading the factory with too much material and allowing expensive resources to sit idle. Schedulers respond to day-to-day fluctuations by modulating wafer starts into the factory. Once a schedule has been set, the shop floor must work to maintain the flow of production. Cross-training provides a flexible labor pool to help the shop floor react to factory variation.

Cross-training Benefits

The success of any business improvement is always measured in terms of the impact on the bottom line. This thesis shows that cross-training does indeed make a positive contribution to the bottom-line of the company by reducing machine interference and thus increasing machine processing time. Increased processing time in turn reduces costs by lowering average inventory levels, increasing equipment utilization (throughput), and preventing late orders. Cross-training is, as expected, most valuable in areas with high workload and high workload variability. A flexible workforce positions the factory to better respond to disruptions in material flow. Efforts should be made to eliminate variation wherever possible, but some type of resource buffer is needed to respond to variability that cannot be avoided. These resources can be additional material buffers, additional equipment capacity, or additional operator capacity (in the case of a dual constrained system). A flexible workforce is a relatively inexpensive way to prevent temporary disruptions from negatively impacting factory performance.

The machine interference metric creates a common understanding of how staffing decisions affect overall factory performance. Machine interference captures how well the factory is providing operators where and when they are needed. One benefit of the machine interference metric is to help create a shared understanding of how critical it is to have the manufacturing resources of material, equipment, and operators all come together. Since machine interference is viewed from the perspective of the equipment, improvements in machine interference can be directly linked to more efficient use of capital equipment. Ideally, labor performance metrics should be captured directly. However, since labor is the most dynamic, flexible, and variable resource in any factory environment, it is difficult to meaningfully catalog operator activities and unambiguously link each activity to overall system performance. The machine interference metric is calculated indirectly in order to focus only on the availability of operators when equipment and material are both ready to run.

The simple act of tracking the cross-training history of all operators is helpful in assessing current skill sets and speeding implementation of a flexible workforce. Accurate records

make it clear who is certified in what areas and increases the awareness and visibility of cross-training options. Also by continuing to communicate the relationship between the availability of equipment, material, and operators, better tactical decisions can be made in operator staffing assignments.

Implementing a cross-training program has several second order benefits that are not easily linked to financial performance. Having a cross-trained workforce creates a variety of job assignments in a repetitive factory environment. This helps improve morale by providing changes to the daily routine and a chance to mingle with different co-workers. Even when there is not a workload spike, rotating operators keeps skills sharp in multiple areas and lowers the ergonomic exposure to repetitive motions. From the supervisor's perspective, having cross-trained operators increases the tactical options available to respond to different factory conditions. By regularly thinking in terms of sharing resources across the factory, the decision to assign cross-trained operators to a given staffing area encourages supervisors to think about improving performance at a system-wide factory level rather than at a local department level.

Data Collection and Performance Evaluation

One of the obstacles in establishing analytical treatment of labor is that current factory systems do not collect and report data from the perspective of staffing areas. It is critical to track progress and evaluate performance to provide meaningful feedback for pilot participants. If individual operators cannot be clearly linked to specific performance on a specific shift in a specific staffing area, then it is hard to demonstrate a clear cause and effect relationship between operator actions and performance measures. Resolving this requires either a major information technology overhaul to handle increased data resolution, programming custom solutions for ad hoc requests, or approximating and inferring information as best as possible from existing reports. As a result of time and budget constraints the latter two solutions are usually pursued, building in additional layers of complexity and confusion for future data collection and reporting efforts. The recommended approach, which is being pursued at Intel, is to make temporary changes on an as needed basis, while pursuing an aggressive data standardization program to create a common

database with the appropriate level of data resolution across the entire manufacturing organization.

Future Direction

Results of the pilot study suggest several areas that require further study. Cross-training is valuable for areas with high workload and high workload variability. This research explored the impact of cross-training on one area (Area B) with low workload and high variability and on one area (Area A) with higher workload and high variability. Greater benefits may be realized from implementing cross-training in areas with even higher levels of workload relative to capacity, and even higher variability relative to average workload. These are areas that merit further study. The machine interference calculation is a first step in establishing a feedback mechanism to link labor deployment to factory performance. Data for this study are primarily collected from tracking equipment and material status. Future studies would be better served by detailed operator tracking to calculate machine interference directly rather than indirectly. If the distribution of the labor and equipment pools shifts frequently over the course of a shift, then information on individual operator and equipment interactions and sequences might be necessary to understand the dynamics of the system. On a strategic level, it would be useful to apply the cross-training analysis to recommend how staffing areas should be formed during plant layout discussions. Continued expansion and tracking of the cross-training program at Intel should provide more information on how best to leverage a flexible workforce.

Appendix A

Model Inputs

First in first out processing, Infinite queue (no starvation)

Scope	Variable	Variable Type
Area	Area name Tool family Number of tools in each family	Label Cross-reference Constant
Operator	Operator class Number of operators Break schedule	Label Constant (Varied between runs) Constant (staggered for each oper)
Operator/Tool PM Events	Mean time between preventive maintenance (hours/pieces) Mean time to repair (hours) First occurrence time (hours) First occurrence time (pieces)	Uniform ($\pm 10\%$) Uniform ($\pm 10\%$) Uniform (MTBPM/2 $\pm 10\%$) Weibull (3, MTBPM/2)
Tool family	Tool family Operator class Operator load time Operator unload time Operator get (from queue) time Operator put (to next queue) time Processing time Lot ID time (where applicable) Tool capacity (lots/tool)	Label Cross-reference Constant Constant Constant Constant Constant Constant Constant
Tool PM Events	Mean time between preventive maintenance (hours) Mean time to repair (hours) First occurrence time (hours)	Uniform ($\pm 10\%$) Constant Uniform (MTBPM/2 $\pm 10\%$)
Tool Failure Events	Mean time between preventive maintenance (hours) Mean time to repair (hours) First occurrence time (hours)	Uniform ($\pm 10\%$) Constant Uniform (MTBPM/2 $\pm 10\%$)

Model Outputs

Scope	Metric
Throughput	Lots complete Pieces complete Average lot cycle time Final lot cycle time
Tool	Utilization (%) = Load (%) + Unload (%) + Processing (%) Down (%) Starved (%) Machine Interference (%)
Operator	Operator Utilization (%) = Load (%) + Unload (%) + Processing (%) + Operator PM (%) Break (%) Idle (%)

Bibliography

- Conway, R.W., Maxwell, W.L. and Miller, L.M. (1967). *Theory of Scheduling*. Reading, Massachusetts: Addison-Wesley Publishing Company.
- Fryer, J.S. (1974). Labor flexibility in mutiechelon dual-constraint job shops. *Management Science*, 20, 1073-1080.
- Gershwin, S.B. (1994). *Manufacturing Systems Engineering*. Englewood Cliffs, New Jersey: Prentis Hall.
- Holstein, W. and Berry, W. (1970). The labor assignment decision: an application of work flow structure information. *Paper No. 291, Herman C. Krannert Graduate School of Industrial Administration (Purdue University)*.
- Hopp, W. and Spearman, M. (1996). *Factory Physics, Foundations of Manufacturing Management*. Boston: Irwin McGraw Hill.
- Johri, P.K. (1993). Practical issues in scheduling and dispatching in semiconductor wafer fabrication. *Journal of Manufacturing Systems*, 12, 474-485.
- Malhotra, M. and Ritzman, L. (1990). Resource flexibility issues in multistage manufacturing. *Decision Sciences*, 20, 673-697.
- O'Ferrell, D.S. (1995). Manufacturing Modeling and Optimization. *IEEE/SEMI Advanced Semiconductor Manufacturing Conference*, 334-339.
- Stafford, E.F. (1988). An optimal solution technique for the operator-machine assignment problem. *Production and Inventory Management Journal*, 29, 25-30.
- Stecke, K.E. (1992). Machine interference: The assignment of machines to operators. *Handbook of Industrial Engineering, 2nd ed*: 460-495.
- Treleven, M.D. (1989). A review of the dual resource constrained system research. *IIE Transactions*, 21, 279-287.
- Uzsoy, R., Lee, C.Y. and Martin-Vega, L.A. (1994). A review of production planning and scheduling models in the semiconductor industry - part II: shop-floor control. *IIE Transactions*, 26, 44-56.
- Uzsoy, R., Lee, C.Y. and Martin-Vega, L.A. (1992). "A review of production planning and scheduling models in the semiconductor industry - part I: system characteristics, performance evaluation and production planning." *IIE Transactions*, 24, 47-61.
- Vining, G.G. (1998). *Statistical Methods for Engineers*. Pacific Grove, CA: Duxbury Press.