Process Time Variation Reduction in a Microprocessor Burn-in Operation

by

Terence G. Emmert

B.S., Aerospace Engineering, U.S. Naval Academy, 1988
M.S., Aerospace Engineering, Naval Postgraduate School, 1995

Submitted to the Sloan School of Management and the Department of Civil and Environmental Engineering in Partial Fulfillment of the Requirements for the Degrees of

Master of Science in Management and
Master of Science in Civil and Environmental Engineering

In conjunction with the Leaders for Manufacturing Program
at the Massachusetts Institute of Technology
June 2001

Signature of Author____
MIT Sloan School of Management
Department of Civil and Environmental Engineering
May 1, 2001

Certified by_________
Donald B. Rosenfield
Senior Lecturer, Sloan School of Management
Thesis Advisor

Certified by_________
David Simchi-Levi
Professor, Department of Civil and Environmental Engineering
Thesis Advisor

Accepted by___
Margaret Andrews
Executive Director of the MBA Program, Sloan School of Management

Accepted by______
Oral Buyukozturk
Chairman, Committee on Graduate Studies
Department of Civil and Environmental Engineering

Process Time Variation Reduction in a Microprocessor Burn-in Operation

by

Terence G. Emmert

B.S., Aerospace Engineering, U.S. Naval Academy, 1988
M.S., Aerospace Engineering, Naval Postgraduate School, 1995

Submitted to the Sloan School of Management and the Department of Civil and Environmental Engineering on May 1, 2001 in Partial Fulfillment of the Requirements for the Degrees of

Master of Science in Management and
Master of Science in Civil and Environmental Engineering

Abstract

Intel uses the Theory of Constraints in the capacity planning process for its assembly and test operations. The test area of the factory consists of two primary operations, Burn-in and Post Burn-in Check (PBIC). While PBIC is the planned constraint for the factory, Burn-in has exhibited constraint-like behavior.

Capacity planning is done using static spreadsheet-based calculations that assume fixed values for tool run rates. An investigation revealed a high degree of processing time variation and queue time variation at each step of the Burn-in operation that is not accounted for in the capacity planning model. The hypothesis of this work was that the degree of variation was significant enough to consume excess capacity installed to account for both process variation and demand variation and thus cause Burn-in to be the bottleneck or at least cause PBIC starvation frequently.

After proving this hypothesis, the project sought to quantify the impact of process time variation reduction in terms of output and unit cost. A discrete event simulation was constructed which included data from the Intel's Costa Rica Assembly/Test Manufacturing Factory and was validated over a four-week period for a high-volume package type.

Studies conducted using the discrete event simulation showed that by reducing the processing time variation attributable to two manual operations a 54.1% increase in capacity and a 32.4% decrease in the unit cost attributable to Burn-in could be achieved. Additionally, an 80% decrease in the standard deviation of the throughput indicated that variation reduction would greatly increase the predictability of the Burn-in process.

This project also demonstrated the proof of concept for using the simulation as a dynamic analysis tool to supplement the existing static capacity planning model.

Management Thesis Advisor: Donald B. Rosenfield,
Senior Lecturer, Sloan School of Management

Engineering Thesis Advisor: David Simchi-Levi,
Professor, Department of Civil and Environmental Engineering

Acknowledgements

I would like to acknowledge the gracious and generous assistance and guidance offered by the employees of the Intel Corporation. Special thanks to Yair Rodrigues, Enrique Saborio, Eric Hernandez, Arelene Say, Scott Headley, Dale Easter, Steven Ledford, and Michael Stewart.

Additionally, I thank Chris Richard, the project sponsor at Intel, for his unwavering support and guidance, and my academic advisors at MIT, Don Rosenfield and David Simchi-Levi, for their technical insights.

I am particularly grateful for the support of my classmate Jon Frommelt whose deep knowledge of Intel's culture and technical infrastructure were critical in allowing me to complete this work. Thank you Jon for your friendship and love of bicycles and Wang's.

Finally, my most heartfelt gratitude is reserved for my wife Cheri and daughter Madeline who have supported me in every way possible during the highs and lows of the LFM program. The project and this thesis are as much their achievement as mine.

Table of Contents

List of Figures

Chapter 1: Introduction

1.1 Assembly/Test Manufacturing

Intel is facing ever-increasing market pressure to produce greater volumes of microprocessors at lower and lower prices. In this competitive landscape, any lost output for Intel is business its competitors will eagerly take. Satisfying its customers by promptly meeting demand is a critical priority for the firm.

To meet this demand, Intel must increase its capacity, but to maintain its historical earnings growth the firm must also be conscious of its cost of production. Microprocessor manufacturing consists of a wafer fabrication step, which occurs in dedicated wafer fabrication facilities or Fabs, followed by assembly and test steps, which occur in a factories specifically designed for these operations. While wafer fabrication has traditionally constituted the vast majority of production costs, the assembly and test processes are growing in complexity and cost. Due to the relative low cost position of assembly and test in the production process the Assembly Test Manufacturing Group (ATM) operates on the guiding principle that it must never constrain Intel's production.

Intel's high volume assembly and test operations are located in Penang and Kulim, Malaysia, Cavite, the Philippines, and San Jose, Costa Rica. Essentially all of the Pentium™ family of products pass through one of these factories, and the highest volume product during the period of this study was packaged using a technology called Flip-Chip Pin Grid Array (FCPGA). This package type was chosen as the focus of this study because of the volume and the maturity of the process.

Newlin (Newlin, 2000) offers a comprehensive description of ATM operations, but for the sake of this discussion it can be roughly characterized by two areas, assembly and test, and the test area can be further grossly subdivided into two areas called burn-in and Post Burn-in Check (PBIC). Figure 1.1 describes this structure. The focus of this project is the Burn-in area (BI).

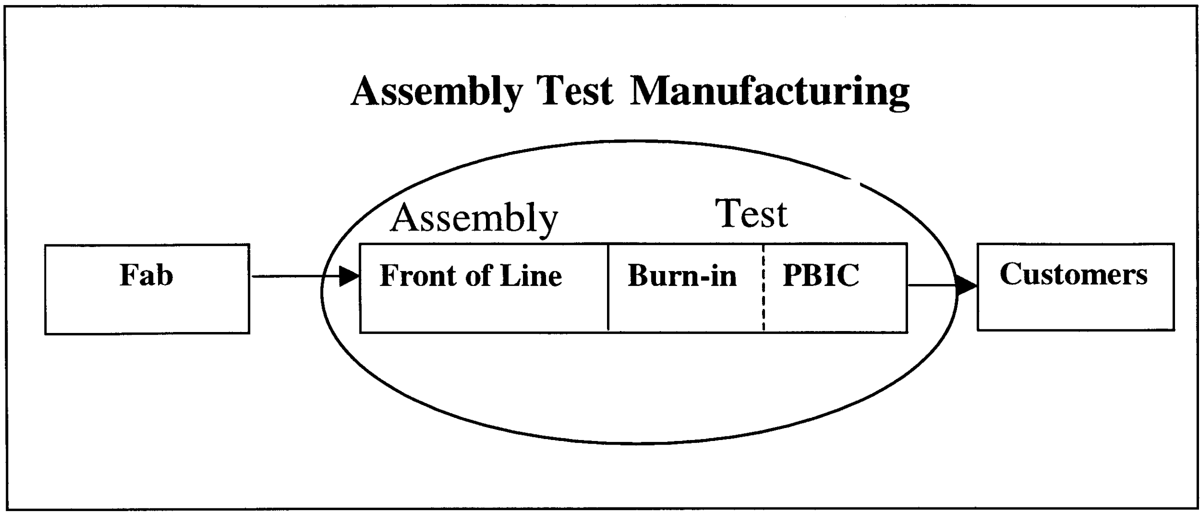


Figure 1.1 – Position of ATM in Intel's Production Process

1.2 Burn-in

The reliability of a microprocessor throughout its life can be described by the "Bathtub Curve" Figure 1.2. The purpose of BI is to accelerate the life of the product by applying a recipe of electrical signals and elevating the ambient temperature. By doing this, product that are inclined to fail in the early stages of life, the "infant mortality stage", will do so and can be identified prior to release to customers.

BI is performed in a Burn-in Oven and the time required for the product to be subjected to this process varies based on the product type and maturity. When a product is first introduced the BI time is relatively long (on the order of 12 hours). But as the product matures and reliability data is gathered the product engineering group will reduce the BI time (on the order of an hour for mature products). Because the product spends a significant amount of time in BI there is an initiative to incorporate a greater degree of functional testing in Burn-in. This trend creates a demand for a next generation of more sophisticated and more expensive Burn-in tools.

The interface between the product and the BIO is a tooling item called a Burn-in Board (BIB). To begin the BI process, the product is loaded onto a BIB using a machine called a Burn-in Loader/Unloader (BLU). Figure 1.3 is a process flow diagram for the BI area.

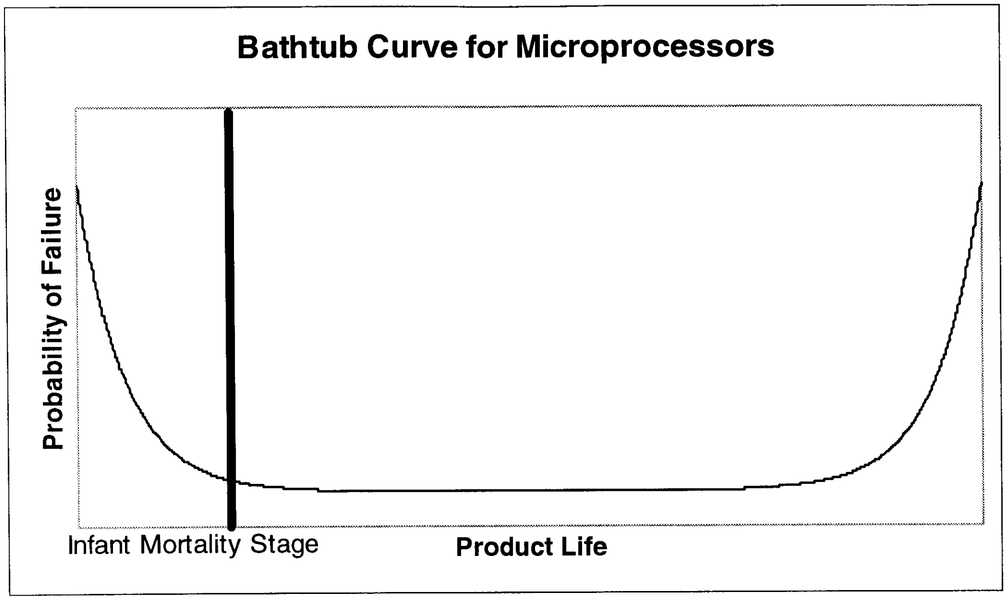


Figure 1.2 – Bathtub Curve for Microprocessors

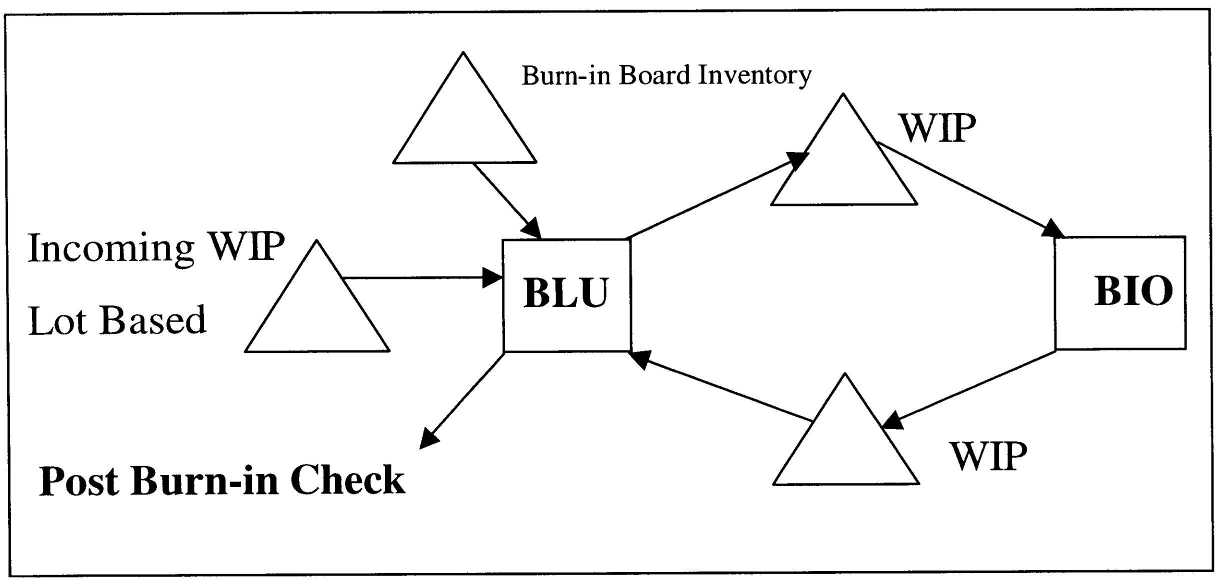


Figure 1.3 – Burn-in Process Flow

A BIB can accommodate a variable number of product units based on the size of the unit and its thermal characteristics. For larger products such as microprocessors used in servers, there can be a few as six units to a BIB. For the smaller products used in PCs there can be as many as 15 units to a BIB. Units are processed through BI in lots based on the lot policy used during assembly and which typically consists of 1000 units. Once the units have been loaded on BIBs, an electrical continuity check, called Vcc-to-Ground check is conducted manually to isolate product units with faults that cause a short in the package. If the ground check indicates a fault, the BIB in question is removed from the cart and a more detailed troubleshooting process is performed to isolate the specific faulty product so it can be removed from the lot. Once all of the BIBs are checked and all faulty units removed, the lot is moved to a queue in preparation for the manual load into the BIO. Lots are pulled from the queue using a first in first out prioritization. After being loaded, a diagnostic, called a pre-signal check is performed. This check is designed to determine if faults exist in the BIB and the BIO circuitry. If the pre-signal check provides a positive indication, the fault is manually isolated and repaired while the lot waits to be processed. Once the pre-signal check is complete, the Burn in procedure is executed. After the units are complete in the BIO, they are manually unloaded from the oven and placed in a queue to await being unloaded from the BIBs using the BLU. The BLU operation is the only re-entrant operation in ATM.

1.3 Capacity Planning

Intel plans capacity and manages its factories based on the Theory of Constraints (Goldratt, 1992). PBIC is planned and managed as the constraint because of the high cost of testers (approximately $2.5M) and BI is planned and managed as a near constraint.

Capacity planning for BI is accomplished by considering demand, yield and rework forecasts, manual processing times, equipment run rates and equipment availability. The run rate for the BIO is a function of the BI time, which is also estimated through forecasts. The capacity planners use an Excel™ spreadsheet to determine the number of BIOs, BLUs and BIBs required to meet demand on a weekly basis. To ensure that ATM always has sufficient production capacity to process the output from the Fabs, capacity planning buys enough tools to provide production capacity that is 25% above forecasted demand.

Additional processing capacity is added to account for assembly and test process variation based on a policy called Gap. The Gap policy specifies the amount of additional excess capacity required at a constraint, near constraint and non-constraint operations. For PBIC, which is planned as the constraint, the Gap is 10%. BI is considered a near constraint and should be allocated a Gap of 15%. However in capacity planning BI is treated as two independent operations. The BIO is allocated a 10% Gap and the BLU is allocated a 20% Gap. The BIB inventory is planned by adding 25% excess to account for demand variation but does not include any excess to account for process variation. Figure 1.4 is a depiction of the allocation of excess capacity among the Test operations.

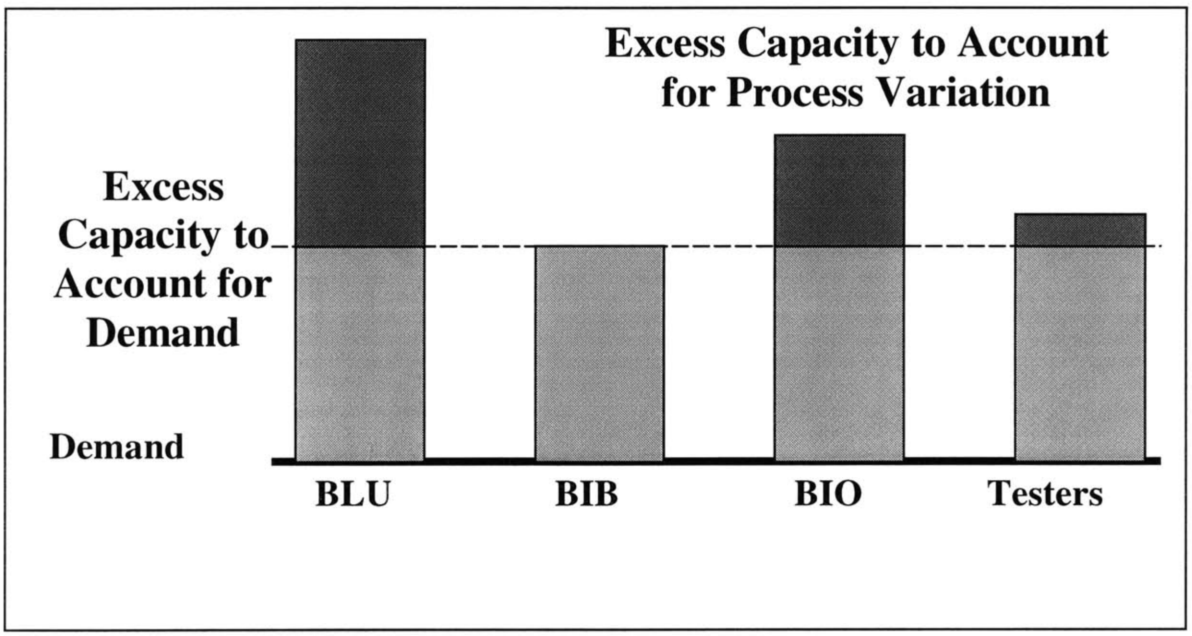


Figure 1.4 – Planned Excess Capacity

1.4 The Problem – Burn-in is a Constraint

At the outset of this project there was significant anecdotal evidence based on interviews with industrial engineers and process engineers that, while BI was planned as a near constraint, the operation was in fact the bottleneck for the factory. There was also a cost element associated with the BIBs that was attracting increased attention. BIBs, as a tooling item, are treated as a factory level expense, as opposed to BLUs and BIOs, which are treated separately as capital expenditures. Based on capital expenditures BI is a less expensive operation than PBIC, but if the cost of the BIB inventory, which can exceed the cost of all BI area tools, is considered BI surpasses PBIC in cost. Two questions flow from this circumstance. First, is there a way to reduce the BIB inventory to help reduce the cost of BI? Second, if BIB costs cannot be reduced should BI be planned for and managed as the constraint rather than PBIC? This project will address the first question, while an answer to the second question is out of scope. The scope of this project also addressed the applicability of any recommendations should Intel choose to treat BI as the constraint rather than test.

An investigation of BI revealed substantial variation in the processing time stemming from manual troubleshooting and time that lots spend in queues. The manual troubleshooting was performed to isolate faults in the BIO and the BIBs. During troubleshooting the entire lot waits and the BIO remains idle while the faults are isolated and repaired. Queue time variation differed from factory to factory, showing that differences in execution could be contributing to the variation.

The high degree of variation explained why BI constrained the factory despite the existence of a significant excess in production capacity. In the spreadsheet-based capacity model, the set-up time for the BIO, the run rate for the BLU, and the cycle time for the BIBs were fixed values. In reality these parameters were highly variable. There was a high probability that that BI would be a capacity constraint because the fixed values used in some instances accounted for as little as 50% of the occurrences found in the data analyzed. Changing the fixed values used in the capacity model so that they provided, say a 95% confidence that capacity would be sufficient was not prudent because the variation was so extreme that in the case of BIBs a nearly 300% increase in inventory would have been necessary. This project therefore focused on quantifying the variation, identifying sources of variation and determining the potential benefit of reducing variation.

1.5 The Project

I began this project with a period of extensive quantitative and qualitative data gathering. Two data sources captured the timing of the manual and automated operations that comprise BI. Extensive interviews characterized the processes and decision criteria used on the factory floor. The data gathering effort required traveling to the high volume manufacturing sites. In conjunction with my Intel supervisor, I chose the Costa Rica factory as the focus of this study because it was closer, geographically and temporally, than the Asian factories. With the variation characterized, I built a discrete event simulation based on a four-week period of operation at the Costa Rica factory for the FCPGA package. Historic data allowed me to validate the simulation and establish confidence that relevant dynamics were modeled. The simulation provided insight, through a series of experiments, into the impact of reducing processing time variation for the BLUs and the BIOs. In the course of this study, I observed that there were benefits to treating BI as a system of dependent processes rather than independent processes and that the simulation offered insight into the system's dynamics. The potential therefore existed to use the simulation as a supplement to the static capacity planning process.

Chapter 2: Background

2.1 The Impact of Variation on Capacity

Factory Physics Law 11 (Hopp and Spearman, 1996):

> *If you cannot pay for variability reduction, you will pay in one or more of the following ways:*
>
> 1. *Long cycle times and high WIP levels.*
> 2. *Wasted Capacity (low utilization of resources).*
> 3. *Lost throughput.*

This law alludes to the interchangeability of variation and capacity. In choosing to reduce variation rather than increase capacity in order to achieve the desired throughput, one must consider the manufacturing strategy. For ATM in planning capacity for BI, the practice has been to add excess tool capacity to ensure that throughput goals are met. This was a reasonable policy because BI capacity was historically inexpensive and variation reduction is by its nature relatively difficult, requiring concerted management attention. However, as the product mix has increased driving the need for larger BIB inventories, and tool complexity has driven up the cost of equipment, cost has become more of an issue.

Variation reduction may now be worth considering because the trend in BI related costs may have made the managerial difficulty of implementing variation reduction more appealing than simply buying more capacity. There are additional benefits of variation reduction. Because variation reduction is difficult, when it is done well, it increases a company's manufacturing competitiveness. This clearly aligns with Intel's goal of manufacturing excellence.

To achieve variation reduction requires a thorough understanding of the process in question. Pursuing variation reduction initiatives forces managers and operators to consider manufacturing processes from a different perspective and ultimately increases their understanding of these processes. Additionally, the learning from one variation reduction initiative can be applied to subsequent initiatives.

For this project, initial interviews revealed symptoms that pointed toward this idea of the tradeoff between variation and capacity. Among these symptoms were that BI had constrained the factory despite significant excess capacity. BIB inventories were an issue because they are expensive and because operators complained that the inventories were inadequate. Yet inadequate BIB inventories could imply a few things, among them that there is a lot of WIP in the system. BIB shortage could also indicate that the tooling item is very unreliable so effective inventory is much less than the reported inventory. In reality both of these factors contributed to the observed problem. Translating this anecdotal evidence into quantifiable data was an important first step in understanding the nature of variation and capacity balance for BI.

2.2 Quantifying the Degree to which Burn-in is a Constraint

Quantifying the assertion that the BI area had constrained a specific factory in the past was difficult because of the lack of data available and because of the installed excess capacity. To understand how often BI constrained the factory one would look for evidence that showed a buildup of WIP waiting to be processed in BI while at the same time the PBIC area was starved for WIP. Such data was not explicitly available. However, queue times were available and could be used to infer WIP levels.

Figure 2.1 shows the distribution of queue times for a typical week in one factory for all the Assembly/Test operations. This graphic shows that lots waited the longest to be processed through BI load, the first BI operation and that the lots had a relatively short wait prior to PBIC. However, this in not the whole story. Figure 2.1 also shows that a number of tools were bagged during this week. Newlin (Newlin, 2000) discusses the bagging policy in detail.

For the purpose of this discussion, bagging involves removing tool from use to avoid a negative impact on utilization measures. Because significant excess capacity is installed in ATM factories, one would expect to see low utilizations during normal operations. To avoid this, tools are shut down or bagged, effectively reducing capacity and increasing the utilization of tools in use. The significance of this policy in understanding if BI is a constraint is based on the following rationale. During the week shown in Figure 2.1 PBIC was able to bag one tester and two operations in assembly were also able to bag tools. This indicates that the demand for that week was not significantly above forecast and thus did not require the excess capacity built into the process. Based on interview data, the BI area very rarely bags tools. This means that during this week in which demand was not significantly above forecast, the BI area was using all of its excess capacity and still causing delays in the entering queue.

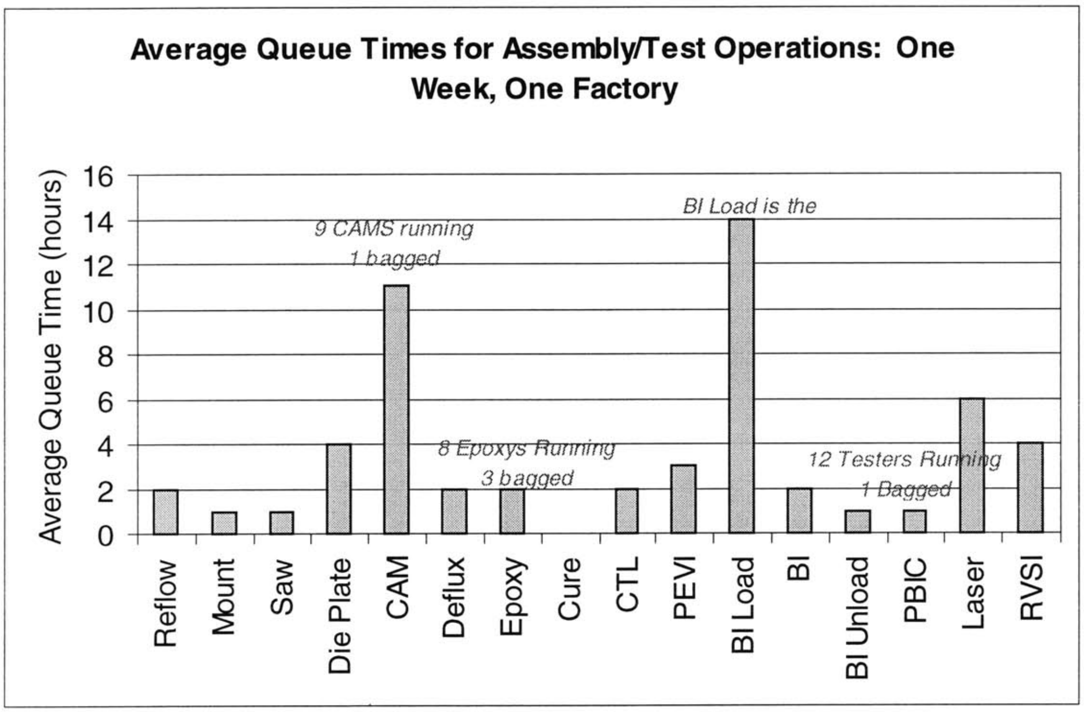


Figure 2.1 – Average Queue Times for ATM Operations

2.3 Detailed Capacity Planning Description

At the factory level, capacity planning is the responsibility of the Industrial Engineering group. Formal agreements on data to be used in the capacity planning process are reached during a weekly meeting called Resource Requirement Planning (RRP). Demand data for the entire product mix is presented by the Site Planning group, which uses a forecast, generated at the corporate level, called the Build Plan. The Build Plan has a nine-month horizon and is updated weekly. There tends to be substantial variation from forecast to forecast, so at the RRP, a buffer is added to the Build Plan demand numbers to ensure that Assembly Test Manufacturing has sufficient capacity to process all material generated by the Fabs. This revised demand forecast is called the Capacity Roadmap.

Additional data relevant to capacity planning are the Product Heath Indicators (PHIs). This forecast impacts the run rate and yield, and therefore the required capacity for BI. Specific parameters relevant to BI capacity planning are the Burn-in time (the amount of time the product must be subject to high temperature in the oven), the Re-burn-in rate (some units that fail later stage testing require a second pass through BI) and the K-lot percentage (K-lot is discussed later in Section 2.6.5).

PHIs are derived from specific product technology reliability testing. In general, PHIs vary by product and by the product maturity. When a product is first introduced the PHIs are chosen conservatively to ensure that unreliable product does not reach the market. Conservative PHIs imply long run rates and low yields. As the product matures and reliability data is gathered, PHIs tend to become more optimistic and thus more favorable from a capacity planning perspective. However, occasionally unexpected product reliability data mandates that PHIs be changed back to more conservative levels. When this happens, it is usually contrary to the forecast and results in the factory having insufficient capacity.

A centralized quality and reliability (Q&R) group provides the PHIs that are used by the site Q&R group to determine the Burn-in time, Re-burn-in rate, and the K-lot percentage. A locally resident analytical tool called the Baseline Reliability Analysis Tool is used to determine the values of these parameters. The local Q&R group has some flexibility in setting, for example, the Burn-in time. If the central Q&R group sends data to the local Q&R that indicates a conservative change in Burn-in time from a previous forecast, the local Q&R group can cushion the impact of this change on capacity by varying other Burn-in parameters such as the Burn-in temperature or by modifying the thermal characteristics of the Burn-in oven.

The Capacity Planner in the Industrial Engineering group uses the demand and PHI data gathered in the RRP to run a spreadsheet based capacity model. This model is locally maintained, but consistency in values used for run rates and utilizations is ensured through joint engineering teams that exist as part of a forum called the Virtual Factory. The Virtual Factory concept facilitates the necessary information flow to ensure that Intel's geographically dispersed manufacturing facilities operate as essentially one factory.

2.4 Variables in the Capacity Model

The capacity model determines the required number of BLUs, BIOs and BIBs based on the following relationships:

$$\text{Number of BLUs} = \frac{\left(\frac{D}{Y}\right)(1+r+k)}{(uph)MU}$$

$$\text{Number of BIOs} = \frac{\left(\frac{D}{Y}\right)(1+r+k)(BIhrs+BIset)}{(u_b * b_o)MU}$$

$$\text{Number of BIBs} = \frac{\left(\frac{D}{Y}\right)(1+r+k)(BIhrs+BCT)}{(u_b * b_o)MU}$$

Where D is demand based on the Capacity Roadmap, Y is yield, r is Re-burn-in rate, k is K-lot percentage, uph is the run rate in units per hour, MU is the machine utilization, u_b is units per BIB, b_o is BIBs per oven, $BIhrs$ is Burn-in hours, $BIset$ is Burn-in setup time or the time that a BIO spends occupied with a lot exclusive of the Burn-in hours, and BCT is BIB cycle time or the time from when a lot is loaded on to BIBs to the time the lot is unloaded from BIBs exclusive of the Burn-in hours. It should be noted that for small values of r and k (on the order of .1) these formulas are reasonable approximations. A more accurate formulation would replace the existing parenthetic term in the numerator accounting for r and k with the following term in the denominator: $(1-r-k)$.

For the purpose of this study I have categorized these parameters as external and internal variables based on their relationship to BI as a process. Demand, yield, Burn-in hours, Re-burn-in rate and K-lot percentage are external variables and *uph*, utilization, BI setup time and BIB cycle time are internal variable. The units per BIB and BIBs per oven variables were a function of the technical characteristics of the BIB and the BIO design and were thus considered external. In practice rather than using the quantity $u_b * b_o$, the lot size was used. This categorization was useful because it offered a means of bounding the project based on data availability, and the sphere of influence of the central ATM Industrial Engineering group.

Understanding this sphere of influence was important from the standpoint of obtaining access to the right people and data, and from the standpoint of implementing the recommendations of this study. With a firm understanding of the impact of both internal and external variables, the question became which variables could be substantively changed by the central ATM Industrial Engineering group. As an example, demand forecasts were highly variable and reducing this variation would have been beneficial from a capacity planning perspective, but a more detailed study of this phenomenon and offering suggestions that could be acted upon would have required this project to be based in one of Intel's central planning or marketing groups. A similar line of reason applied to the PHIs, which were driven by the product engineering side of the business. I therefore chose to focus on the internal variables, which offered a depth of academic material and fell within the purview of the Industrial Engineering function in ATM.

In examining the internal variables, MU is a good place to start. MU is a derived value based on the Gap policy and tool availability. Gap is an excess capacity policy developed in the Fab environment and based on the Theory of Constraints and intended to provide a buffer against lost capacity due to internal variation. The Gap policy used during the time of this study was called 10/15/20, which meant that 10 percent extra capacity was planned for constraint operations, 15 percent extra capacity for near-constraint operations and 20 percent for non-constraint operations. The Gap percentage equals availability minus utilization. The capacity planners used an availability value for each tool that was sometimes measured and sometimes derived, and subtracted the appropriate Gap percentage to determine the utilization to enter into the capacity model. Newlin (Newlin, 2000) discusses the Gap policy and its limitations in some depth as well as Intel's paucity of availability data for Assembly/Test tools. BI was considered a near-constraint for capacity planning purposes, but the Gap policy was curiously applied as 10 percent extra capacity for the BIOs and 20 percent extra capacity for the BLUs. Gap was not applied to the BIB inventory, though a utilization value was derived from BIB reliability information.

The remaining three internal variables were the primary focus of this study. The *uph* for the BLUs was measured directly from tools on the factory floor. Industrial engineers and process engineers in different factories at different times and with varying degrees of statistical rigor conducted time studies to establish the units per hour processed by a typical BLU. The results of these studies were presented to cross-functional teams representing all the factories that comprised the Virtual Factory. In this forum a fixed value for the BLU *uph* was considered based on the data presented, which became the validated number to be used by all of the capacity planners across the Virtual Factory.

The Burn-in setup time was measured to determine the run rate for the BIO. This setup time included all of the time the lot spent in the oven from load to unload, excluding the Burn-in hours. Like the BLU *uph*, the Burn-in setup time was determined through time studies and validated through the Virtual Factory.

The BIB cycle time represented the equivalent of tool run rate for the BIBs. BIB cycle time was the amount of time a unit spent on a BIB excluding the Burn-in hours. This parameter was comprised of processing times for the BLU load and unload sequences, queue time waiting to be loaded into the BIO, queue time waiting to be unloaded in the BLU, time to conduct the manual package short test (Vcc-to-Ground Check) and time devoted to loading and unloading the BIO, conducting the pre-signal checked, troubleshooting and ramping the temperature up and down in the oven. The value used for the BIB cycle time was validated through the Virtual Factory and was a primary focus of ongoing cycle time reduction efforts.

2.5 How well does a static capacity model represent a dynamic process?

A key question arises if one considers the primary motivation behind this study, which is to determine why Burn-in has constrained the factory despite the excess capacity designed to account for external and internal variation. The questions is, in using fixed values for the internal parameters tied to equipment run rates, how well do these values represent the performance of the actual Burn-in operation? The hypothesis of this work is that the BI consists of highly variable process times, which make a static capacity model an overly simplistic representation of the real system. To understand this hypothesized schism between the capacity planning process and the dynamics of the factory floor requires a more thorough treatment of the Burn-in process.

2.6 Details of the Burn-in Process

2.6.1 General

The general layout of the Burn-in area consists of a WIP area for lots arriving from the front of the line assembly operations, a common area in which the BLUs are located and a separate area where the BIOs are located. The BIO area is separate because it requires a less stringent clean room that BLUs and the other Assembly/Test operations.

Lots arrive at BI on a single cart with a set of plastic trays which each hold ten units. The standard lot size is 1000 units, though the lot size can be smaller for a variety of reasons including engineering test lots or K-lots. A color-coded (based on product type) plastic envelope containing documentation on the lot's processing history is attached to the cart. The arrival rate varies substantially. Interviews suggested a weekly periodicity as the front of the line operations push to meet end of week output requirements, but data did not substantiate this assertion. Figure 2.2 shows the arrival pattern for at one factory for four weeks.

Processing time and queue time data is captured in a factory floor tracking system called Workstream into which operators enter time stamps for various events. The first step for a lot as it enters BI is for the operator to enter the time into Workstream. This timestamp denotes the beginning of the BLU load process.

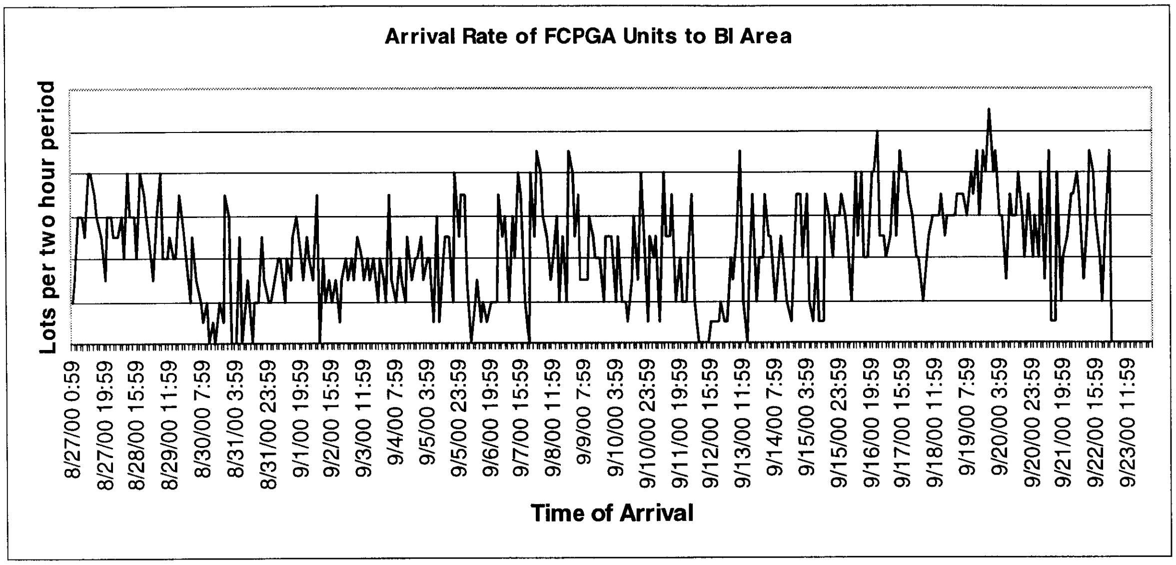


Figure 2.2 – Arrival Rate for FCPGA units to BI

2.6.2 The Burn-in Load Process

BLUs are organized in sets of two with a single operator working two BLUs. The load or unload operation is easily monitored because the manipulating arm and BIB are visible through glass sides on the machine. A BLU is about five feet high and there is a red/yellow/green status light mounted on a pole about three feet above the top of the BLU so it can be seen from anywhere in the BLU area. The operator splits the lot between the two BLUs and pushes two trolleys loaded with empty BIBs into position, one trolley for each BLU. BIBs are heavy and bulky so a single lot must be split between two trolleys to minimize the risk of over-exertion in manipulating the trolley. During the load sequence the BLU pulls a BIB from the trolley, inserts the units into sockets on the BIB, and replaces the BIB in the trolley.

A BIB is a printed circuit board approximately 15 inches by 30 inches. Mounted on the BIB is a matrix of sockets, which are the interface between the unit under test and the printed circuit board. The number of sockets that can fit on a BIB is a function of the size and thermal characteristics of the of the product. There can be as few as six sockets or as many as fifteen. The sockets are fairly sophisticated mechanical system that must be opened by the BLU before the unit can be inserted. The sockets are made by a different supplier than the vendor who makes the BIB and constitute approximately half of the cost of the BIB.

There is a common type of BLU used for all products across the Virtual Factory, but the BLU requires a product specific 'kit' that includes hardware that interfaces with the product as it is loaded into the BIB and a vision system to guide the mechanical arm. There is also software included in this kit that must be installed in the BLU. The changeover time for different products is on the order of ten hours so BLUs tend to be dedicated to specific product types.

With the BLU load sequence complete, the Vcc-to-Ground check is performed. This procedure varies from site to site. One factory assigns the BLU operator the responsibility of conducting the Vcc-to-Ground check while the trolley remains in the proximity of the BLUs used in the load sequence. Another factory assigns one or two BIO operators the unique role of conducting the Vcc-to-Ground check in an area adjacent to the BLU area. In both cases the timestamp used to indicate completion of BLU load follows the Vcc-to-Ground check.

The BLU operator makes the decision whether to process a lot entering BI or a lot preparing to leave BI. Operator performance was assessed based on output from their respective pair of BLUs. The standard practice that has evolved is for the operator to chose the opposite of the previous activity. If a set of BLUs has just completed a load sequence the operator will chose a lot that is ready to be unloaded. This logic, which I have termed "toggle logic", is based on local expedience. If an operator has just completed an unload sequence there is are two trolleys of empty BIBs in position to be immediately loaded, all that is required is for a new incoming lot to be loaded into the BLU. If the operator chose instead to unload another lot, she would need to push the two trolleys of empty BIBs to a staging area and then push two more trolleys of a lot to be unloaded into position at the BLUs. Then the operator would need to find a cart with empty plastic trays to which the lot could be unloaded.

One factory had a variation on the toggle logic that involved a position called the Material Handler. This person reported to the BI area shift manager and was responsible for directing the BLU operators in which lots to load or unload. The Material Handler's decisions on lot priority were tied to the downstream WIP area in front of the PBIC. PBIC was the planned constraint for the factory so the Material Handler's job was to ensure that this queue was never empty.

The factories had experimented with different lot prioritization schemes in the past. One notable example, notable for its poor results was to tie the BLU operators load and unload decision to the shift timetable. In this scheme, at the beginning of the shift BLU operators would choose lots to load. As the shift wore on, an increasing number of unload lots would be processed until by mid-shift there was an even mix of load and unload lots being processed. By the end of the shift, all lots being processed by the BLUs were unload lots. This scheme proved difficult to manage in practice with WIP bubbles growing and dispersing leaving tools fully loaded or starved at an unpredictable periodicity. Accounts from interviews indicated that at times a variant of this prioritization scheme occurred. Because BLU operators' performance were assessed based on local output for their individual shift, it was in the operators best interest to focus on unload lots toward the end of the shift if their output for the day appeared likely to be short of the goal. An unforeseen maintenance issue could cause such a shortcoming in output for a particular shift. The result of this response was that the next shift might be left with very little WIP in the BI area and respond by focusing on load lots. A cycle then begins in which WIP levels rise and fall unpredictably and with increasing amplitude.

Pairing BLUs to process a single lot has interesting ramification in the event that a BLU breaks. If during the course of a load or unload sequence, a BLU ceases operation and requires maintenance, one half of a lot is obviously delayed in processing. Since the BLUs are paired, however, once the other half of the lot is processed it must wait until the first BLU is repaired and the lot completely processed before it can be moved to the next operation. Additionally, the BLU that was functioning properly must sit idle until the broken BLU is repaired. This practice is tied to the necessity of maintaining lot integrity so that unit's can be traced through the entire manufacturing process.

2.6.3 The Vcc-to-Ground Check

The Vcc-to-Ground check is designed to isolate short circuits in the unit package. The check begins with the operator checking for indications of a short circuit on each BIB using a hand-held multimeter. If no BIBs show indications of shorts the Vcc-to-Ground check is complete and the operator pushes the trolley to the WIP area for the BIOs. If one or numerous BIBs show indications of short circuits, the BIBs are removed individually and inserted into a diagnostic bench called an X-Y table. Using the multimeter and the X-Y table, the operator isolates the specific unit that has the fault and removes it from the lot. The operator documents the number of faulty products discovered during the Vcc-to-Ground check. The obvious questions arises, why do such a manual step when the lot will be processed in the BIO and the shorts identified at that point? In fact the Vcc-to-Ground check has been proven to save total BI processing time by identifying package faults that would prolong the troubleshooting process performed prior to running the BIO.

2.6.4 The Burn-in Oven Process

When the lot has completed the BLU load sequence and Vcc-to-Ground check is satisfactory, the BLU operator or Vcc-to-Ground check operator enters the time in Workstream and pushes the lot to a WIP area to wait for loading into a BIO. A different group of operators are assigned to the BIOs with a manning ratio (number of ovens for which the operator is responsible) being a function of the BI time. For products with very long BI times one operator can manage as many as ten ovens. For shorter BI times, one operator will be assigned as few as three ovens.

BIO status is not as easily monitored as the BLU status. A BIO stands about eight feet tall with an extensive ventilation system extending from the top of the oven to ductwork in the ceiling. The only means of observing the status is through a display console that faces directly forward with respect to the oven. Because the BIOs are so large the positioning is designed to optimize the use of space. BIOs are arranged in rows and oriented as in Figure 2.3.

An operator is often responsible for ovens in different rows and can only observe the oven status by standing directly in front of the tool. A few factories have implemented a system called the oven monitoring system (OMS) that includes a large monitor screen located at the periphery of the BIO area. This system is directly tied to an automated database that records BIO status. This same database also captures status information on the BLUs.

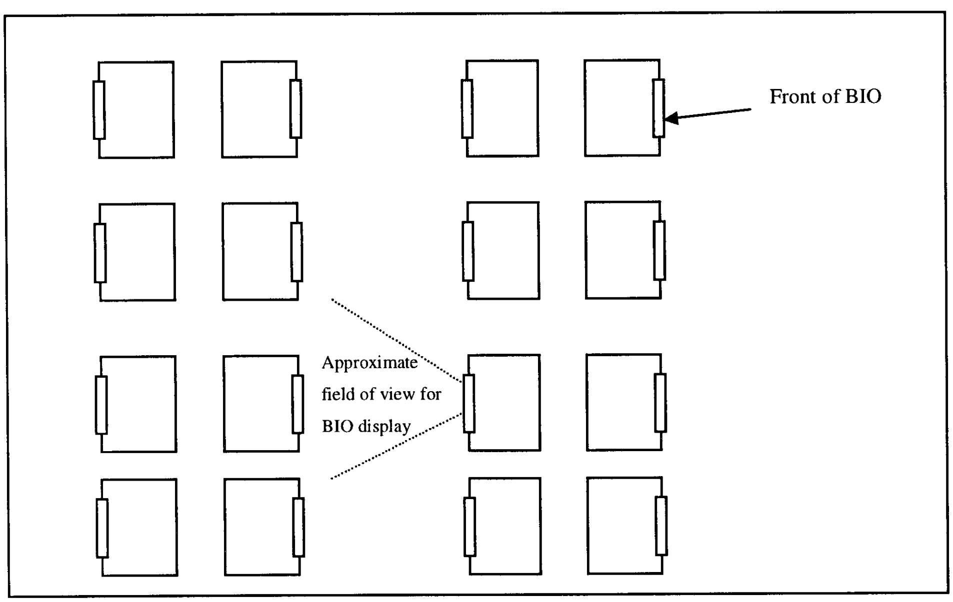


Figure 2.3 – Orientation of BIOs in Factory

The BIO operation begins with an operator pushing a lot from the WIP area to the BIO. The operator inserts the BIBs one at a time into the BIO. The BIBs slide into vertically oriented slots in the BIO. Once all of the BIBs are inserted, the operator uses a custommade hand tool about the size of a small hammer to lever the BIB into seating properly into the back of the BIO. The oven doors are then closed and the operator runs a software-driven diagnostic called the pre-signal check. The purpose of the pre-signal check is to identify faults in the BIBs or the BIO. When the pre-signal check is complete, the system provides a map of the BIO and identifies slots that have failed the check with a color-coding. If there are no indications of a fault, the operator enters the approval for the BIO to continue with the BI sequence. At that point a time stamp is entered into Workstream indicating that BI has begun. It is important to note that from Workstream's perspective the load, diagnostic and troubleshooting operations are considered part of queue time prior to the BIO.

If the pre-signal check reveals faults, the operator then opens the oven and follows a documented troubleshooting logic tree designed to isolate the fault to the BIB or the BIO hardware. Once the operator has isolated the fault, she either repairs it, or contacts a manufacturing process engineer to assist in the repair. When all of the faults have been rectified, the operator re-runs the pre-signal check. If all of the faults were fixed, the pre-signal check should return indications that no faults are present, but occasionally, the second (or perhaps the third or forth) run of the pre-signal check reveals a faulty slot that was good on the first pre-signal check. Comprehensive data is not gathered on the delays attributable to this troubleshooting process so root causes of faults cannot be tied to delays in process time. Anecdotal evidence from interviews with shift supervisors, BIO operators, and process engineers indicated that fuses on the BIBs failed frequently which required replacement by the on-site BIB vendor technician. The frequency of fuse failure seemed directly related to the increased power usage of newer generation products.

When the pre-signal check has been satisfactorily completed, the BIO is cleared to ramp up to the Burn-in temperature. The lot is burned in for the specified Burn-in time and then the BIO temp is ramped down to room temperature. The operator then unloads the lot back to a pair of trolleys, enters the completion time into Workstream, and the lot is pushed to a WIP area to await BLU unload. It is possible for there to be an interrupt in the BI sequence that requires the operator to take action before it can be resumed. The only indication of an interrupt is a text message on the console. An operator can be occupied troubleshooting a BIO for a prolonged period of time during which she is unaware that action is required on an interrupt or that a lot is ready to be unloaded on a different BIO.

2.6.5 K-lots and the Burn-in Unload Process

The purpose of BI is to identify product that is inclined to fail during the infant mortality stage of the lifecycle. If such is product is identified in the BI sequence it is tagged in a database so that the BLU can identify and separate the failed unit from the lot. Such units are segregated into a category called K-lot and sent to the Semi-finished goods inventory (SFGI). In SFGI K-lot units are accumulated until a lot large enough to be sent through BI again is assembled. In addition to failed units, any units that for some reason did not complete BI the first time are designated K-lot. An example of such a case is units that were involved in a very lengthy pre-signal troubleshooting process and the operator chose to run the lot rather than repair all the faults. It is possible for an entire lot to be designated as K-lot if more than a specified percentage of the lot fails BI. The throughput time for a K-lot is based on the time when the first units that constitute the lot are received at SFGI. If it takes a month to accumulate enough units to complete a K-lot, all of the units appear to have been delayed by a month. K-lot units that fail are disposed of, while units passing are sent on to PBIC.

The K-lot process is perceived as wasteful and BI area shift supervisors are urged to keep the K-lot percentage as low as possible. This incentive translates into thorough troubleshooting to ensure that any units that are not bona fide failures in BI do not get sent to K-lot. Another incentive issue tied to K-lots is that they are not counted as output for the BLU operators which means that, depending on the circumstances, operators may chose not to load a K-lot when a first-pass lot is available in the queue.

When the lot has been unloaded from the BIO, the operator moves the lot to the WIP area for BLU unload. The unload sequence in the BLU is virtually identical to the load sequence, though data suggests that the unload operation requires slightly less time than the load operation for same sized lots. When the unload is complete, the BLU operator pushes the lot to the WIP area to await processing in PBIC.

Chapter 3: Data and Variation

3.1 Data Sources

When comparing the detailed description of the BI area to the capacity planning process two inconsistencies are noteworthy. First, the capacity planning process treats the BLU load/unload sequences and the BIO sequence as independent processes when in reality, because of the reentrant nature of the process, there is a dependent relation. The BIOs rely on being fed WIP from the BLUs, and produce WIP to be unloaded by the same BLUs. BI is thus a coupled system consisting of two machines and three steps with two internal queues. An assertion that will be addressed later in this paper is the merits of treating BI as coupled system in the capacity planning process as a means of rationalizing the existing scheme that treats the process as two independent steps. Second, there is a high level of manual operation and troubleshooting that introduces variation into the BI process that is not explicitly accounted for in the static method of capacity planning. While the Gap policy is an effort to buffer against internal process variation, the specific excess capacity values are based Fab experience not ATM experience, let alone the specific nature of variation in the BI area.

To address this second inconsistency is was necessary to quantify the variation in processing and queue time for the BI area. There were four sources of data available, the Workstream system, the Sybase system, time studies preformed by industrial and process engineers, and personal interviews. Newlin (Newlin, 2000) comments on the accuracy of Workstream in capturing process times and output. To gain the most accurate picture of the variation in BI it was necessary to use these data sources together as a means of reconciling inconsistencies in the individual sources.

Workstream was the primary source of data used in quantifying BI area process time and queue time variation. The boundaries between processes and queues therefore was a function of timestamps entered by operators. For example, the processing time for BLU load was the difference between the timestamp entered at the beginning of the load and the timestamp entered when the lot completed being loaded and the Vcc-to-Ground check was complete. The queue time waiting to be loaded into the BIO was the difference between the timestamp entered when the lot was complete with BLU load and the timestamp that was entered to begin Burn-in following the load and pre-signal troubleshooting. Figure 3.1 shows the data coverage provided by the Workstream and Sybase systems mapped to a timeline of Burn-in events.

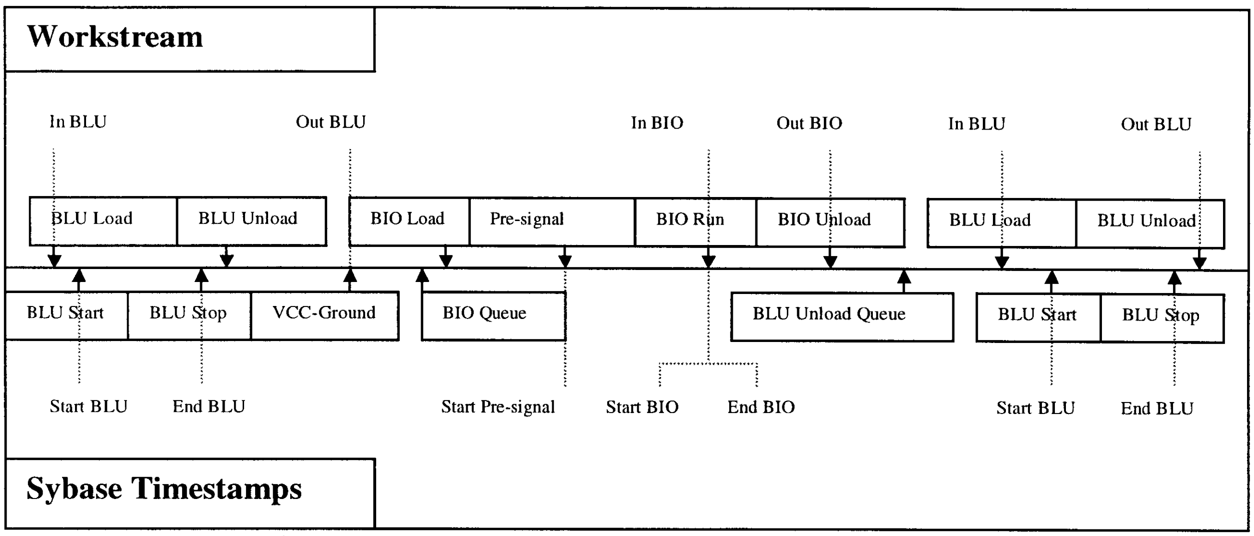


Figure 3.1 – Event Coverage by Workstream and Sybase

In addition to isolating the data according to process step, the data was categorized based on the factory of origin. To minimize the impact of variation rooted in the product, a single package type was studied. This package was called the Flip Chip Pin Grid Array and was used to package the highest volume product for the factories being considered during the time frame under study. The factories considered were the four primary ATM facilities used for microprocessor production located in Costa Rica, Malaysia and the Philippines. The time period chosen was four weeks during late August and early September 2000. This timeframe was a compromise between the need to have enough data to capture weekly cycles and need to keep database query times manageable.

3.2 Comparison Between Process Time Variation and Fixed Values used in the Capacity Model

Appendix A is a graphical display of the process-to-process and factory-to-factory comparison of process and queue time variation. The histograms show a high degree of variation at each step of the BI process. In addition to coefficients of variation generally between around 1 and at times much greater than 1, the distributions are positively skewed. These coefficients of variation are indications of high-variability processes (Hopp and Spearman, 1996).

The high degree of variation observed highlights the schism between the assumptions underlying the capacity planning model and the reality of the dynamics of BI. Given the excess capacity added to account for variation, however, the questions is what level of confidence should the capacity planners have that there is enough capacity to meet demand? Figure 3.2 and Figure 3.3 answer this question by comparing the fixed values used in the capacity model to the cumulative probability function developed from the factory floor data.

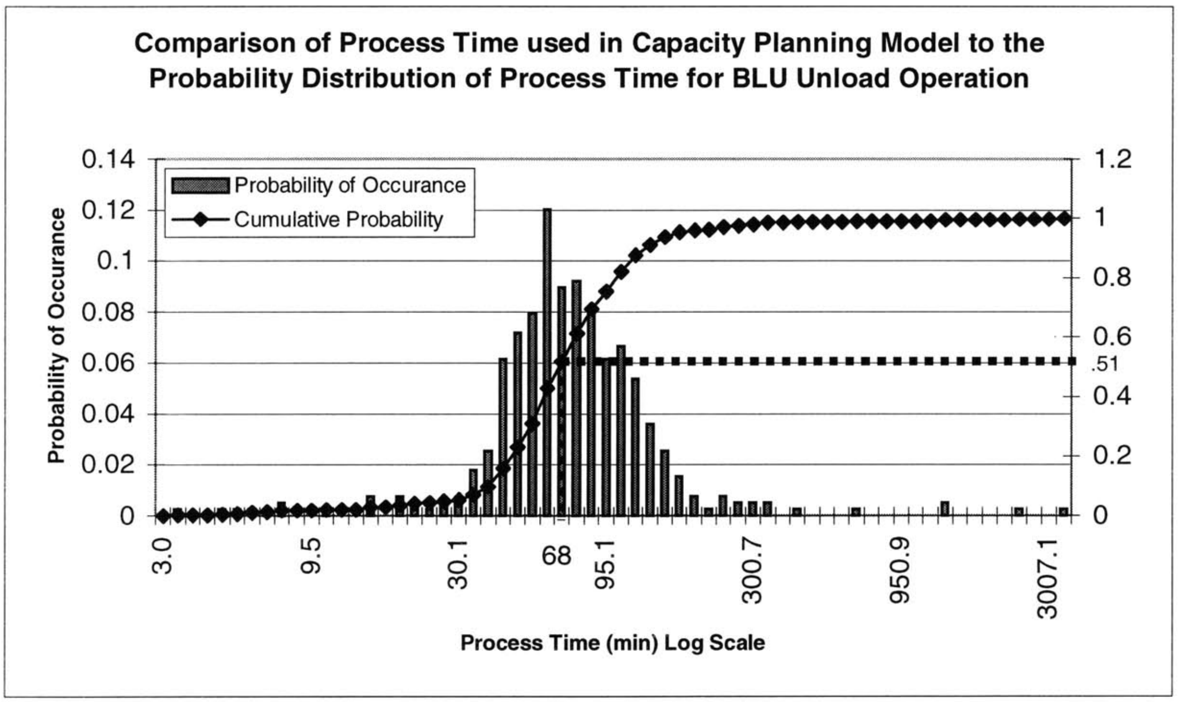


Figure 3.2 – Comparison of BLU Unload Time Distribution to Capacity Model

The BLU unload process was used to do this comparison because the BLU load data included time spent conducting the Vcc-to-Ground check. Based on Figure 3.2 there is a 51% probability that the BI area will have adequate BLU capacity to handle forecast demand. This is based on a processing time of 68 minutes which was the value used in the capacity model for the time period under study. The BLU process was a good place to start this comparison because its run rate is not dependent on the BI time of the product. In making the same comparison for the BIB inventory one must consider BI time, which varied during the time period considered, and which could not be explicitly extracted from the data. It is possible to get a sense for the range of probabilities that the BIB inventory would be sufficient by considering the range of BI times and plotting this range on the cumulative distribution for the BIB cycle. Figure 3.3 shows this comparison.

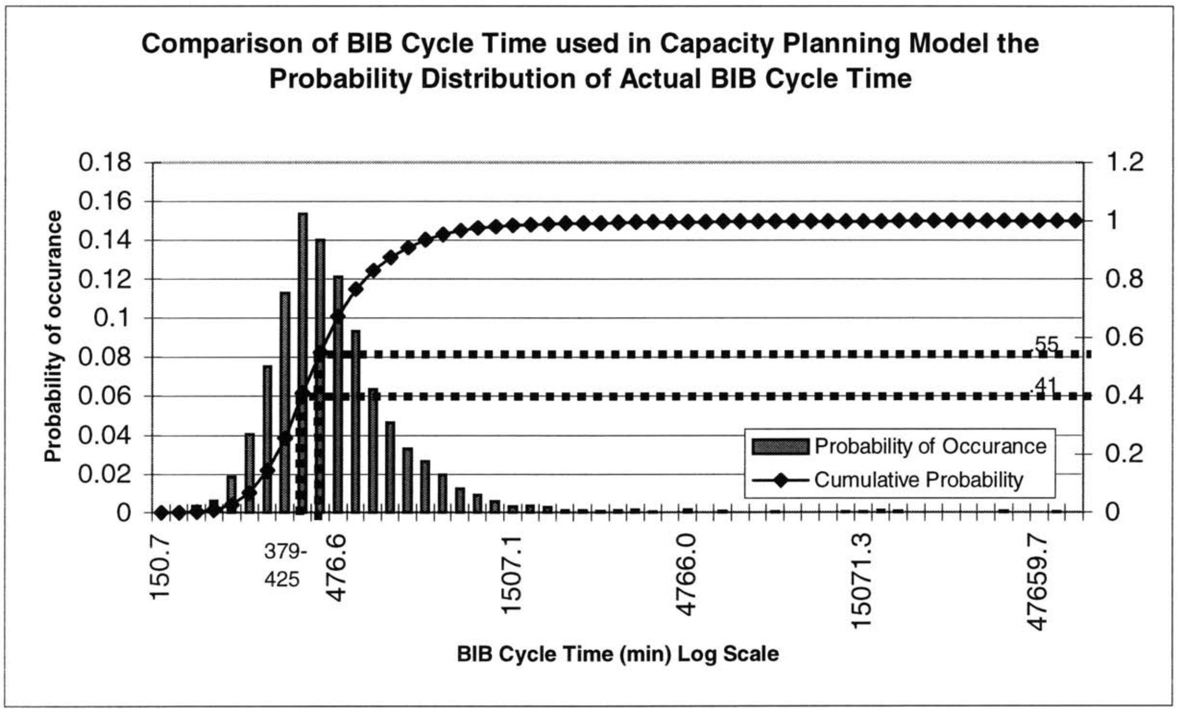


Figure 3.3 – Comparison of BIB Cycle Time Distribution to Capacity Model

The values for BIB Cycle Time used in the capacity model varied between 379 and 425 minutes. Mapping these values to the cumulative distribution shows that the capacity planners could be between 41% and 55% confident that the BIB inventory would be sufficient to service demand.

It was not possible to make this type of a comparison for the BIOs because the time associated with the load and pre-signal troubleshooting steps, which constitute a significant portion of the BIO setup time, is contained in data represented by BIO queue time.

3.3 Sources of Variation

The detailed process description suggests that the BI area includes a number of potential sources of variation that are based on manual operations. These potential sources included the Vcc-to-Ground check, the pre-signal troubleshooting, and the internal queues. The pre-signal troubleshooting time had a secondary effect that deserved study. Because a single operator was typically responsible for multiple ovens, it was quite possible for a lot to be delayed while waiting to be unloaded from the oven while the operator was troubleshooting a different BIO. These sources were chosen as points of investigation because they were a function of labor practices and decisions, and appeared to be significant in their impact on the operation. These sources were also present in the next generation of the BI process despite the introduction of expensive new tools.

Additionally, a political issue drove this choice in the investigation. While mean cycle time reduction for BI was of critical importance to the industrial and manufacturing process engineers assigned to the BI area, process time variation reduction was not considered a high priority. However, the implication of this study is that the focus on mean cycle time reduction without reducing variation, fails to address the problem that BI is the bottleneck when PBIC is planned as the bottleneck. This is because the rationale behind mean cycle time reduction efforts is cost avoidance.

For instance, BIB cycle time is a prominent target for mean cycle time reduction initiatives. Process engineers develop changes that shorten the mean cycle time and validate this new BIB cycle time through a joint engineering team. Industrial engineers then insert this new value into the capacity model, which results in lower requirement for future BIB inventories. By referring to Figure 3.3, one can see that mean cycle time reduction without addressing variation is the equivalent of shifting the distribution to the left, and reducing the process time value used in the capacity model is equivalent to moving the vertical dashed line to the left by a comparable amount. If the variance remain the constant, the intersection of the capacity model value and the cumulative distribution curve will remain the same, and the risk of running of out of capacity will be the same as before the mean cycle time reduction effort was undertaken. In reality one would anticipate some change in the variance as a result of changing the process on the factory floor. However, without explicitly studying variation, it cannot be known if such process changes increase or reduce variation. If a change were implemented that reduced mean cycle time, but increased the variation of the cycle time, the risk would increase that tool or BIB inventory capacity would be insufficient to meet demand.

The data supported the claim that some of the sources of variation identified were significant. However, it was not clear that the Vcc-to-Ground check had a significant impact on process time variation. Figure 3.4 presents a comparison between the BLU load and BLU unload processes. Since the Vcc-to-Ground check time was captured as

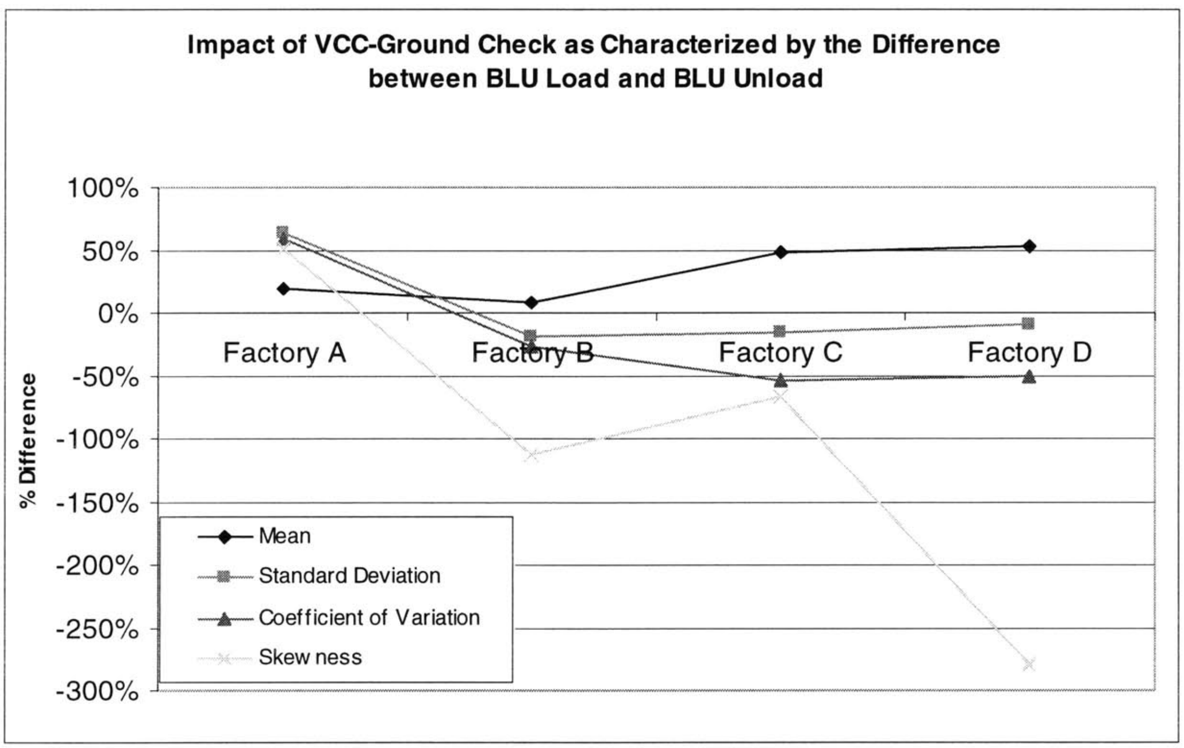


Figure 3.4 – Impact of Vcc-to-Ground Check on Variation

part of the BLU load sequence, and the BLU load and unload were very similar operations the impact of the Vcc-to-Ground check could be estimated by looking by the differences between the means, standard deviations, coefficients of variation, and measures of skewness for the two operations. In Figure 3.4 the BLU unload process time statistical measures were subtracted from the BLU load process time statistical measures. Positive values indicate that BLU load measures were greater than BLU unload measures and negative values indicate that BLU unload measures were greater than BLU load measures. While all of the factories show a positive difference between mean processing time, indicating BLU load took longer than unload, the same indication is not so clear when looking at the measures of variation. Factory A is the only factory in which BLU load sequence is more variable that the BLU unload sequence and this might be attributed to its unique method of conducting the Vcc-to-Ground check. In all of the other factories BLU unload is more variable than BLU load. While this comparison does not account for other differences between the two operations, it casts doubt on the assertion that the Vcc-to-Ground check has a relatively large impact on the process time variation for BLU load. This uncertainty about the relative impact of the Vcc-to-Ground check surfaced during the simulation development also, and will be addressed in Section 4.3 .

The Workstream data shows a high degree of variation attributable to the combination of queuing and troubleshooting associated with the pre-signal check. Figure 3.5 and Figure 3.6 compare the queue time prior to the BIO and the queue time prior to the BLU unload. It was not possible to differentiate queue time from troubleshooting time so Figure 3.5 includes both in its accounting of queue time. One factory's data was not was not credible and was left off this comparison. These figures also show that there are significant factory-to-factory differences in the queue time and the variation attributable to the queues. This data supports chartering a rigorous benchmarking study among the factories to catalog best-known methods for managing BI.

This type of comparison is not as useful in understanding variation associated with the BIO processing time because the BI time during the period of study varied between the two variants of the product. The data depicted in Appendix A for the BIO does show a

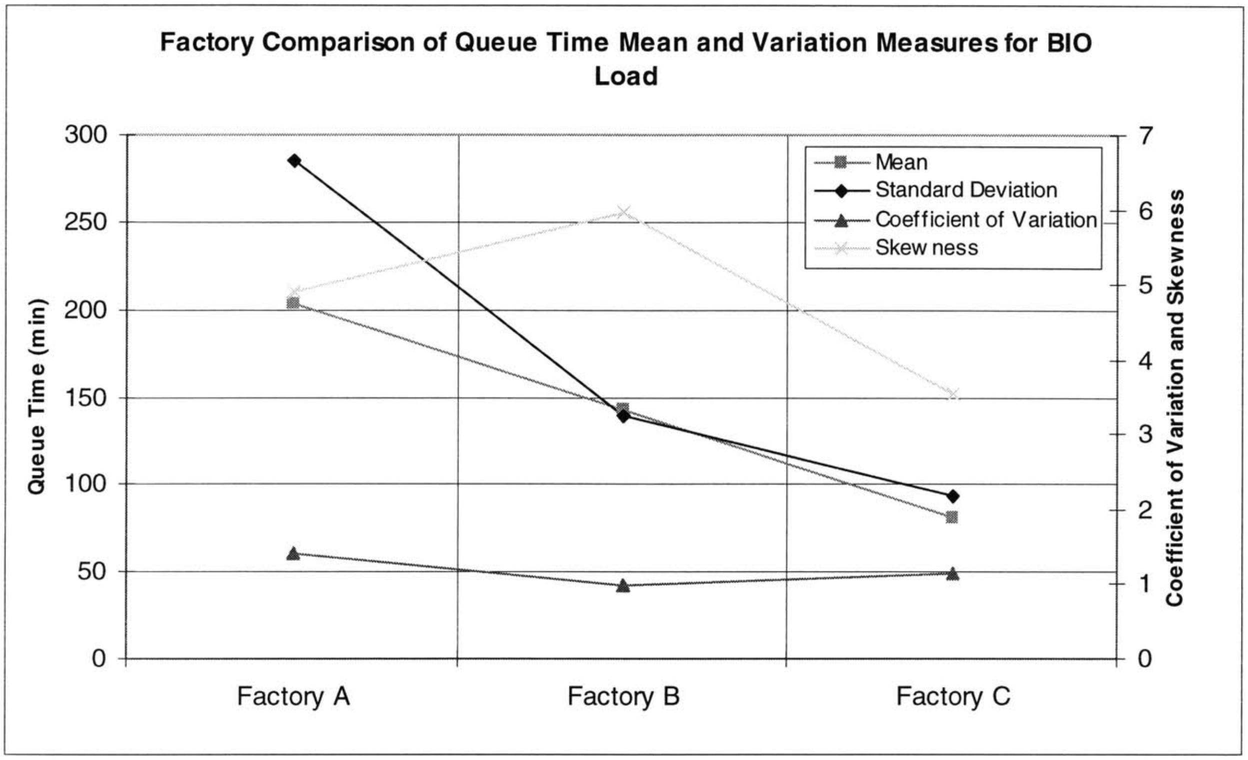


Figure 3.5 – BIO Load Queue Time Variation

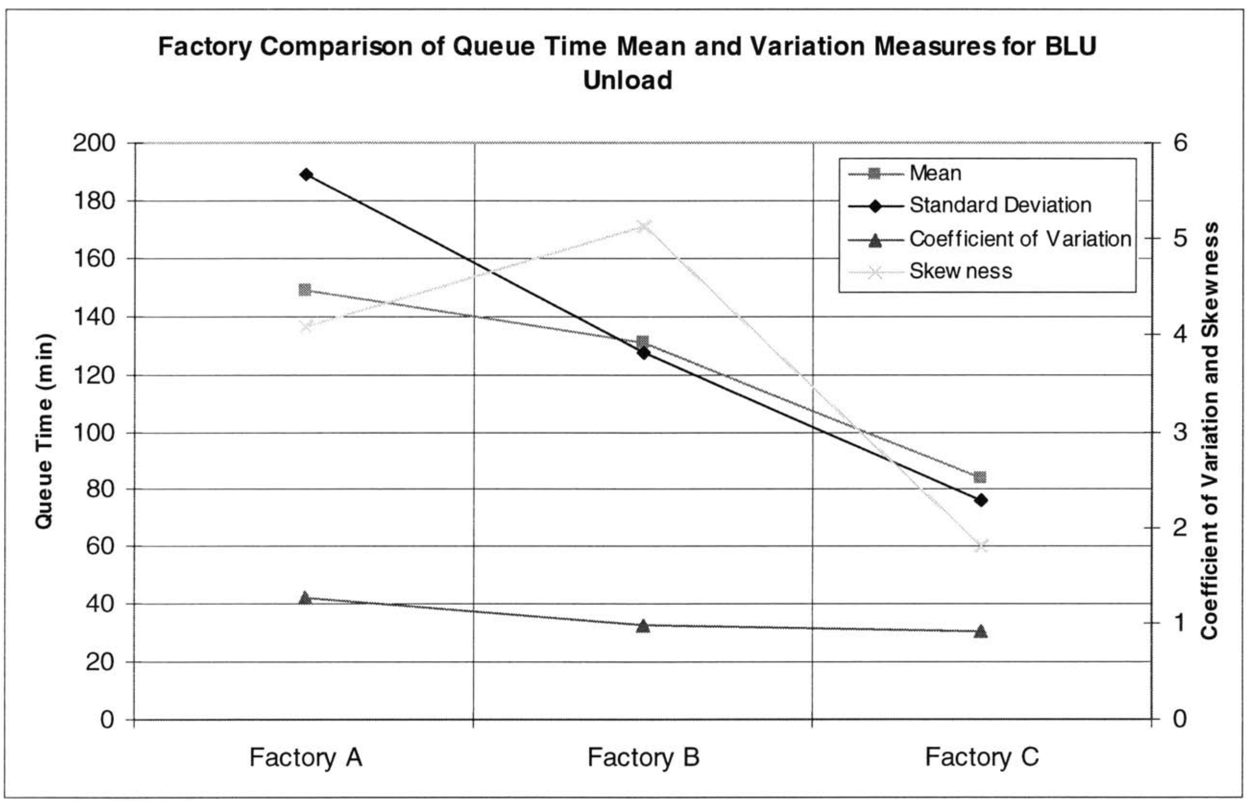


Figure 3.6 – BLU Unload Queue Time Variation

lower coefficient of variation for the BIO than for the other processes. This is because, as it is characterized in Workstream, the BIO is almost purely an automated procedure, with the exception of interrupts in the BI sequence.

3.4 Data Accuracy

There are a few issues associated with the data used in this study that should be noted. The first is that because Workstream data is manually entered, it is subject to human error. One such error is a practice that is common to all of the factories in which a operator enters a timestamp for a lot as it enters a process and then enters a timestamp indicating the lot has completed the process almost in the same instant. This practice usually occurs when the lot has been completed in an operation in which the operator failed to initially enter the starting time. Lots that have been subject to this practice are easily identified because they appear to have been processed in nearly zero time. While this erroneous data was purged from the process time data, the same logic could not be applied to the queue time data. Removing such data could have introduced bias, but this risk was considered appropriate.

Data associated with K-lots were also removed because of the difference in the manner in which K-lots are handled when compared to first-pass lots. As mentioned in the detailed process description, K-lot units can spend substantial periods of time in the Semi Finished Goods Inventory and appear in the data to have taken an inordinate length of time to be processed through BI. Within BI they are handled differently as well. Because of these differences, K-lots were removed from the data considered. Appendix A provides graphs that show the percent of useable data, where K-lot and near-zero-processes-time lots were unusable.

While both Workstream and the Sybase system data were available, inconsistencies in the time reference used in the two systems made comparison between the two data sources problematic. However, this comparison was required to derive distributions for the BIO load and pre-signal troubleshooting time for use in the simulation.

3.5 The Incentive to Quantify the Impact of Process Time Variation

While the data clearly shows substantial process time variation and the analysis demonstrates that effective capacity is negatively impacted, the question remains as to what the potential payoff is for reducing this variation. During the course of the investigation I suggested changes that involved either the reallocation of resources or investments in new process technology and labor. However, with each suggestion I faced questions on the appropriate level of resources to devote to variation reduction initiatives and the potential benefits of such initiatives. Rodrigues (Rodrigues, 2000) details specific changes to factory floor operations that will facilitate process time variation reduction and one of my goals was to augment that work with a quantification of the appropriate resource levels to dedicate to such initiatives. A discrete event simulation was build to aid in analyzing this problem, and is discussed in the next chapter.

Chapter 4: Simulation

4.1 The Simulation Strategy

I chose discrete event simulation as the analytical tool to aid in quantifying the impact of reducing the process time variation for BI. BI consists of three process steps with multiple machines at each step, and each machine has a random element associated with its processing time. The processing time variation in this area is a function of the total variation accumulated at each step of the BI process. Discrete event simulation offered a means to characterize microprocessor lots as discrete items with unique packets of information, called attributes. The attributes are modified in sequential steps and those modifications impact down-stream processing. A lot can therefore have an attribute that specifies the BI time, which can vary among the different types of products being analyzed, and the BI process will adjust itself to account for this difference.

The simulation strategy that was followed during this investigation was to first build a simulation that would serve as a baseline. The baseline was constructed based on a specific subset of operations, at one factory, during specific time period (Shannon, 1998). The fidelity of the simulation had to be of such a level that it captured the relevant characteristics of the process under study as they pertained to the questions being asked about the process. This becomes a subjective decision because the amount of detail engineered into the simulation is a function of how closely the builder would like the simulation output to match the selected data that represents the real system. However, the level of detail must be balanced with the resources available to develop the simulation and the run time needed to conduct studies. Additionally, the validation criteria was based on matching the basic character of process time and queue time probability distributions. Throughput for the process was also considered. A more qualitative verification and validation of the simulation was performed at the factory being with managers and technicians who were expert in the details of the BI process to ensure that relevant dynamics were captured (Sargent, 1998 and Robinson, 1997).

Following the simulation construction and validation, I conducted experiments and compared the output to the baseline. My conclusions were based on percentage changes from the baseline. This approach was consistent with the level of fidelity in the simulation, and I was confident that the percentage changes observed through this study were indicative of those changes possible in the real BI area. The primary focus of the experiments was variation reduction, but application of the simulation to capacity planning was also considered.

The simulation package chosen for this project was Extend™ (Imagine That Inc.), which offered a friendly user interface, the capability for the user to define random distributions based on data tables, the flexibility to change the source code of need be, and reasonable cost (Nikoukaran, 1998). Although Extend offered a fine suite of output tools, Excel was chosen to aide in presenting the output from the simulation and to couple that output with a financial model so that unit cost comparisons could be made.

4.2 Modeling Assumptions for Processes and Features External to Burn-in

A significant assumption in the simulation was the choice of the Flip Chip Pin Grid Array package type as the focus of this study. This assumption was made because it simplified the model of the operations feeding into the BI area without sacrificing accuracy, because the common practice was to dedicate tools to a specific package type. One exception was that BIOs were configured to run both FCPGA and the Organic Land Grid Array (OLGA) package. However, during the time period of this study the OLGA package was of sufficiently low volume to make this assumption reasonable.

FCPGA offered the benefit of being the highest volume product, so small changes in unit cost could equate to significant impact on Intel's costs. FCPGA was also being used to package a mature product with short BI times. This was important because, assuming that process time variation is independent of the BI time, the longer the BI time the less significant the variation when compared to the total BI process time.

The time period chosen for this simulation was the four workweeks that equated to September 2000. This time period was considered sufficient to capture shift-based daily and weekly cycles. Five days or 7200 minutes were added to the beginning of each simulation run as a settling-out period. The simulation begins with empty queues so a few days are required to permit the simulation to reach steady state performance (Centro and Reyes, 1998).

All of the Workstream data used to develop process time distributions was subjected to the same adjustments used in the variation study. Lots with process time that were impossibly short were removed from the data and K-lots were similarly not considered.

A simple model of the arriving lots from the upstream processes was developed based on Workstream data for the time period under study. The lot arrival program was based on a random number of lots arriving at two-hour increments. A two-hour arrival increment was chosen as the finest level of granularity possible with the program data table in Extend. This arrival increment (120 minutes) represents a small fraction of the total simulation run time of 47520 minutes (.25%). As the lots arrived the BI time attribute was set based on the probability that the product was one of two product variants. Over the course of the four weeks modeled, the mix of the two variants changed and the BI times for each variant changed.

The simulation does not explicitly model variations in lot size, but since one of the determinants of process time variation was the range of lot sizes, the process time distributions derived from Workstream data included the impact of varying lot size. Lot size has the most significant impact on the BLU load and unload sequences and impacts the manual load time as well. Larger lot sizes also increased the probability that troubleshooting in the Vcc-to-Ground check or the BIO pre-signal check will be required. Tool availability is not explicitly modeled, but is similarly implied in the process time distributions. If a tool operation is interrupted for maintenance while a lot is being processed, the delay will be reflected in the process time. Planned maintenance was not modeled because none was conducted during the time period studied. BIBs were assumed to be 100% reliable because no reliability data was available. Unreliable BIBs impacted the process time attributable to the Vcc-to-Ground check and the BIO pre-signal check and was thus modeled implicitly. The simulation provided output on the utilization of equipment and tooling so it is possible to assess the risk associated with this reliability assumption. The implications of the BIB reliability assumption will be discussed in more detail in Section 5.4 which discusses BIB inventory sensitivity studies.

A model was also required of the operation downstream of the BI area as a sink, which absorbed the throughput. This model of the PBIC was very simple and derived from Workstream data.

4.3 Modeling Assumptions for Processes and Features Internal to Burn-in

The inventory of BIBs was treated as existing in 67 unit increments. This assumption was based on the need to batch a lot of BIBs with a lot of units under test and the capacity of a single BIB. The simulation literally treats a lot of units as a single entity and the lot of BIBs as a single tooling item.

The BLU area was modeled as a series of random delays each representing two BLUs processing a single lot. Logic was created to represent the decision criteria of the BLU operator as to which operation to choose, load or unload. The simulated operator uses a "Toggle Logic" which consisted of noting the status of the lot just processed, either burned-in or yet to be burned-in, and chooses the next lot to be processed from the supply queue of the opposite status. So if the last lot processed was a load, the next lot would be an unload. Additional logic is installed so that the operator makes appropriate decisions in the face of empty supply queues or a depleted BIB inventory.

During the time period under study, the BLU load and unload sequences were so similar that it was assumed that just one of the distributions could be used to represent both. This simplified the BLU model and helped keep the run-time manageable. Figure 4.1 shows a comparison of the distributions for the load and unload operation at Factory A during the time period studied. The decision was made early on to chose the unload distribution to represent both load and unload because, unlike the load distribution, the unload distribution did not include the effect of the Vcc-to-Ground check. The Vcc-to-Ground check was planned as a separate model but found to be unnecessary in light of the close similarity between the load and unload distributions.

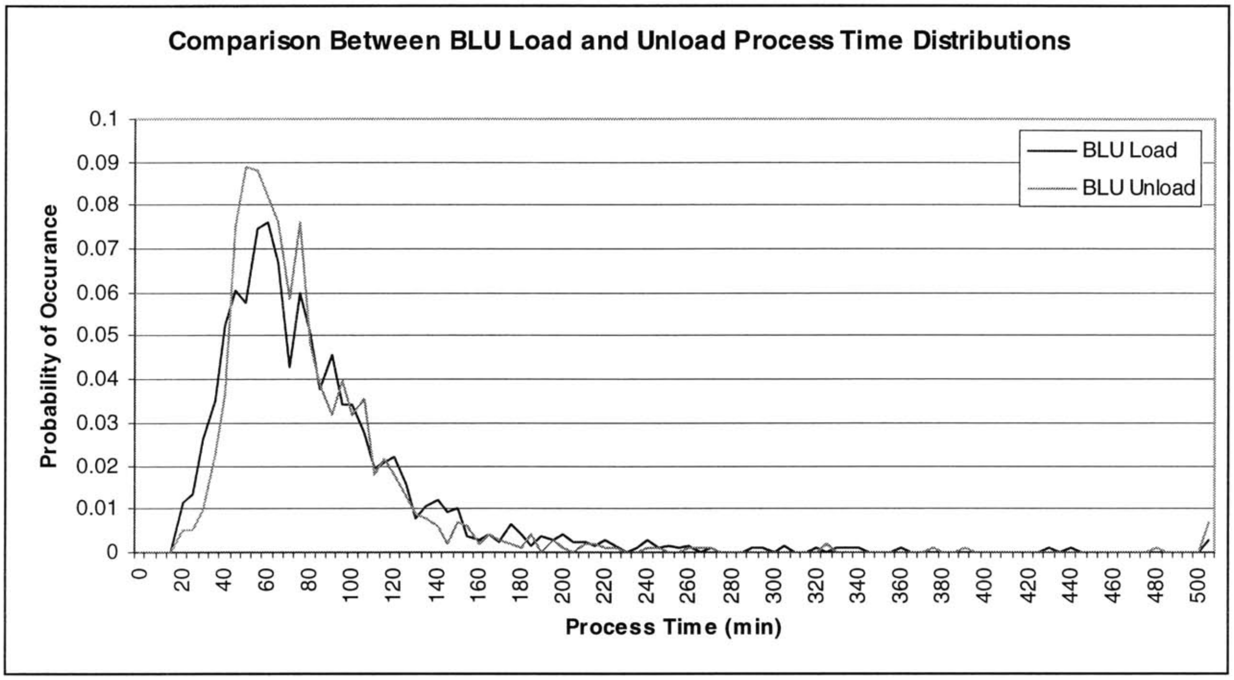


Figure 4.1 – Comparison Between BLU Load and Unload Process Time Distributions

It seems counter-intuitive that the Vcc-to-Ground check would have such an insignificant effect on the load process time, except when one considers the data shown in Figure 3.4 describing the process time differences between the load and unload sequence. An additional factor may have been that because the product was mature, the probability of shorts in the package had decreased to such a level that the Vcc-to-Ground check could be conducted in just a few minutes. This short of a delay would not be perceptible in the distribution given the inherent error associate with the manual entry of data into Workstream. The Vcc-to-Ground check was thus not explicitly modeled in the simulation.

Internal queues to the BI area included the WIP areas prior to the BIO and the BLU unload sequence. Internal queues were modeled as buffers between processes and thus a pure function of the run rates of the prior and subsequent operations. Validating the internal queues of the baseline simulation was a high priority because they reflected the combined effects of the process times and the dynamic interactions among the operations.

The BIO model was the most complex element of the simulation. In this model each step of the BIO process, the load and pre-signal troubleshooting, the ramp to BI temp and the BI time, and the unload, was modeled as a separate delay. Each delay had a unique random distribution. Manning ratio was explicitly modeled through a labor pool that took scheduled breaks during the course of a shift. The BIO processing time was a function of the required BI time read from the incoming lot. During any week in the simulation run there were as many as three unique BI times varying from .8 hours to 1.8 hours and BI time varied from week to week.

Appendix B presents the technical details of the simulation.

4.4 Validation

Three validation measures were used, process time probability distributions, queue time probability distributions, and throughput history. Figure 4.2 and Figure 4.3 show the comparison between the simulation and the real BI area for the BLU load and unload sequences.

This comparison shows that the simulation closely approximates the real system. As one would expect the BLU unload model is a slightly closer match to the data than the BLU load model because BLU unload data was used to generate the model.

The comparisons for the BIO process time and the internal queue times were done using a log scale in time. This was necessary because the distributions were highly skewed and a log scale permitted a more visible comparison between the distributions.

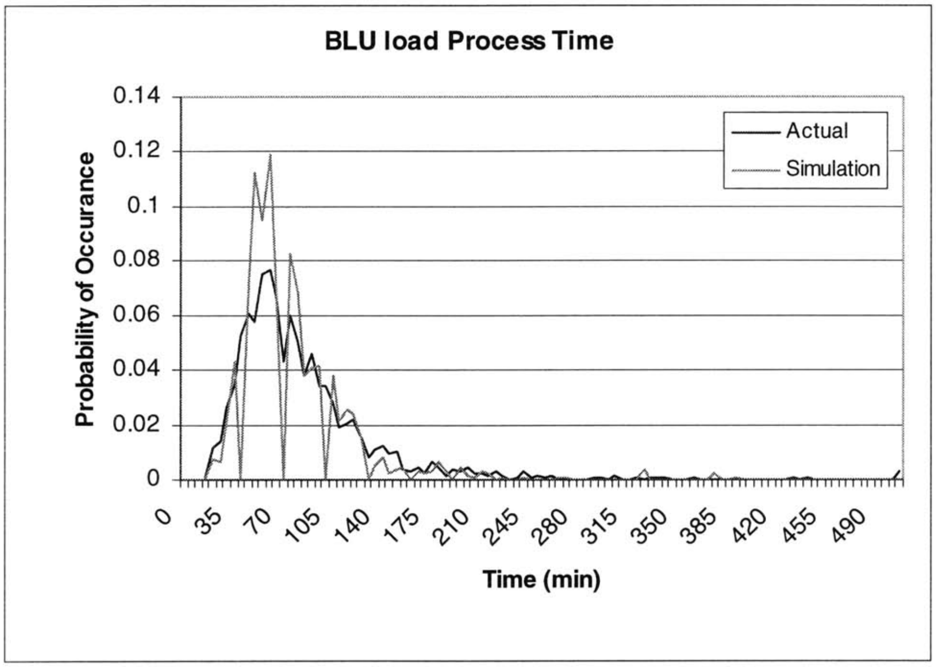


Figure 4.2 – Simulation vs. Actual Comparison for BLU Load Process Time

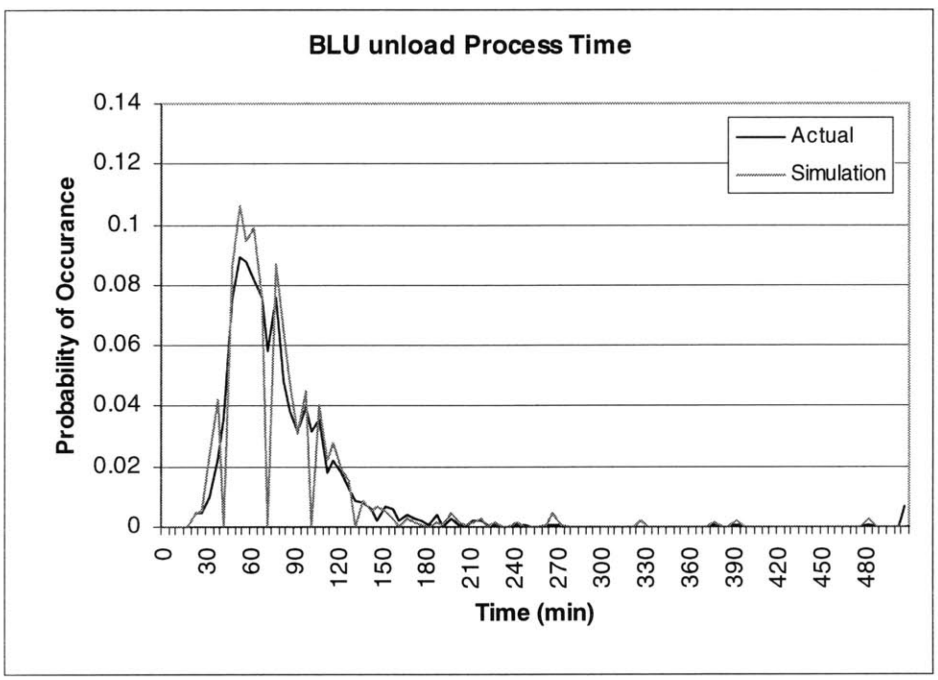


Figure 4.3 – Simulation vs. Actual Comparison for BLU Unload Process Time

Figure 4.4 shows the comparison between the simulated BIO area and the actual. The simulation is pessimistic in that it predicts a higher degree of variation than the actual system. This difference is attributable to inaccuracies in the data used to produce the BIO model. The most difficult component of the BIO model to build was the load and pre-signal check model because no explicit data source existed for this process. It was thus necessary to compare data gathered from Workstream to data gathered from the Sybase system. A source of error in the resulting distribution could be the difference in the reference time used by the different systems.

The BIO unload model was based on observations of operators and labor models which captured mean unload times. The resulting load model was a constant 20 minutes with a Gamma distribution added with a scale factor of 1.5 and a shape factor of 2. The Gamma distribution would generally add about three minutes to the load time, but could add as much as 12 minutes. The BIO model captured the dynamic of an operator being delayed in unloading one lot because he or she is occupied troubleshooting another lot. This dynamic became accentuated when the troubleshooting delay model was pessimistic.

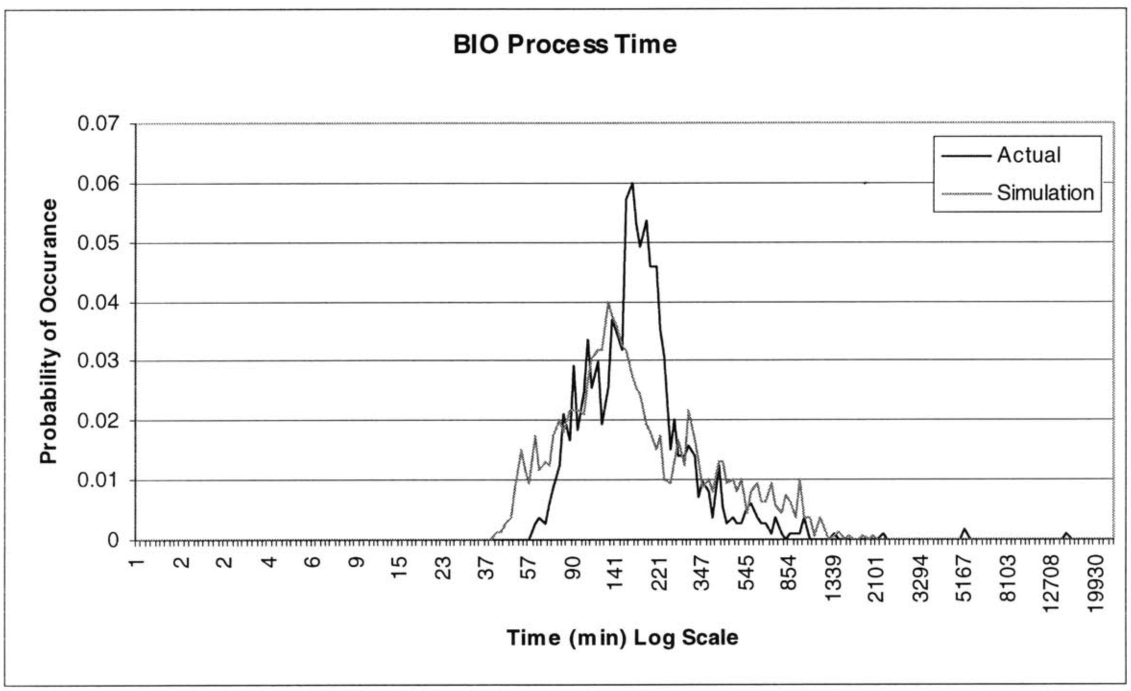


Figure 4.4 – Simulation vs. Actual Comparison for BIO Process Time

This pessimism in the BIO process time model decreased the throughput result for the entire simulation because the simulation has less effective capacity than the real system. This result is apparent in Figure 4.5, which compares the output of the baseline simulation to the real BI area for the time period studied. It is interesting to note that this comparison highlights the fundamental premise of this thesis, which is that variation reduces effective capacity. The difference noted in the BIO validation were considered acceptable because the causes of the differences were understood and that the critical dynamics of the BIO area were captured.

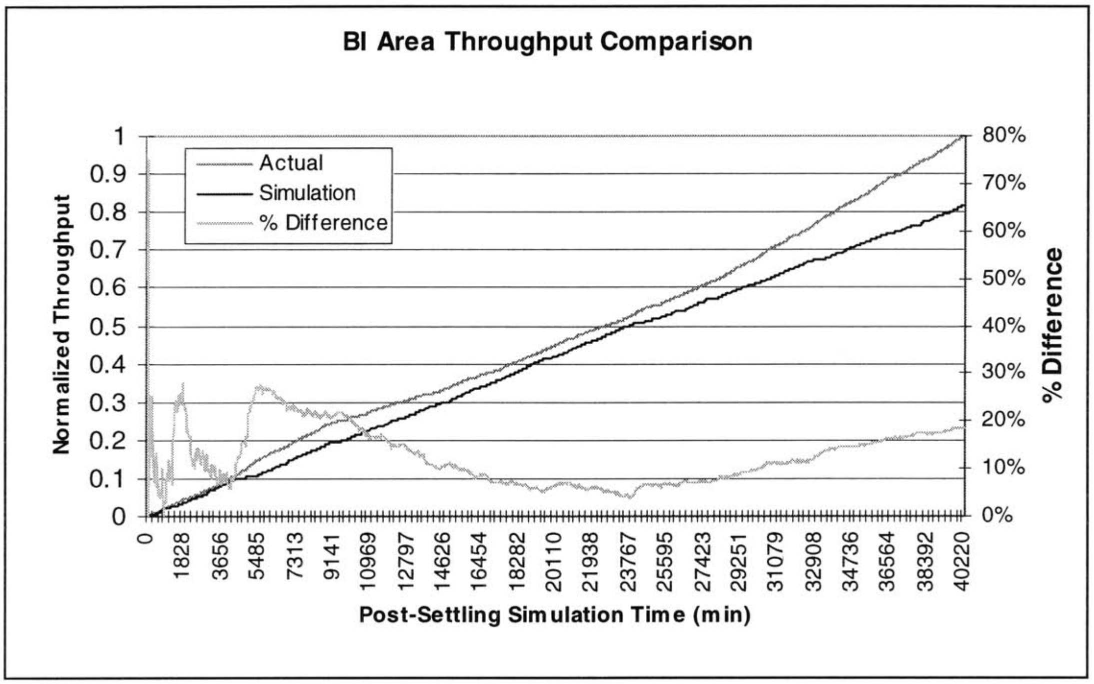


Figure 4.5 – Simulation vs. Actual Comparison of BI Area Throughput

Figure 4.6 and Figure 4.7 show the comparison of the internal BI queues. The queue distributions generated from the simulation are reasonable approximations of the actual BI area. Because the queues were modeled simply as buffers between processes, they are good measures of the validity of the simulation as a complete system rather than, as was the case with the process times, the accuracy of individual machine models. The comparison shows consistency between the simulated distributions and those of the real process that validates the overall simulation for use as a baseline.

This comparison, however, also highlights the data integrity issue with Workstream in which operators enter two timestamps representing the beginning and the end of a process step almost instantaneously. For process time data it was possible to adjust the data based on an understanding of the physical limits of the BI area tools. For queues that same logic could not be applied. It was quite possible for a lot to move immediately from a BLU to a the BIO or from a BIO to a BLU without spending a significant amount of time in the WIP area. According to the data used to generate Figure 4.6, there is a very high probability, between 2 and 3 percent, that lots required almost zero time to move from one operation to the next. It seems very unlikely that this would be the case given that the probability drops off substantially for times of 2 to 10 minutes. In addition, the data in Figure 4.6 includes time for the pre-signal check, which at a minimum requires approximately 10 minutes to complete.

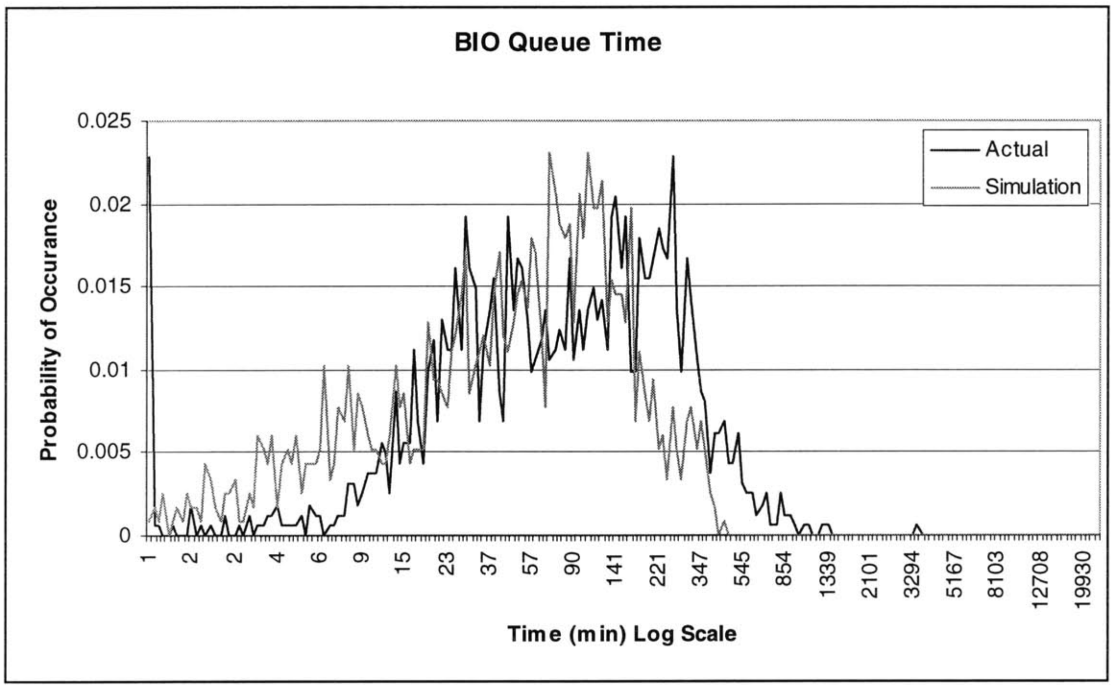


Figure 4.6 – Simulation vs. Actual Comparison of BIO Queue Time

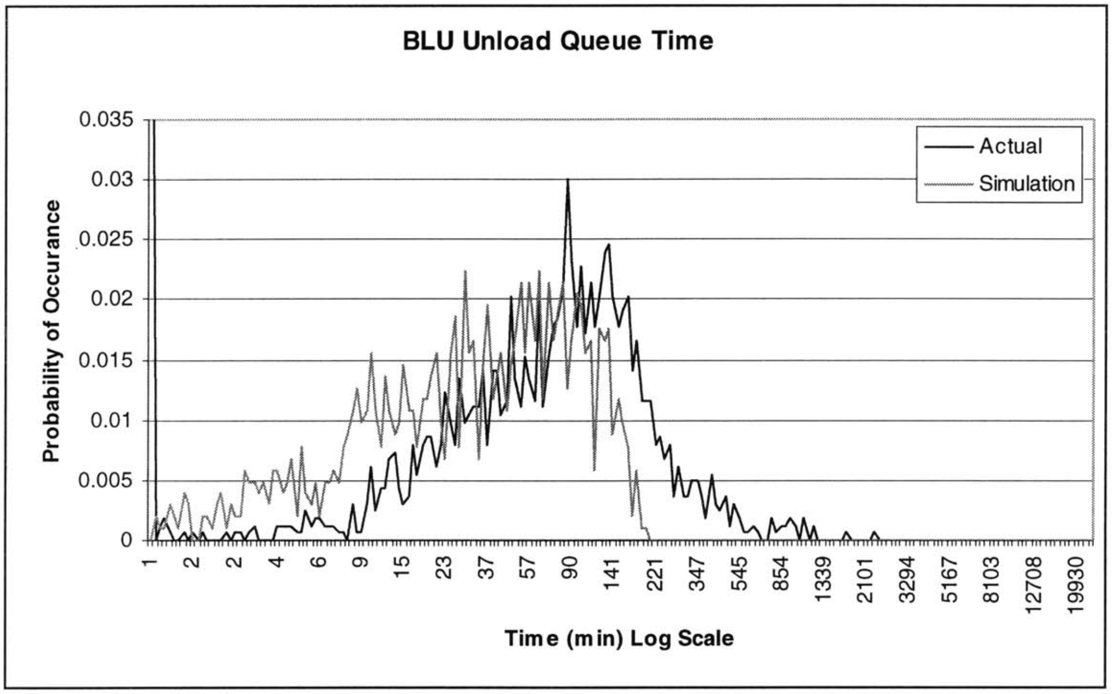


Figure 4.7 – Simulation vs. Actual Comparison of BLU Unload Queue Time

Chapter 5: Experiments

5.1 Coupling the Simulation with a Financial Model

Experimentation using the simulation required the development of a financial model that could be coupled with throughput information provided by the simulation to determine unit cost. This model was a simplification of the detailed financial model used by the Assembly/Test Manufacturing Finance group in that it only included costs attributable to the BI area. Specifics costs considered were also a simplification of the finance group's model in that they included labor, indirect materials and depreciations. Costs not included within the indirect materials cost category were, upgrade expenses, allocations, and spares. Within the depreciation cost category, upgrade depreciation, used equipment depreciation (all equipment was assumed new), and miscellaneous were not considered. The financial model developed for this study included inventory carrying cost, which was not the case with the ATM Finance model.

Figure 5.1 compares the magnitude of costs associated with the BI area. The indirect material cost, which consists exclusively of BIB inventory cost, is 60% of the total BI

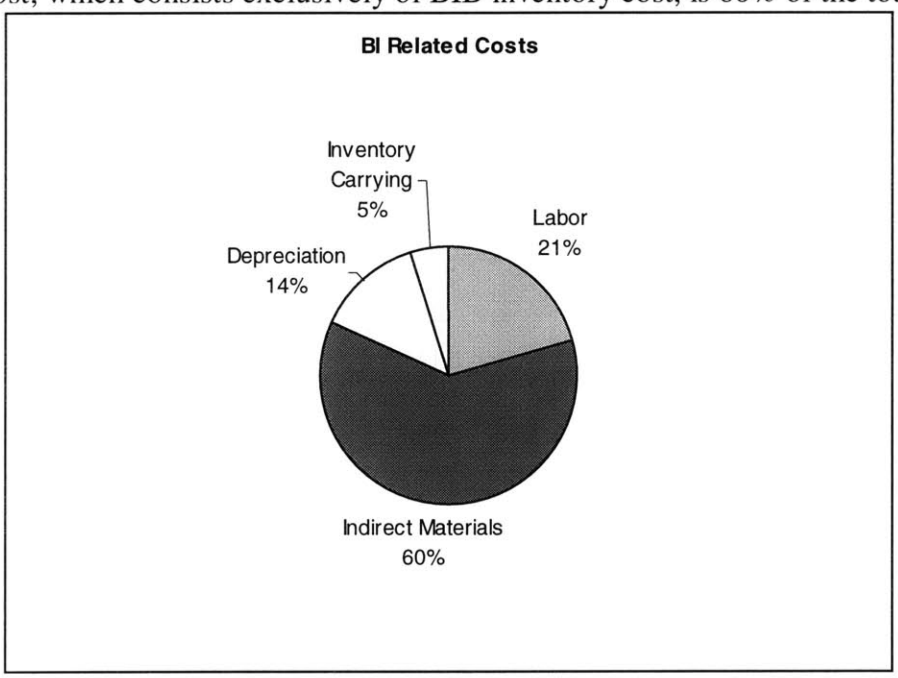


Figure 5.1 – BI Related Costs

area costs. While inventory-carrying cost is only 5% of the unit cost attributable to BI, product inventory and BIB inventory are closely related because the product occupies a BIB while it is in BI. The amount of WIP in the BI area drives the need for an increase in the BIB inventory. It is important to note that large amounts of WIP within BI did not seem to drive decisions at the factory floor level until BIB costs became prohibitive.

5.2 Maximum Capacity Experiments

The experiments were run as Monte Carlo studies with a single study consisting of 32 simulation runs. In each run the random number seed was varied randomly. The following observations are noteworthy based on running the baseline simulation in such a fashion. The standard deviation of the throughput, which offers a perspective on the predictability of the BI area, was 2.7% of the average throughput. The average WIP level internal to the BI area over the course of the four-week simulation was 23.9 lots.

The simulation offered the ability to increase the arrival rate to BI so that a steady stream of WIP was available (the baseline simulation showed no WIP available for BI 3.8% of the time) and the maximum capacity of the system could be determined. Appendix C shows the results of a series of simulation scenarios and the result of increasing the arrival rate to the BI area such that the system is never starved. The maximum capacity was approximately 5% above the capacity at which the system was operating in the baseline scenario. This insight answers a fundamental question that was the impetus for this study, why BI constrains the factory despite large amounts of excess capacity. The answer is that while between 35 and 45% extra capacity is planned into BI, the real system only has 5% extra capacity. This difference is due to high process time variation coupled with a multi-step, multi-machine operation with imbedded queues.

The next set of experiments were based on assessing the impact on throughput and unit cost resulting from reducing the variation associated with the BLU and BIO processing times. Appendix C shows the results of studying variation reduction in the BLU and BIO process times separately and then looking at the coupled impact of both variation reduction efforts applied simultaneously. BLU process time variation reduction alone increased maximum throughput by 3.5% above the theoretical maximum capacity of the baseline and decreased the unit cost by 3%. Variation reduction in the BIO resulted in a throughput increase of 20.2% above the theoretical maximum capacity of the baseline and a decrease in unit cost of 15.4%. The coupled impact of reducing variation in both processes was a 54.1% increase in throughput and a 32.4% decrease in unit cost.

Two important insights can be drawn from these results. First, there is a much greater potential reward by addressing the variation in the BIO process time than the reward offered by reducing process time in the BLU. Second the combined effect of variation reduction to both processes is not a linear combination of the two reductions accomplished separately. The coupled impact is significantly higher than the independent impact. This result highlights the importance of treating the BI process as a single system made up of coupled operations rather than as a set of independent processes.

The standard deviation of the throughput was also reduced as a result of reducing the process time variation in the BLU and BIO processes. The combined impact of reducing variation in both the BLU and the BIO processes reduced the throughput variation to .4% of the average throughput for the time period simulated. This compares to 2.7% for the baseline scenario, an 80% reduction.

5.3 Proof of Concept – Capacity Planning Experiments

The simulation provided the average queue time for the internal queues and this information was useful in understanding the interaction between the BLU and the BIO processes. In the baseline simulation, the average queue times for the BIO and the BLU unload were 96.6 and 68.6 minutes respectively. In the simulation where process time variation was reduced for both processes, the BIO and BLU unload queue times changed to 9.9 and 80.0 minutes respectively. The shift in WIP from the BIO queue to the BLU unload queue is a logical outcome of increasing the run rate of the BIO to a greater degree than the BLU. Thus process time variation reduction caused a capacity imbalance between the BLUs and the BIOs.

If maximum throughput for minimum unit cost is the goal for the BI area, however, the question becomes can BLU capacity be added to correct the capacity imbalance and increase the throughput without increasing the unit cost. Appendix C shows that this result can be achieved. By adding two additional BLUs and one operator, a combination, which can be considered a single unit of BLU capacity, the throughput for the BI area is increased by 69.5% and the unit cost decreased by 35.4% when compared to the baseline running at maximum capacity. Adding an additional unit of BLU capacity increases throughput to 73.6% above the baseline but results in only a 36.3% reduction in unit cost. This scenario also showed a very high utilization for BIBs indicating that having added BLU capacity the next local constraint for BI was the BIB inventory. A final scenario shows the impact of increasing the BIB inventory to accommodate more WIP. Throughput is increased to 78% above the baseline running at maximum capacity, but unit cost is decreased by only 33.9%. These results show that there is an optimal combination of tool capacity and BIB inventory that will minimize unit cost and that the simulation can aid in determining this combination. Such a study would make useful follow-on research.

The importance of this results lies in insight provided about the dynamics of the system. The simulation tool can thus be a valuable supplement to a static spread sheet-based capacity planning model. When variation reduction efforts are undertaken, the simulation can highlight instances in which run rate mismatches are likely to occur and then offer a method for adjusting tool capacity to achieve the highest possible throughput for BI at the lowest cost.

5.4 Burn-in Board Sensitivity Studies

Thus far the analysis has focused on how variation impacts maximum throughput for the BI area. In reality, ATM does not want BI running at maximum capacity, but rather to carry some excess capacity as insurance against an unforeseen spike in demand. It is a simple enough matter to use the simulation to plan for excess capacity in tools, but a questions remains as to the appropriate BIB inventory to maintain. This question can be addressed by conducting sensitivity studies with BIB inventory as the parameter being varied. Figure 5.2 shows a BIB sensitivity study for the baseline simulation. This graphic shows that as BIB inventory is increased, BI area throughput increases, but only to a point. At approximately 1500 BIBs, the throughput curve plateaus and additional BIB inventory does not result in higher throughput. The thick vertical lines show the range of BIB inventory during the time frame of the study. Based on the graph, it appears that the BIB inventory is adequate because it does not constrain the throughput. However, an important assumption to note is that the simulation assumes a perfectly reliable BIB. In reality, BIBs are significantly less than 100% available, and while the capacity model assumes an 87% available BIB, this estimate may not be accurate. The simulation helps to some extent with this dilemma in providing utilization information during the period under study. Figure 5.2 shows that if the BIB inventory is approximately 95-97% available, the inventory will be sufficient to ensure that BIBs do not constrain the BI area. The fact that BIBs are much less than 95% available explains the numerous complaints volunteered by operators that the BIB inventory was inadequate.

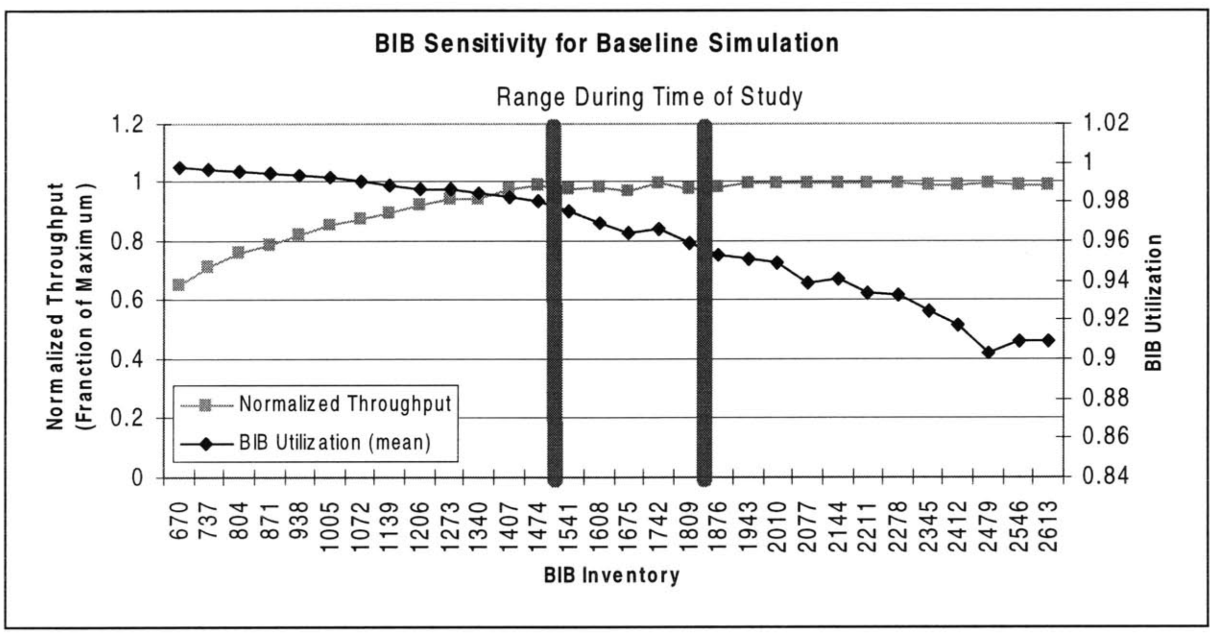


Figure 5.2 – BIB Sensitivity Study for Baseline

Ideally this study would be conducted numerous times, varying the random number seed, to get a sense for the statistical sensitivity. Such a set of studies would produce bands rather than the lines depicted in Figure 5.2. The increases in utilization at the 1742 and 2546 BIB inventory levels are within the noise of the simulation.

Based on the cost model discussed, there is a strong incentive to reduce the BIB inventory, which is directly related to WIP. In the variation reduction simulation studies conducted without increasing the arrival rate (so the process was not running at maximum capacity), it was observed that the average WIP levels in the BI area decreased substantially, from 23.9 lots in the baseline to 16.3 lots. This offered an opportunity to study the impact of reducing the BIB inventory. Figure 5.3 shows the results of this sensitivity study. The throughput curve plateaus at a much lower BIB inventory level than the in the baseline case shown in Figure 5.2.

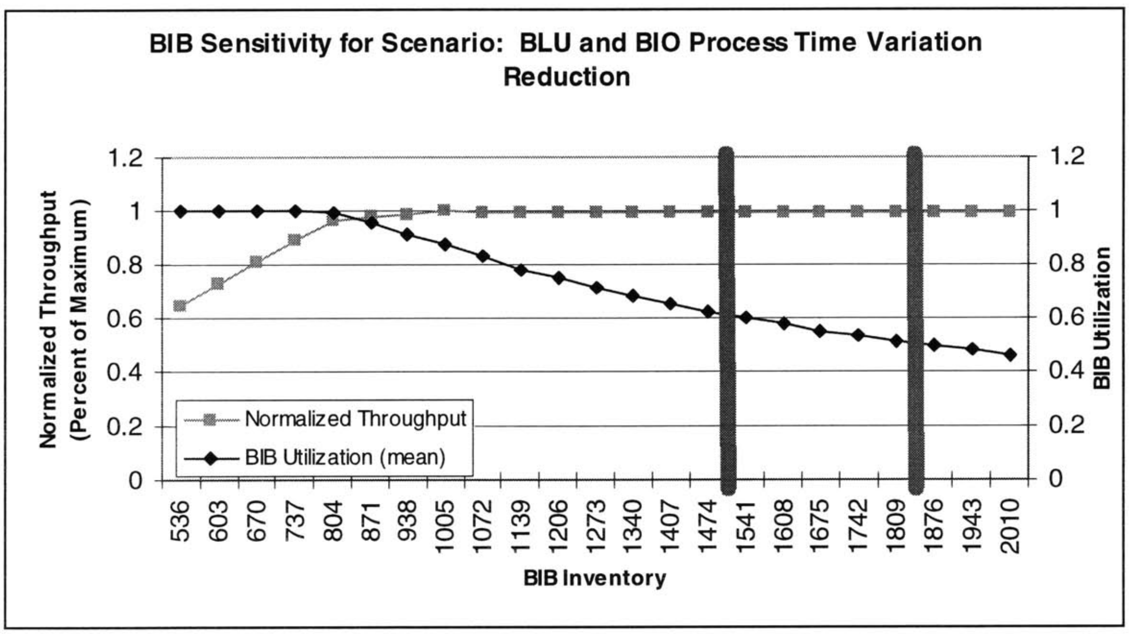


Figure 5.3 – BIB Sensitivity Study Reduced Variation Scenario

If one were using this graphic for capacity planning, the first step would be to enter the chart with an assumed BIB availability. For example, assume the BIB inventory is 80% available. Entering the chart at .8 intersects the utilization curve at approximately 1100 BIBs. This level of BIB inventory is safely above the plateau point for the throughput and well below the actual BIB inventory range during the time period studied. The cost implication of such a reduction in BIB inventory is a 31.3% decrease in unit cost attributable to the BI area.

Chapter 6: Conclusions and Recommendations

6.1 Proving the Hypothesis – The Impact of Variation

The hypothesis of this work was that variation in processing time contributed to BI becoming an unplanned constraint for Intel's Assembly/Test Manufacturing Factories. In addition to proving this hypothesis, this work aimed to quantify the potential impact of decreasing the process time variation, in terms of throughput and unit cost. A final, and less explicit, goal of the project was to build an analytical tool that offered insights that were useful beyond the scope of this work.

The hypothesis was proven by comparing the fixed process time values assumed in the capacity model to probability distributions of actual process time for lots processed in Intel's factories. Because the variation was significant, the capacity planner's confidence in having sufficient capacity should be as low as 50%. This type of comparison was new for the industrial engineers and process engineers in the factories because process time variation reduction was much less a focus than mean process time reduction.

A discrete event simulation was built to aid in quantifying the potential rewards of reducing process time variation. The simulation output when coupled with a financial model showed that maximum throughput could be increased by over 54% and unit cost attributable to the BI area decreased by over 32% through conservative variation reductions efforts.

The simulation also offered insights into the dynamics of the BI operation that suggested using the simulation to supplement the capacity planning process. Specifically, the simulation provided information on the capacity balance between the BLUs and the BIOs. Such information could be used in conjunction with the spreadsheet-based model to achieve the maximum possible throughput for minimum unit cost from the BI area through a closer alignment between BLU and BIO capacity. This use of the simulation was beyond the scope of this work, but the proof of concept was demonstrated.

6.2 Recommendations

6.2.1 Reduce Process Time Variation

There are two general recommendations of this thesis. First, the Assembly Test Manufacturing people with an interest in the effectiveness of the BI area should focus their efforts on reducing the process times variation associated with BLUs and BIOs. To this end, first priority should be given to reducing the delays attributable to manual operations such as pre-signal check troubleshooting. Second, the capacity planning process should include a dynamic analysis of the BI area that includes treating the operation as a system of dependent and coupled processes rather than independent tools. The simulation is a tool that offers insight into these system dynamics.

In the course of this study a number of issues surfaced that would need to be addressed as part of an implementation of process time variation reduction. Data integrity and data availability are the most critical to issues. Data integrity means ensuring that the data available is accurate. Workstream data is subject to inaccuracies attributable to the operator's practices. Operators should be educated on the value of accurate data and their critical role in capturing this data. Incentive systems should be modified so that data accuracy as it is tied to process improvement is rewarded. Another source of data inaccuracy is inconsistency among different systems. It is difficult to compare data from two sources such as Workstream and Sybase, because they lack a common time reference.

Data availability means having data that is relevant to the problem one is trying to address. In this case, data that must be available includes lot based time measurements that are of sufficient granularity to capture key events. Such granularity would permit differentiation of the Vcc-to-Ground check from the BLU load sequence, and differentiation of the BIO load, pre-signal check and troubleshooting as discrete process steps separate from the BIO queue time. To echo the recommendation of Newlin (2000), tool and BIB availability and utilization need to be directly measured. Where reliability is an issue, root cause analysis should be accomplished. Data should be gathered that facilitates such root cause analysis. For example, if BIB reliability contributes to the delay associated with troubleshooting during the BIO pre-signal check, data should indicate which lots required troubleshooting and if that troubleshooting could be attributable to a BIB fault. Additionally, the cause of BIB faults should be tracked. If BIBs are failing frequently, is it because of blown fuses and if so, is there a trend in blown fuses that can be tracked to the power consumption of newer generation products?

The most critical step in a process time variation reduction effort is to implement a data collection scheme that provides meaningful and accurate insight. In the short term, a team that assumes responsibility to reduce variation will depend on manual data gathering. However, consistent manual data gathering is difficult to sustain, so in the long-term automated data sources should be implemented. This project provided a means of quantifying the appropriate level of resources to devote to a variation reduction project, which ultimately must include improved data gathering systems.

The obvious benefit of gathering accurate and relevant data in a variation reduction effort is that this data provides feedback on the impact of the changes being implemented. The less obvious benefit is that data gathering directed at a specific problem such as variation reduction highlights root causes that can then be addressed as part of the variation reduction effort.

With a data system in place, the next step should be to chose a set of variation reduction targets such as ensuring that no lot takes longer than 100 minutes to be processed through BLU load and that no lot takes longer than 75 minutes to be processed through the BIO pre-signal check. The operators incentive system should be reviewed to ensure that such variation reduction goals are not in conflict with their existing goals. Once again, the simulation offers a way to quantify the level of resources that is appropriate to devote to such efforts.

6.2.2 Use Simulation to Augment the Static Capacity Model

The existing capacity planning model is well suited for a process that is static or has achieved steady state. However, because Intel is continually introducing new products, capacity planning must consider dynamic capacity ramps. The capacity model attempts to capture these demand driven dynamics by calculating capacity on a weekly basis, but does not account for process variation. A process time variation reduction initiative would add additional dynamics to the existing BI process that could not be accounted for in the static model.

Because process time variation reduction increases the effective output of a set of sequential tools, the effect can be limited when variation is reduced by different amounts in different operations. Such imbalances should lead capacity planners to consider the interaction among tools in addition to current method that treats BI as independent operations. The simulation showed that in one set of circumstances, in order to maximize the throughput from the BI area, BLU capacity in excess of the amount determined by the static model was required.

The simulation should be used to help industrial and process engineers understand the dynamics of BI, and the impact of variation reduction on effective tool capacity. Variation reduction should be considered an iterative process in which an initiative is implement, data gathered, and the impact assessed before additional initiatives are undertaken. The simulation should be an integral part of this iteration. Having used the simulation in this thesis to identify areas of improvement, the next step is to implement these suggestions (Ishiwata, 1984). Data that is gathered should be used, in addition to understanding the impact of the variation reduction initiative, to make the simulation more accurate. Then the simulation can be used to assess further areas of potential improvement. At this point capacity planners can use the simulation to assess the potential impact of ongoing and future variation reduction efforts on capacity.

6.3 Implication for the Next Generation of Burn-in

Assembly/Test Manufacturing is currently implementing the Next Generation of Burn-in (NGBI) tools and processes. This thesis, while not directly addressing issues associated with NGBI, addressed processes that will be part of NGBI. For instance, NGBI uses the same BLUs as the existing process. More important, the subject of process time variation has a direct bearing on any process and should be considered in the implementation of a new process and the installation of new equipment.

6.4 Opportunities for Future Research

This thesis is not the first to highlight the value of using simulation to gain better understanding of a manufacturing process (Dieter, 1998, Domaschke, 1998, Perrin, 1997). What was unique for Central Industrial Engineering group at Intel was the use of an off-the-shelf simulation package with a friendly user interface. Simulation is now a tool that can be used by an analyst close to the factory floor problems, rather than through a centralized group of simulation experts that must prioritize support requests.

Opportunities for future research include:

Using simulation to explore variation in other ATM operations from a tactical perspective. A specific study that has near-term implications is to understand the relationship between BI and PBIC. A simulation of the combination of both areas might add insight that would allow factory planners to determine which area should be planned as the constraint. It would also be valuable to simulation NGBI once sufficient data has been gathered on the new process's performance.

In considering the use of simulation as a tool in capacity planning, a study that addressed the combination of optimization with simulation would be useful. The proof of concept explored in this paper was limited to varying just a few parameters. A design of experiments or a software-driven optimization routine would provide information that would be valuable in understanding the dynamics of BI and in improving the simulation.

Appendix A

This appendix includes data describing process time and queue time distributions for four Assembly/Test Manufacturing factories from 8/20/00 – 9/2/00.

The following designations are used to represent BI processes in this appendix:

7221 QT – BLU load queue time
7221 PT – BLU load process time
7231 QT – BIO queue time
7231 PT – BIO process time
7241 QT – BLU unload queue time
7241 PT – BLU unload process time

Figure A.1.1 - Factory A – Probability density and cumulative distributions.

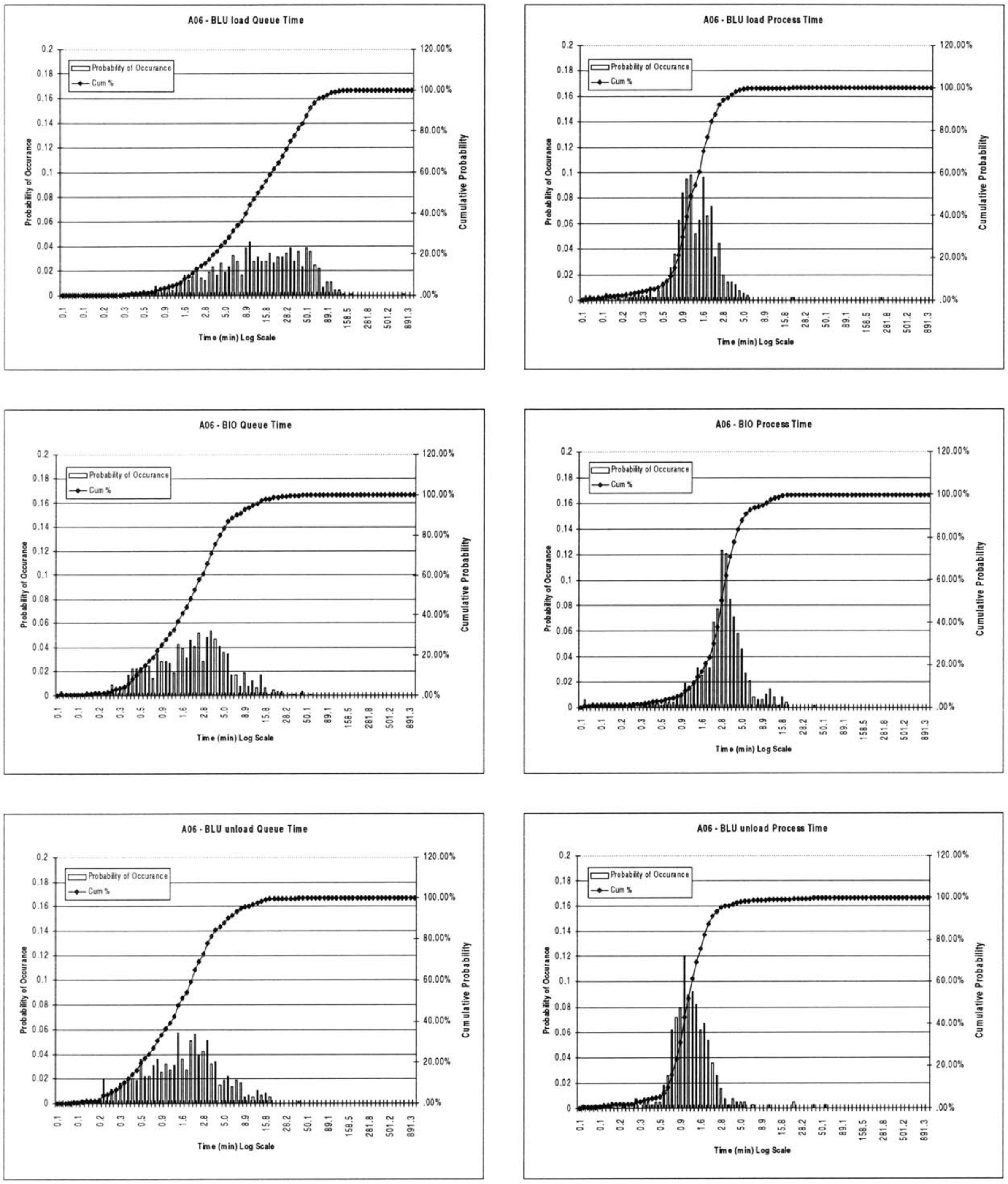


Figure A.1.2 - Factory A – Process-to-process comparison:

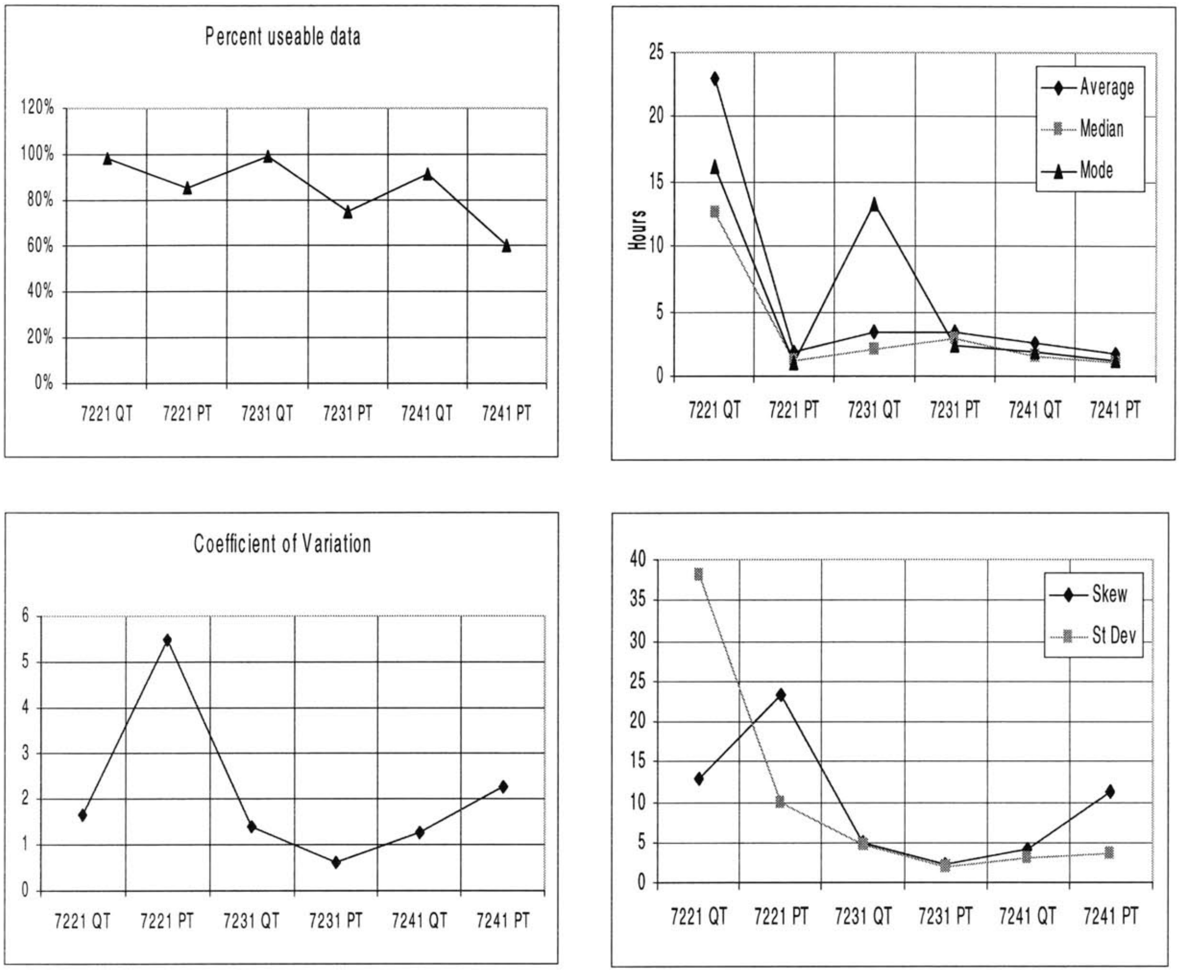


Figure A.2.1 - Factory B – Probability density and cumulative distributions.

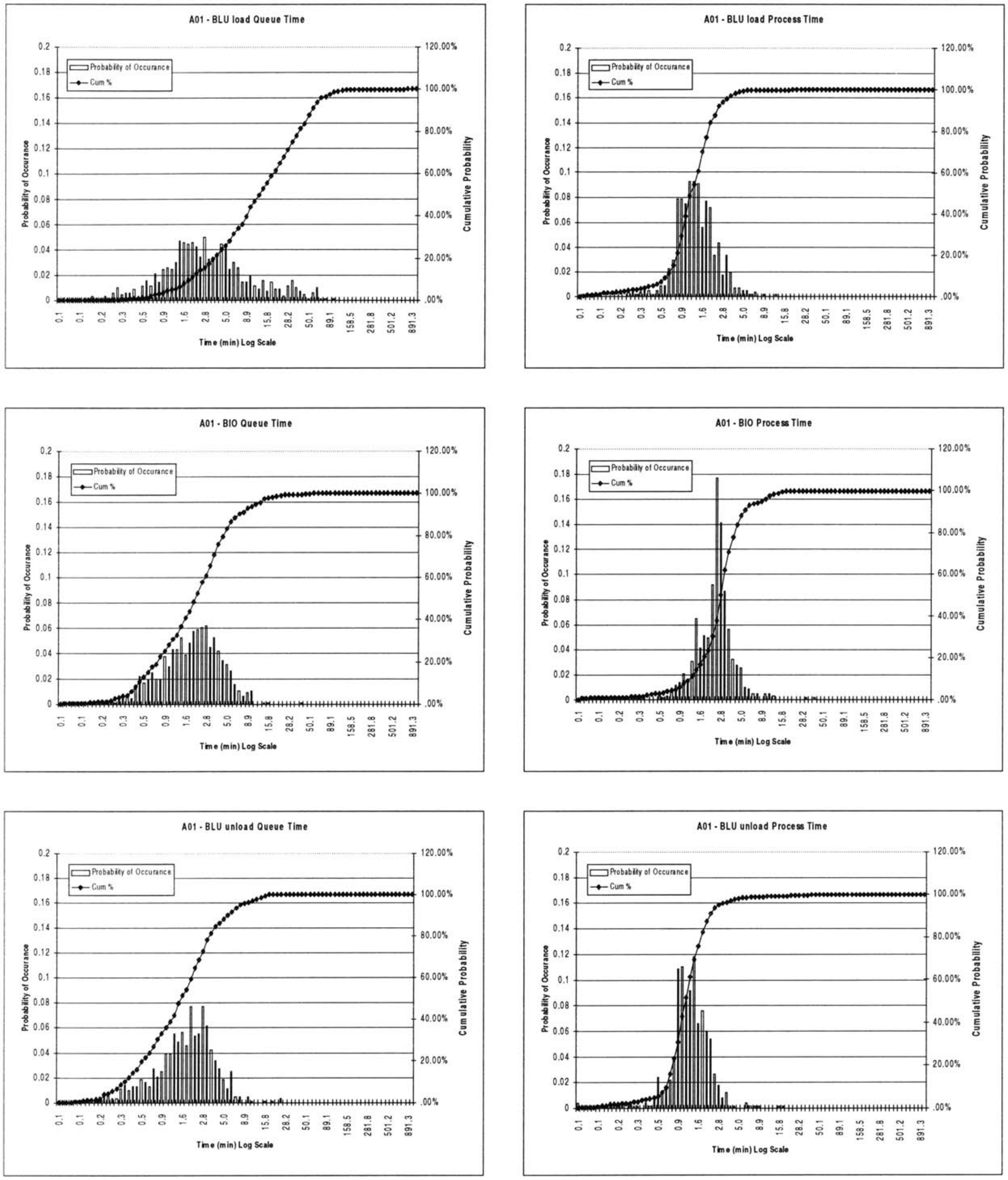


Figure A.2.2 - Factory B – Process-to-process comparison:

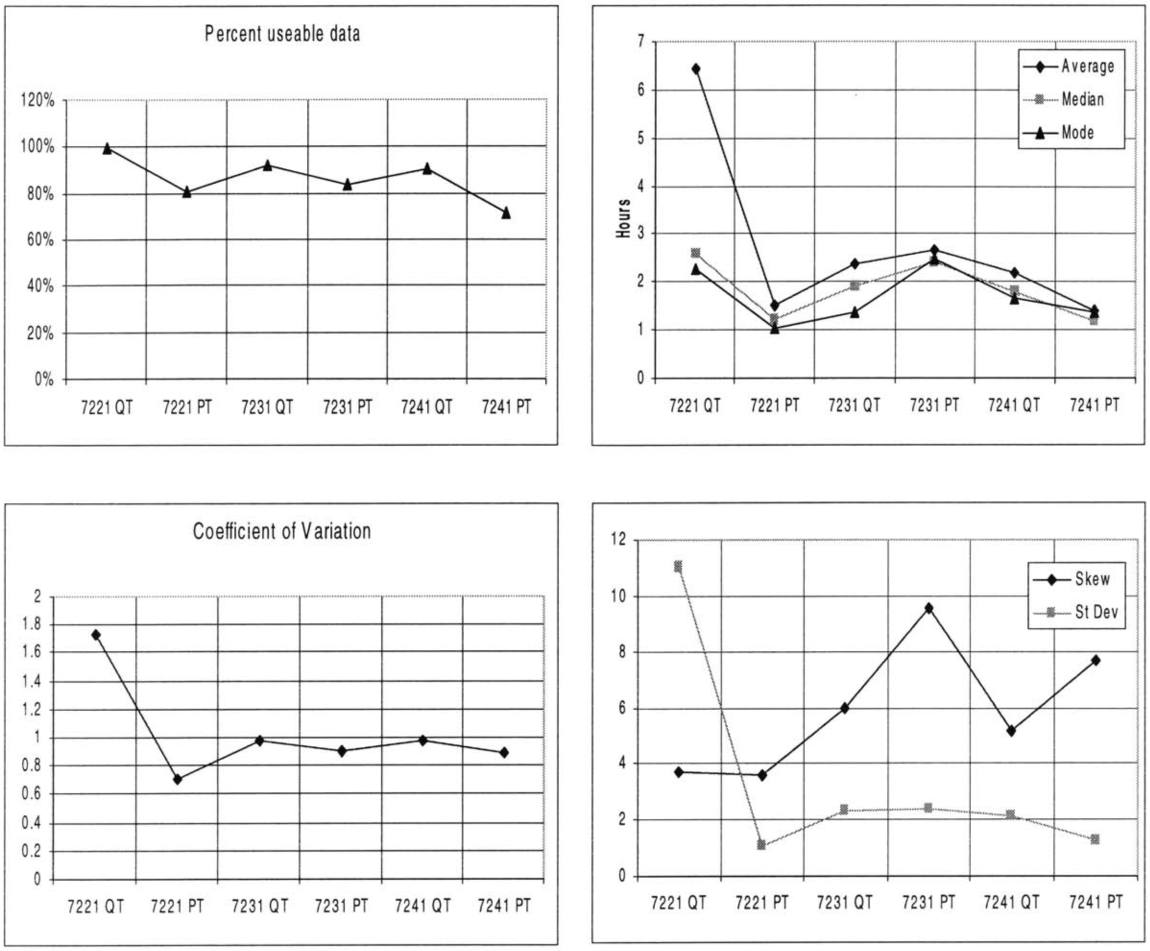
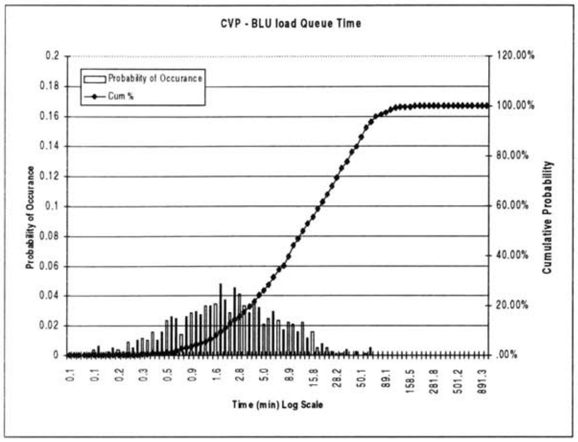
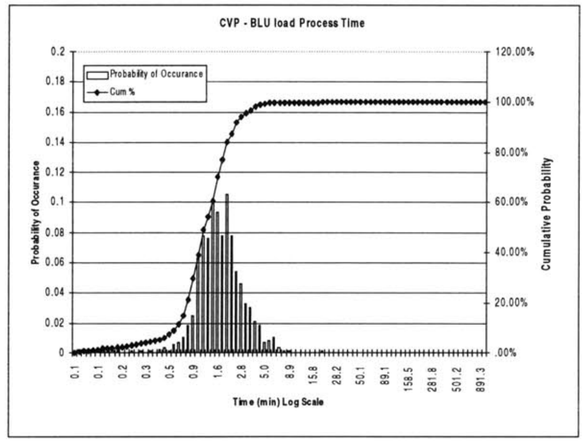
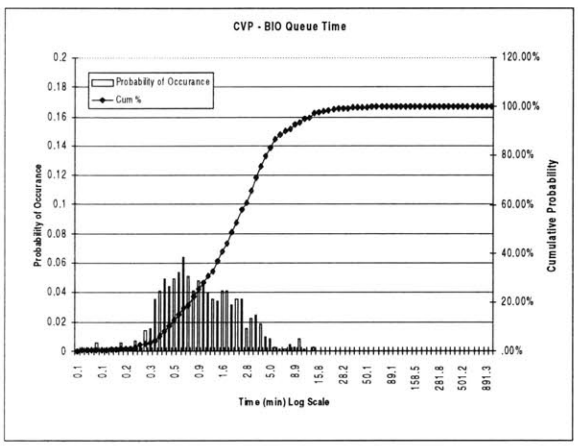
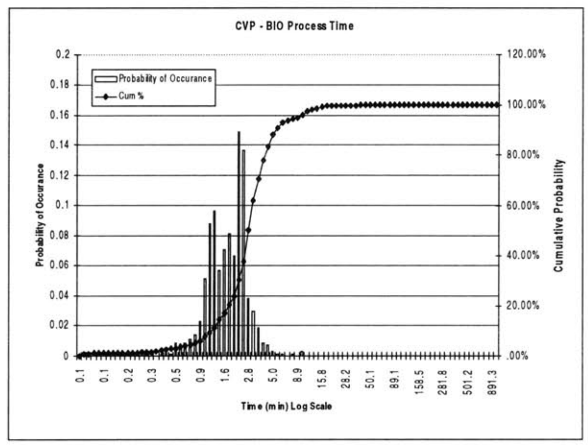


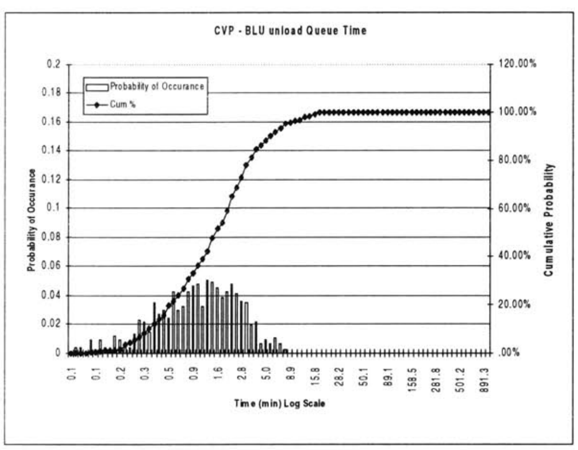
Figure A.3.1 - Factory C – Probability density and cumulative distributions.











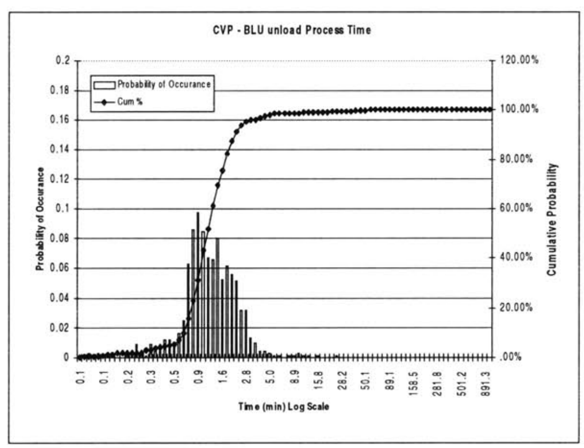


Figure A.3.2 - Factory C – Process-to-process comparison:

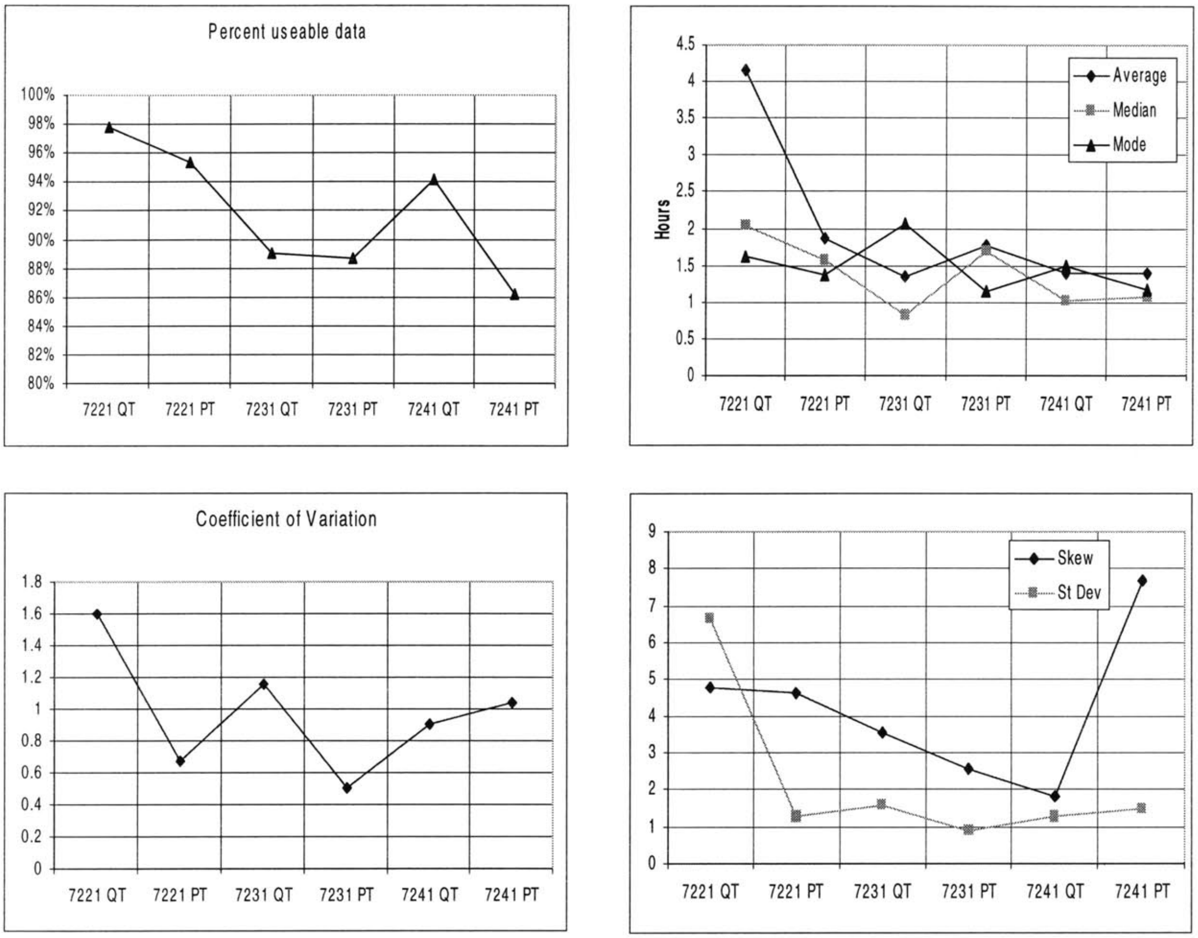


Figure A.4.1 - Factory D – Probability density and cumulative distributions.

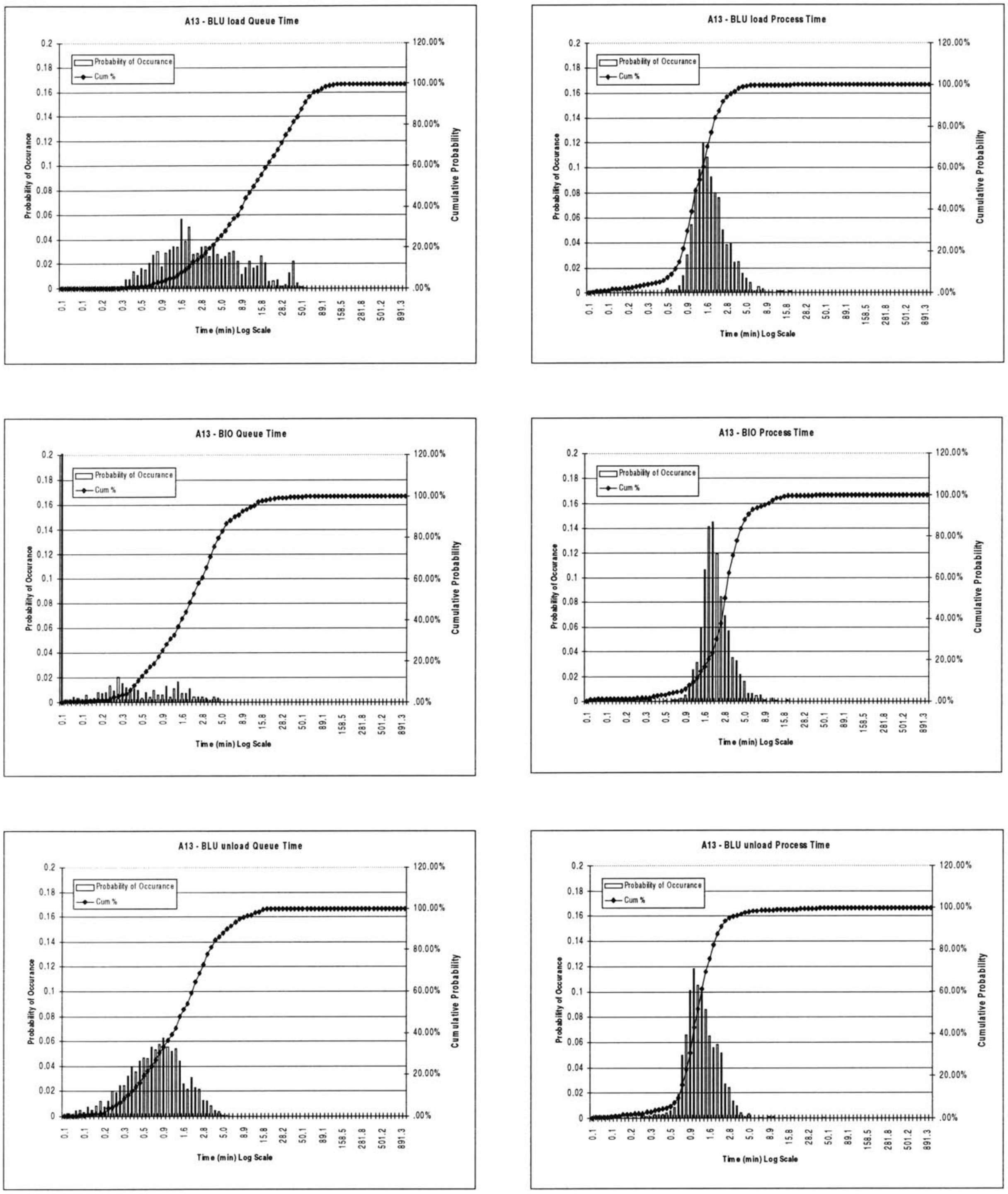


Figure A.4.2 - Factory D – Process-to-process comparison:

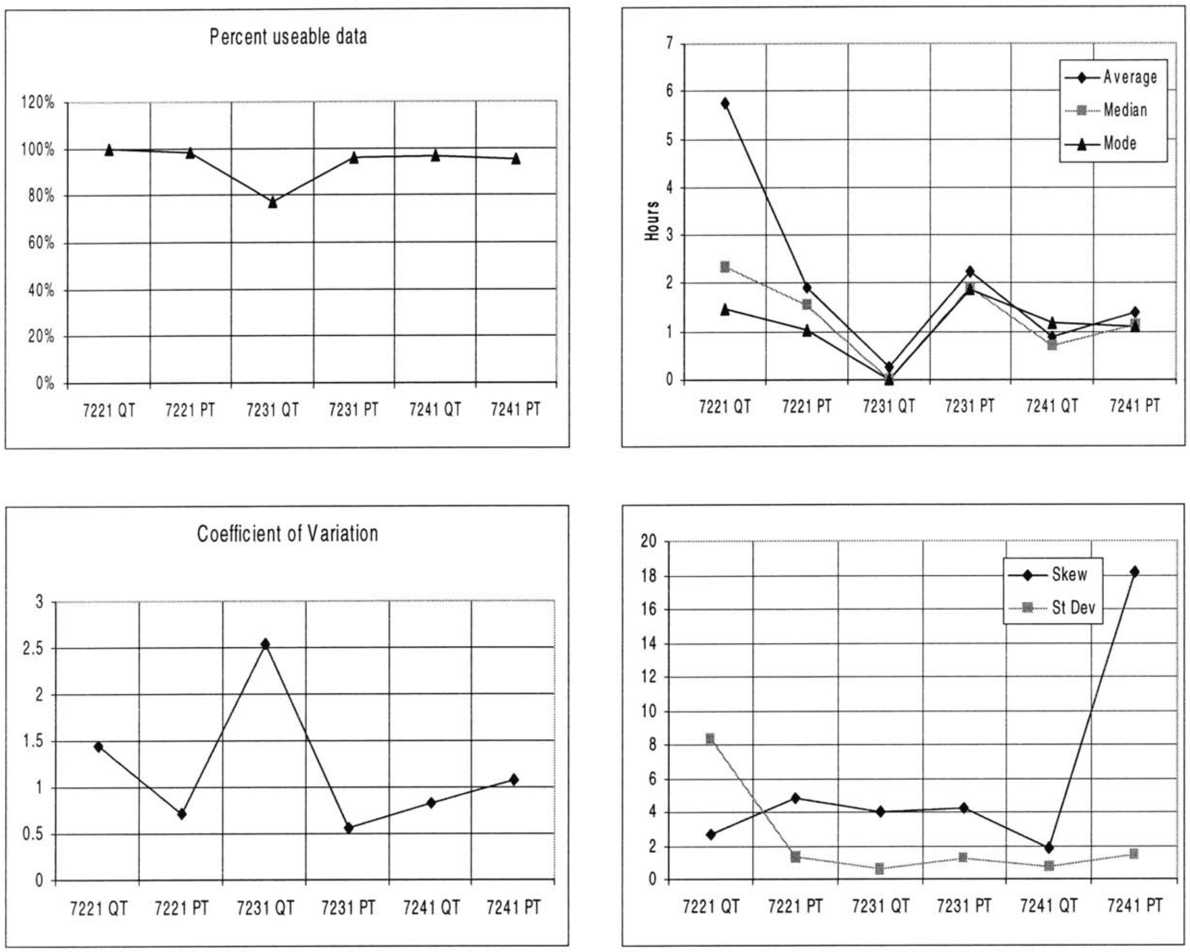


Figure A.5 - Factory-to-Factory comparison for 7221 QT:

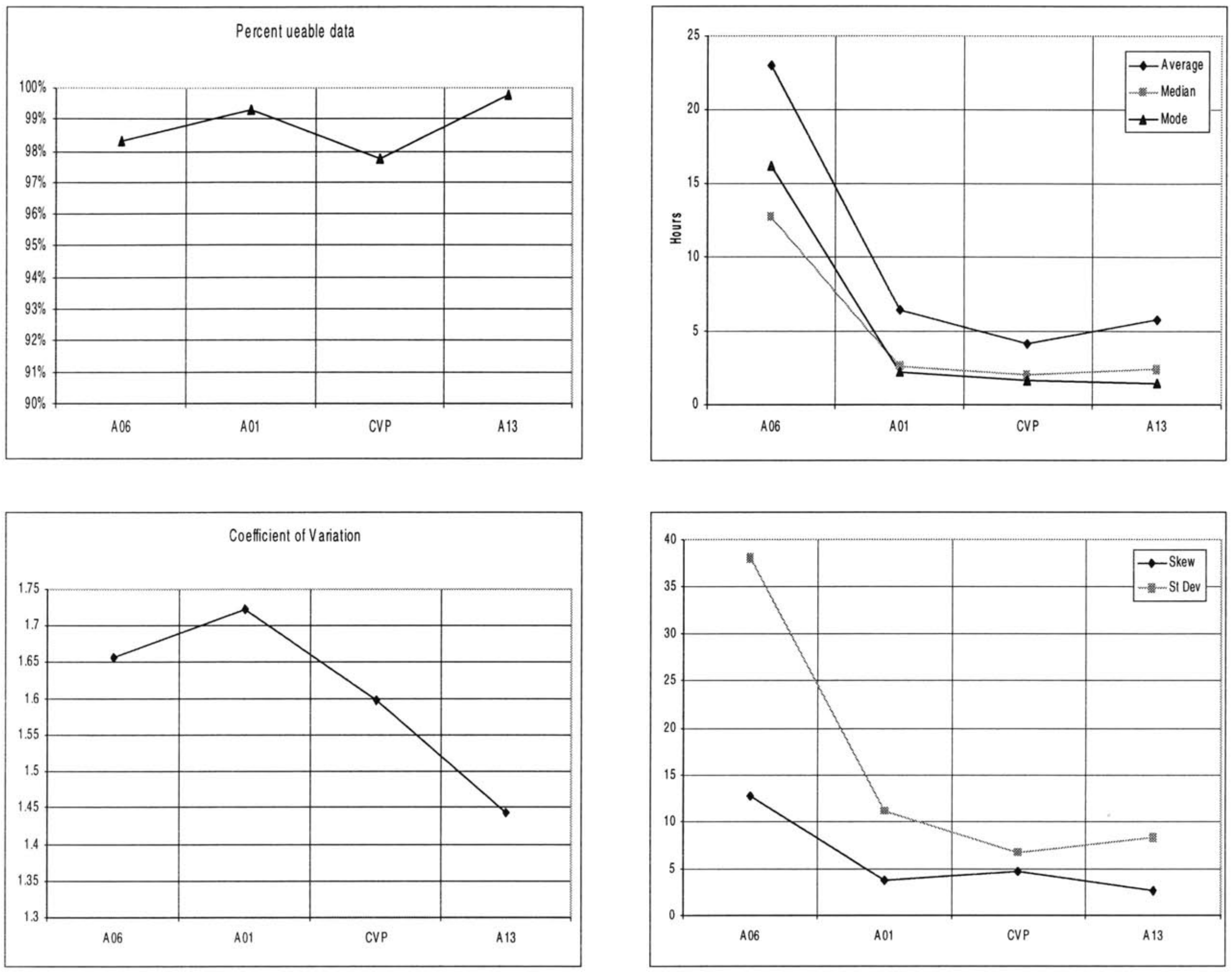


Figure A.6 - Factory-to-Factory comparison for 7221 PT:

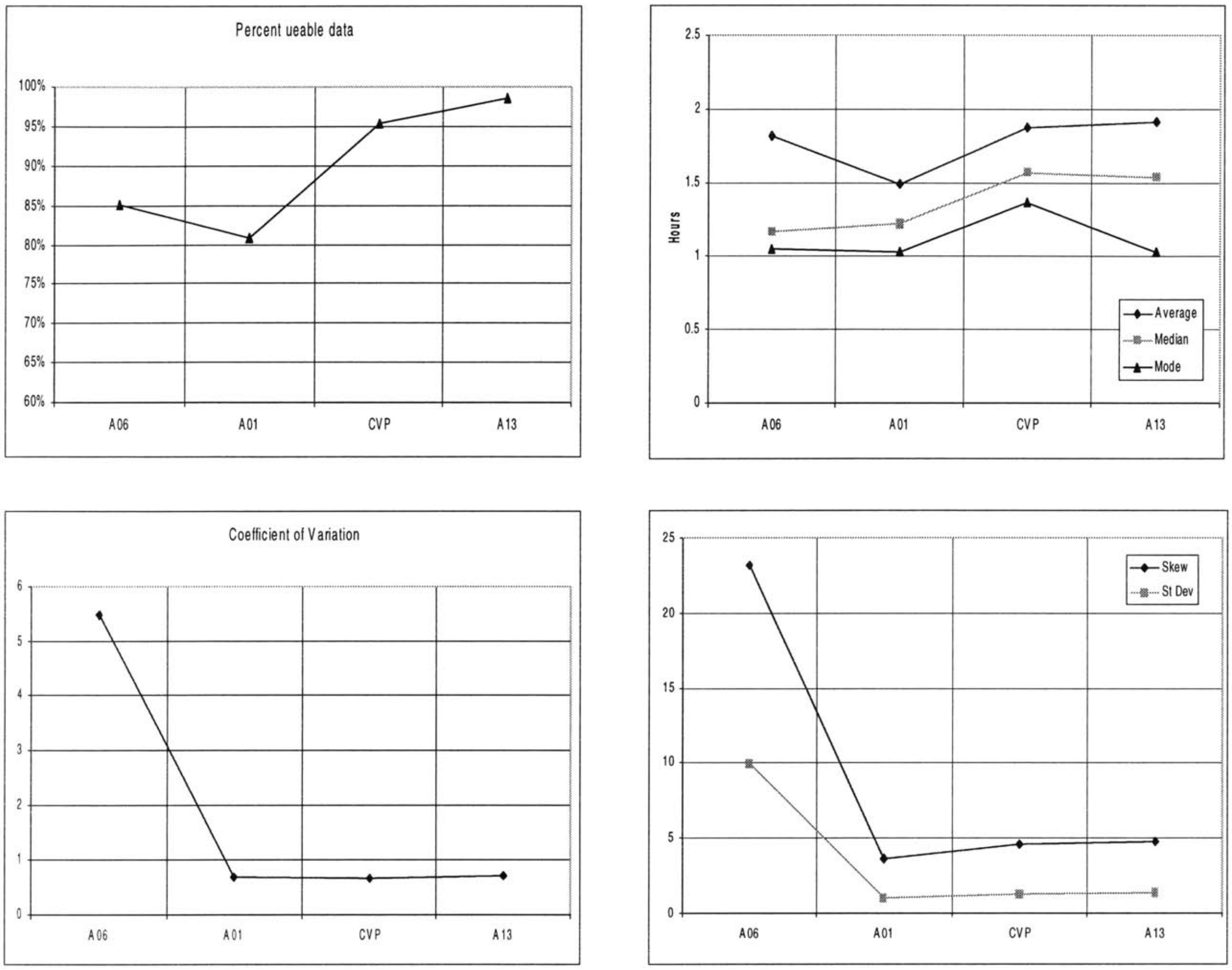


Figure A.7 - Factory-to-Factory comparison for 7231 QT:

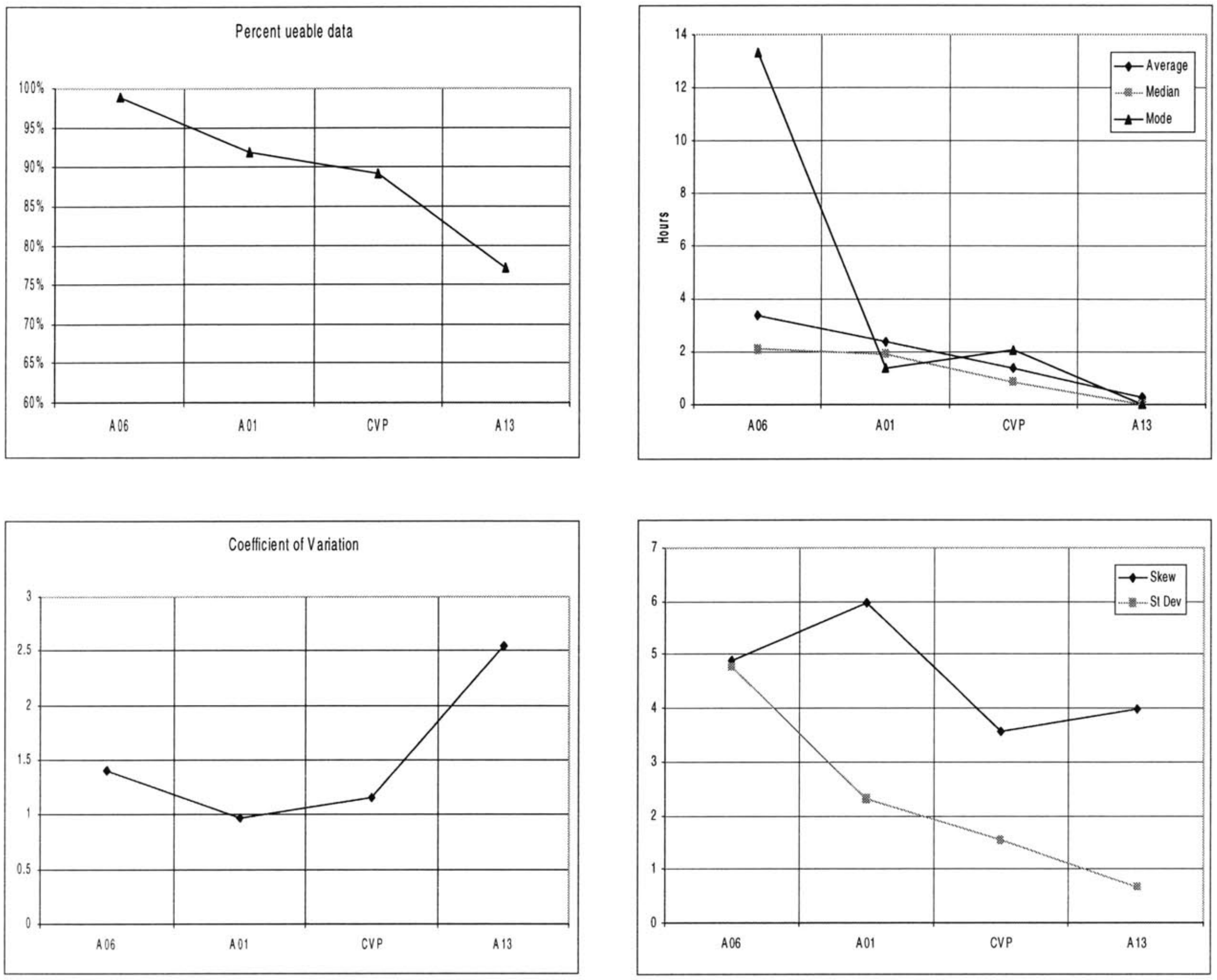


Figure A.8 - Factory-to-Factory comparison for 7231 PT:

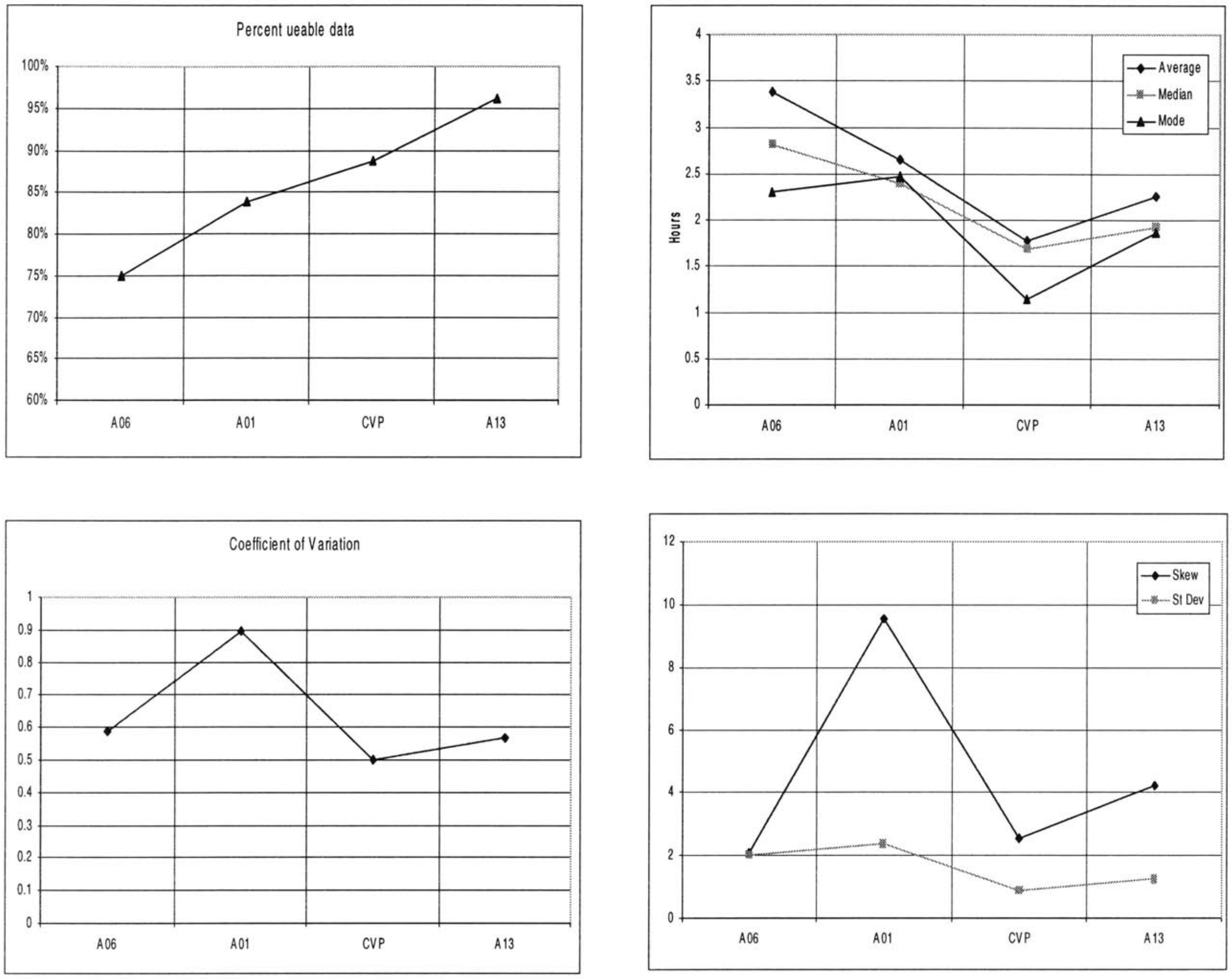


Figure A.9 - Factory-to-Factory comparison for 7241 QT:

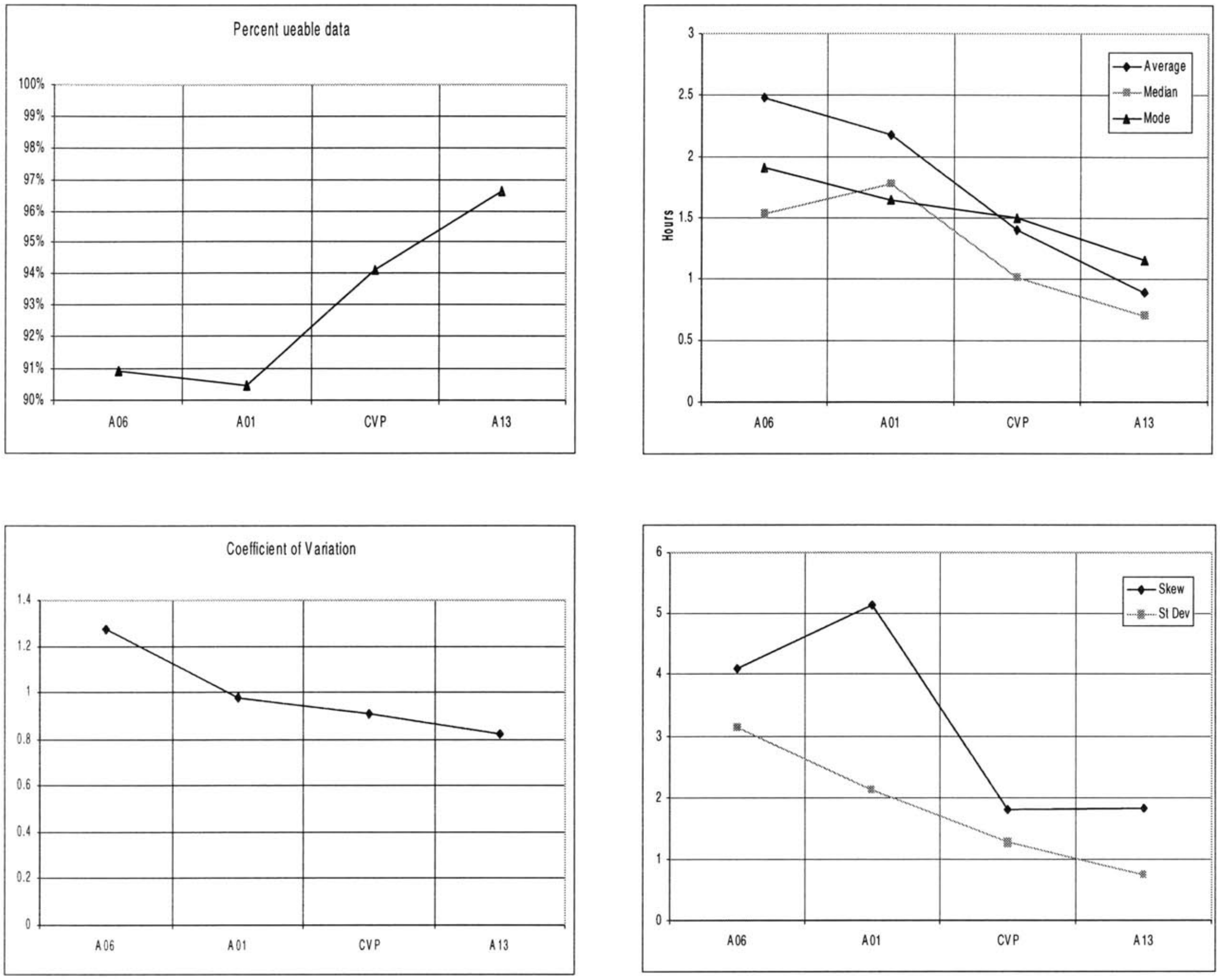
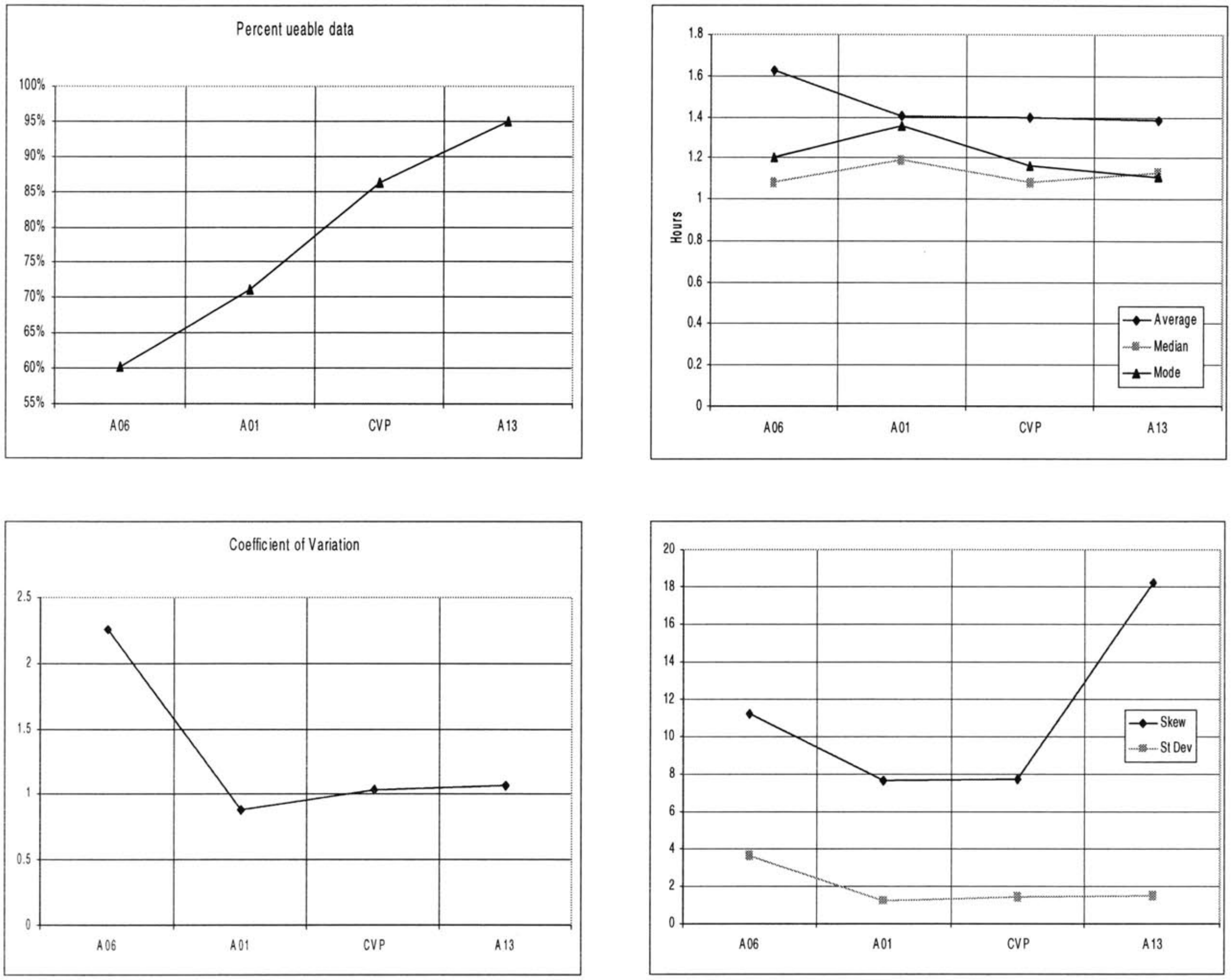


Figure A.10 - Factory-to-Factory comparison for 7241 PT:



Appendix B

Simulation details, Extend^{TM} Version 4.0

Figure B.1 – Entire Simulation

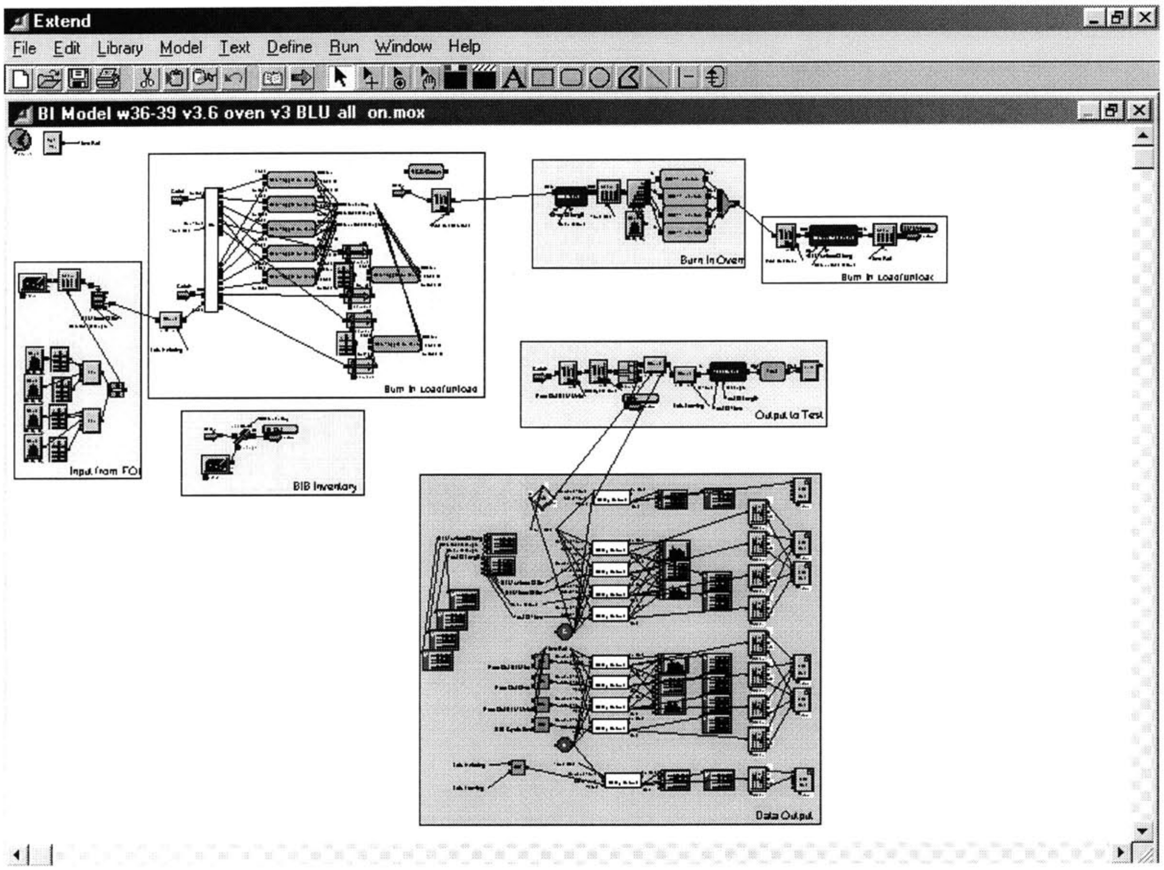


Figure B.2 – Arrival Rate Model

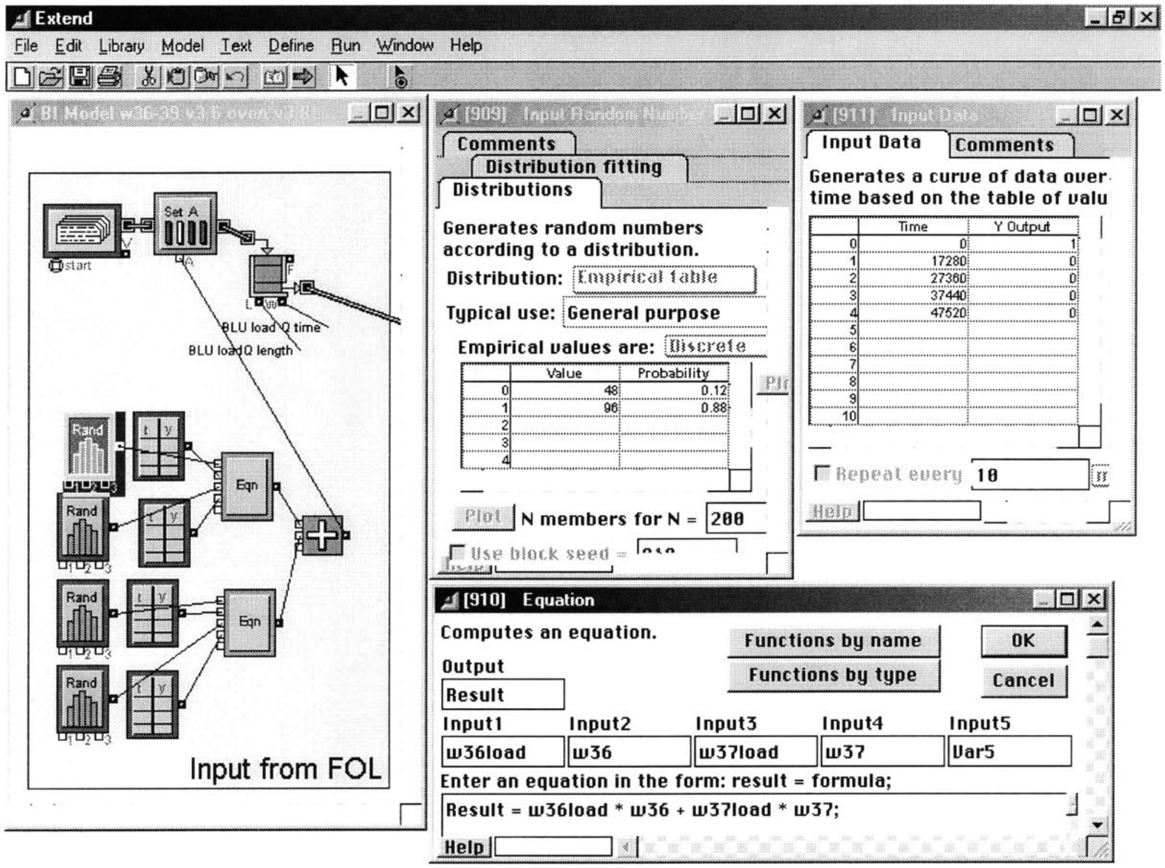


Figure B.3 – BIB Inventory Model

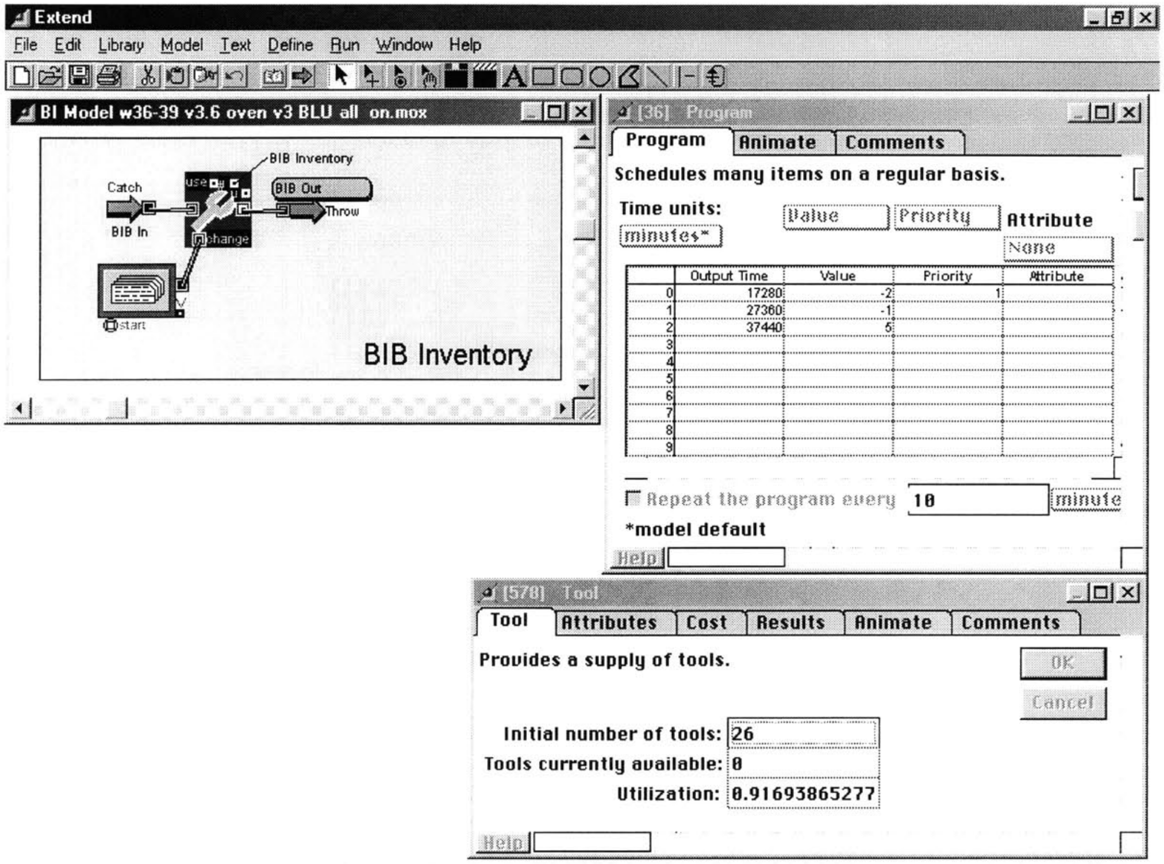


Figure B.4 – BLU Model

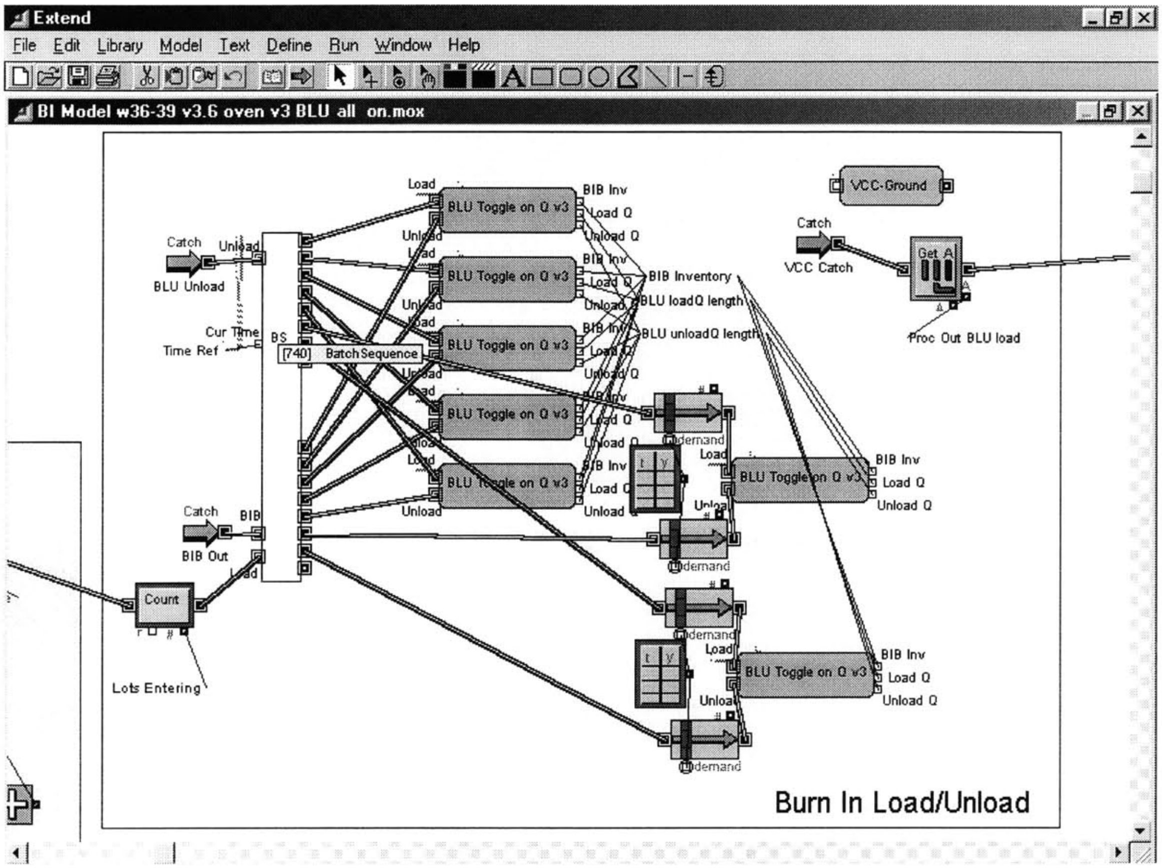


Figure B.5 – BLU Model Detail

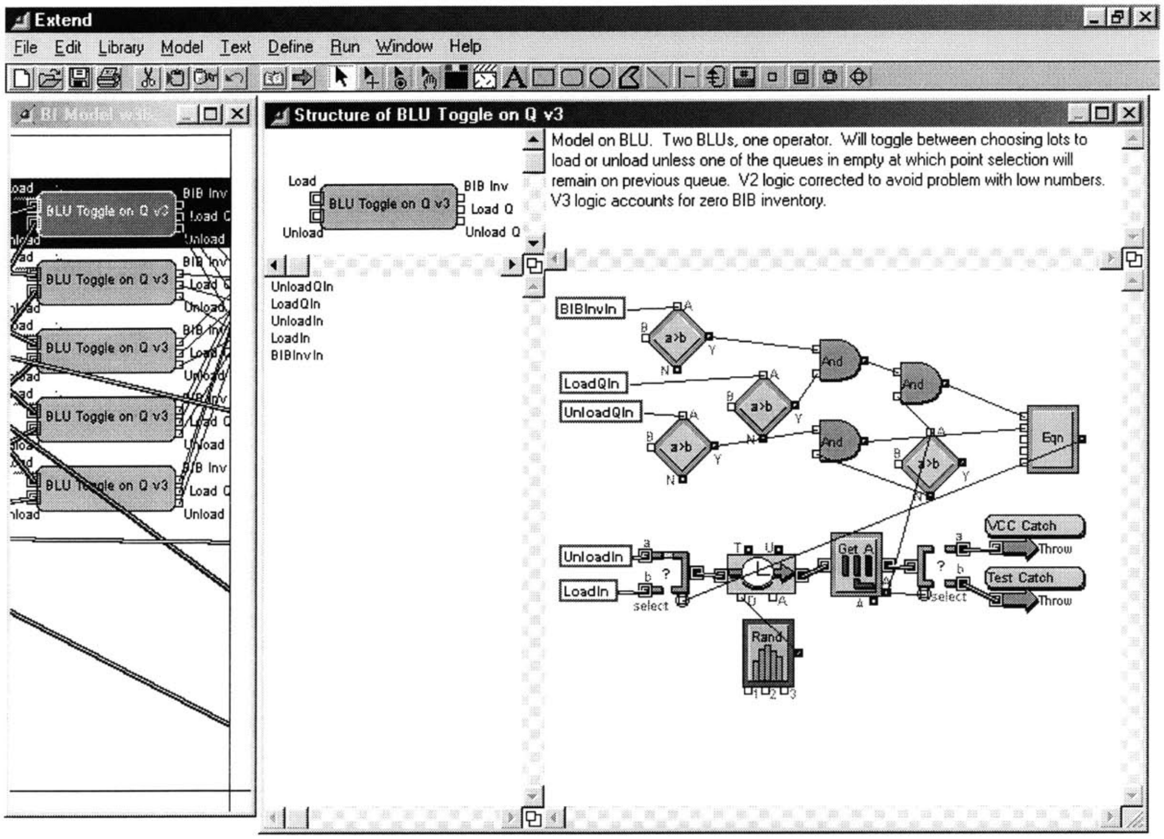


Figure B.6 – Batch and Sequencing Block Prior to BLU

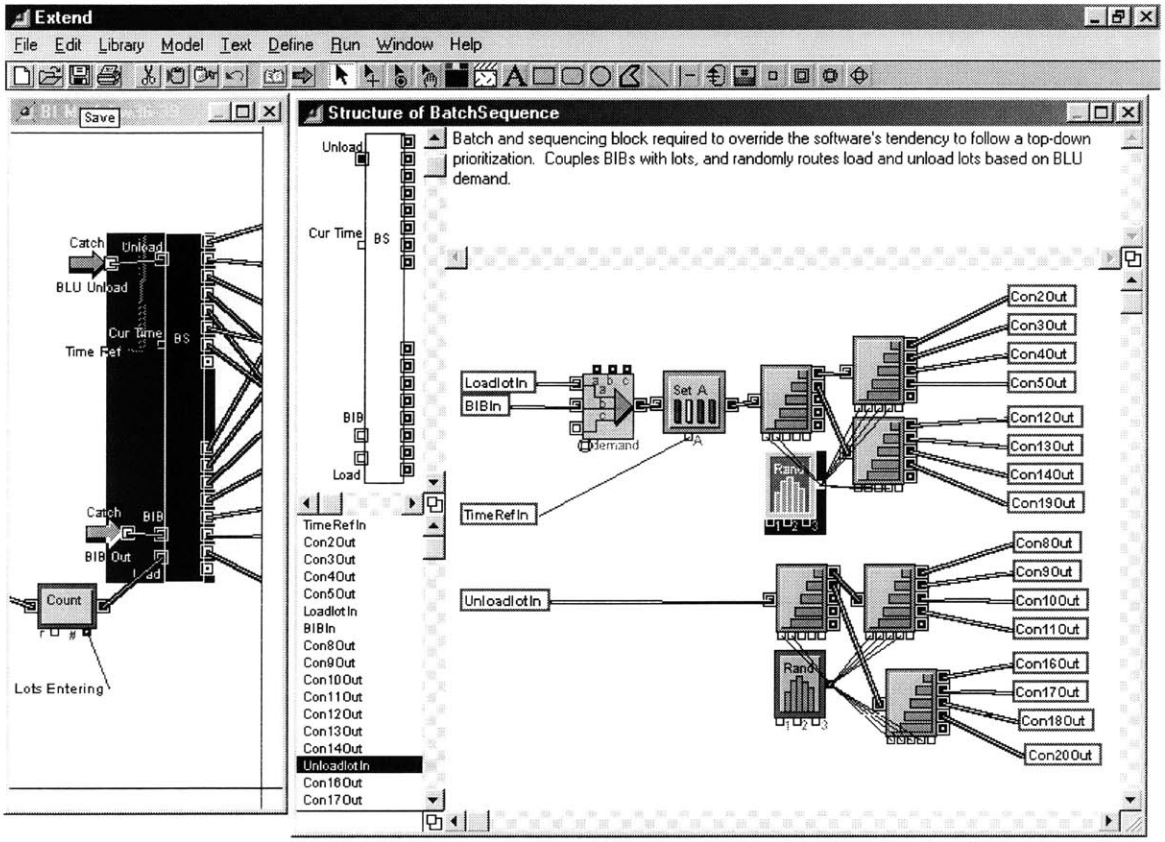


Figure B.7 – BLU Process Time Distribution

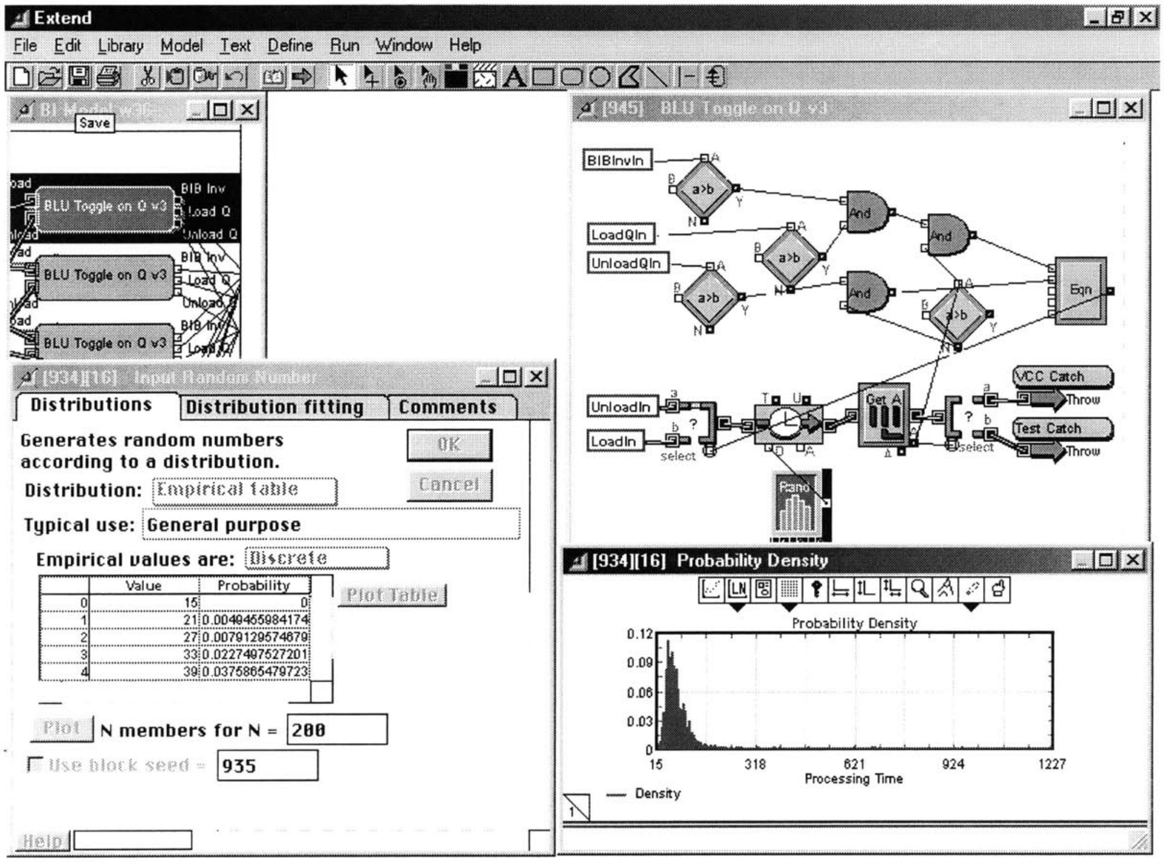


Figure B.8 – BIO Model

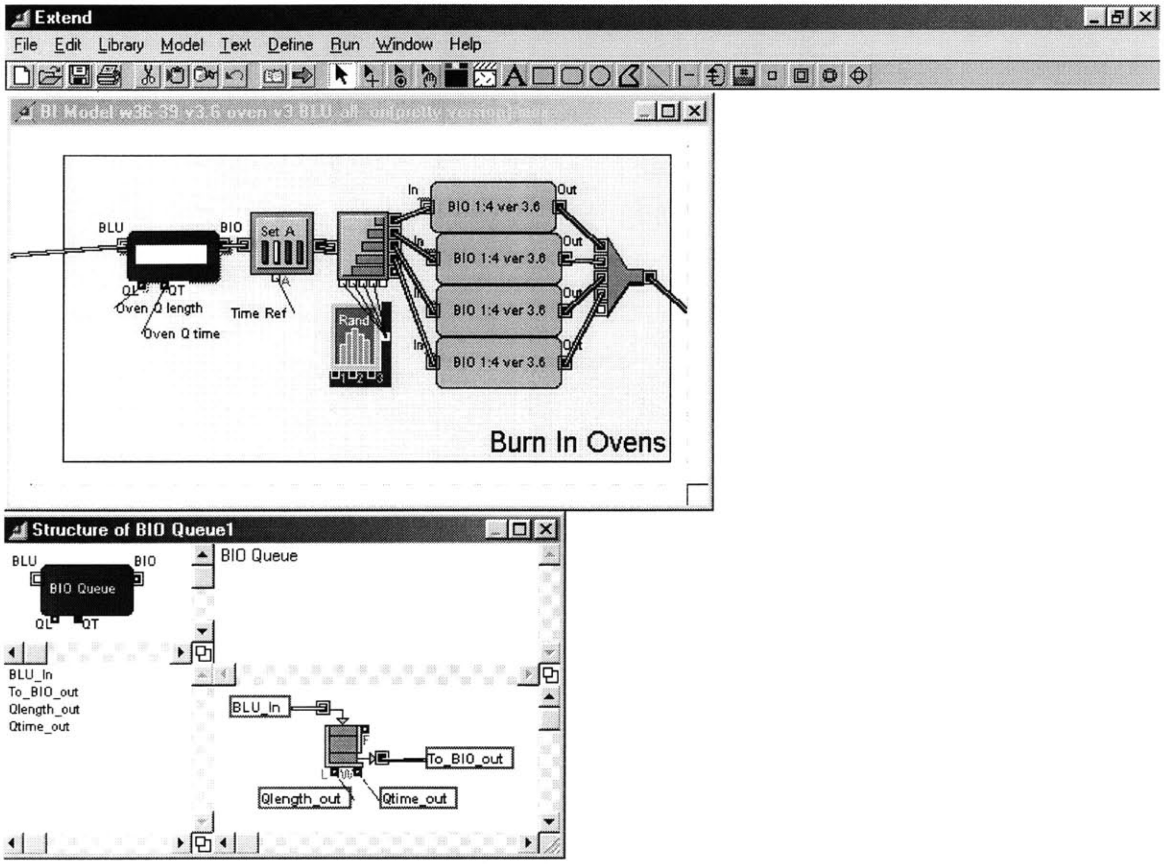


Figure B.9 – BIO Model Detail

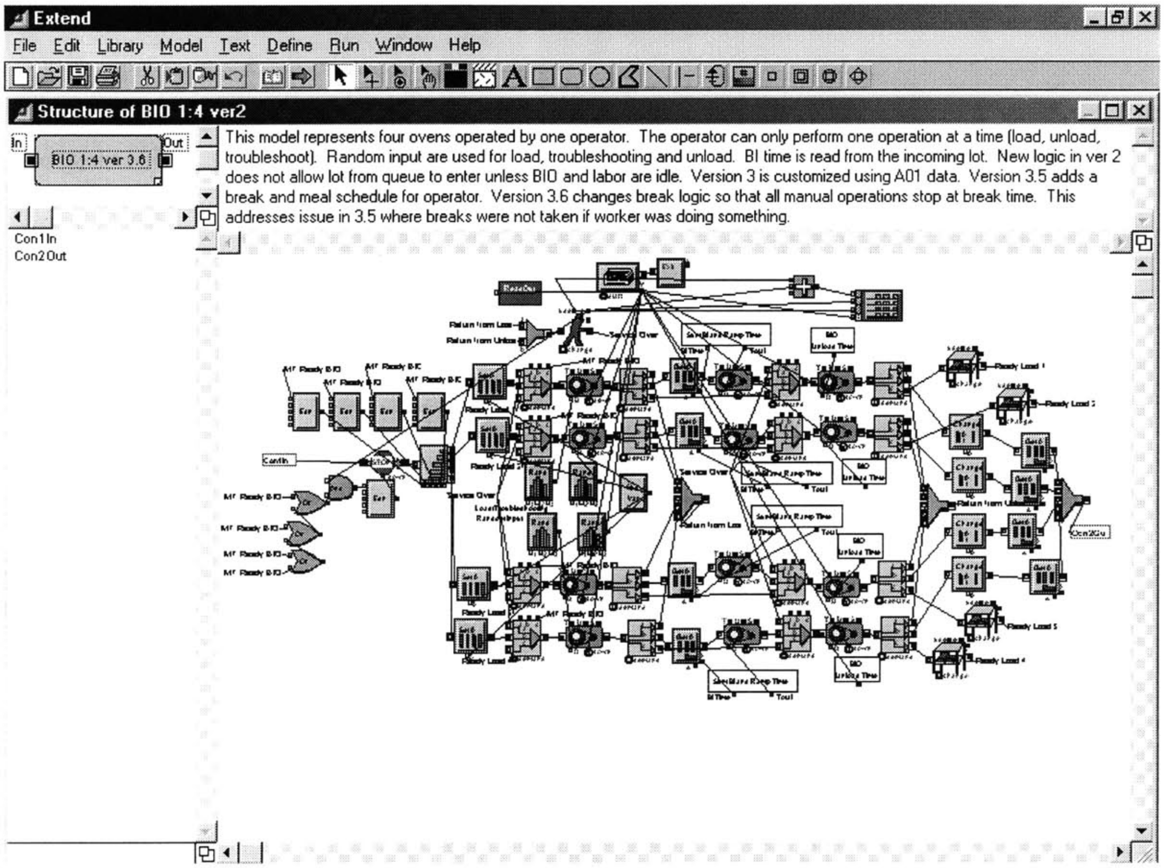


Figure B.10 – BIO Load and Troubleshooting Time Distribution

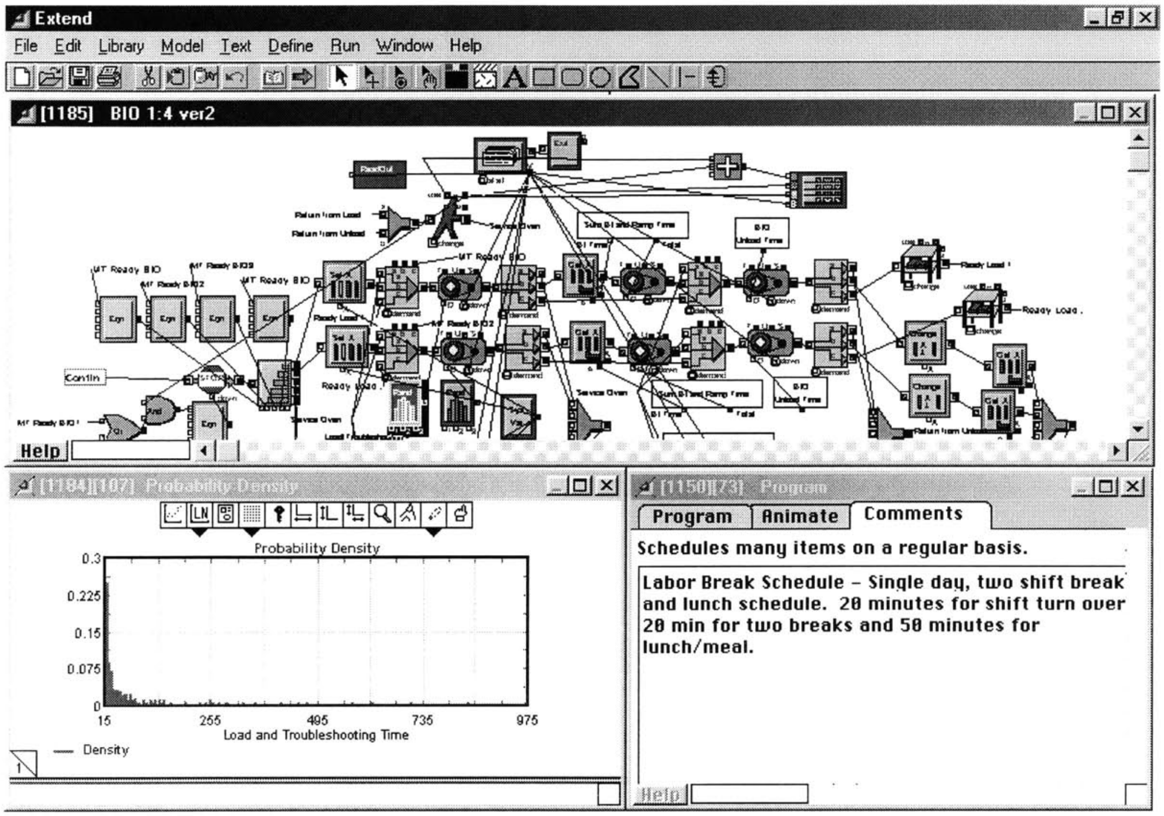


Figure B.11 – BIO Unload Process Time Model

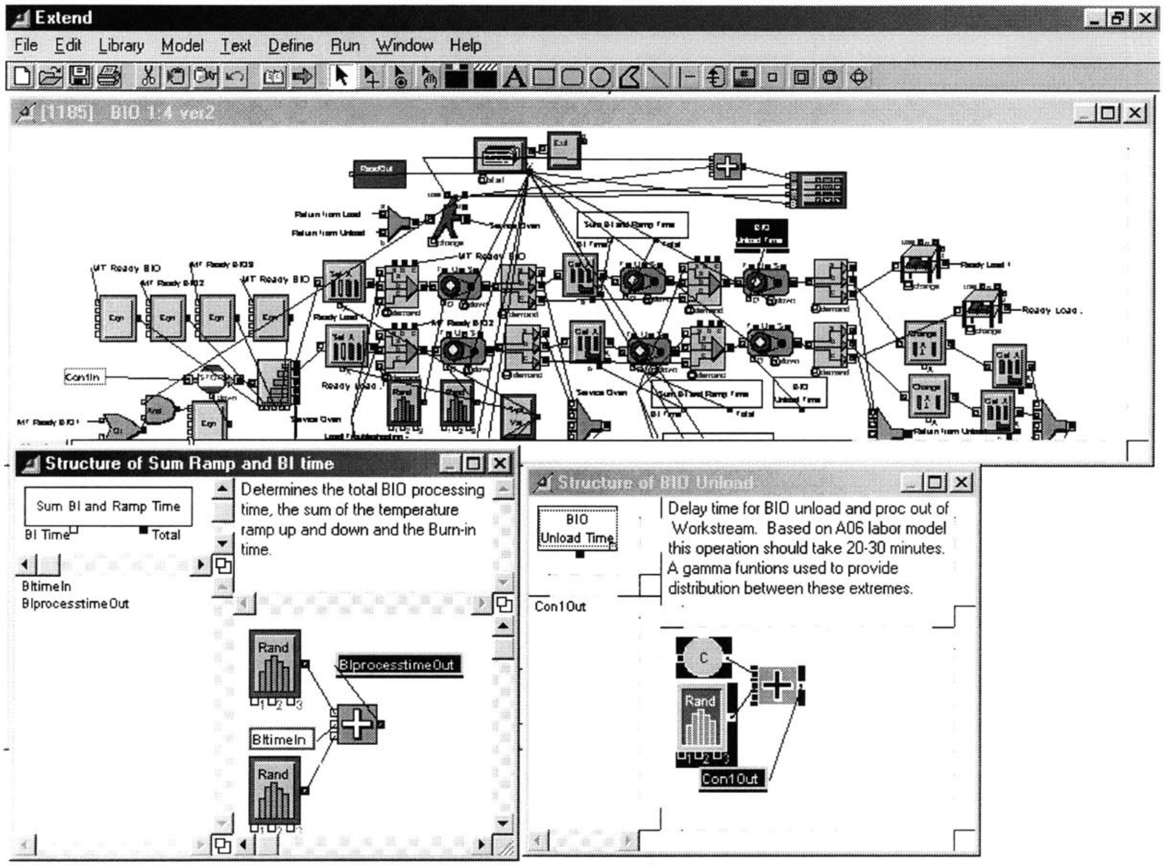


Figure B.12 – PBIC Model

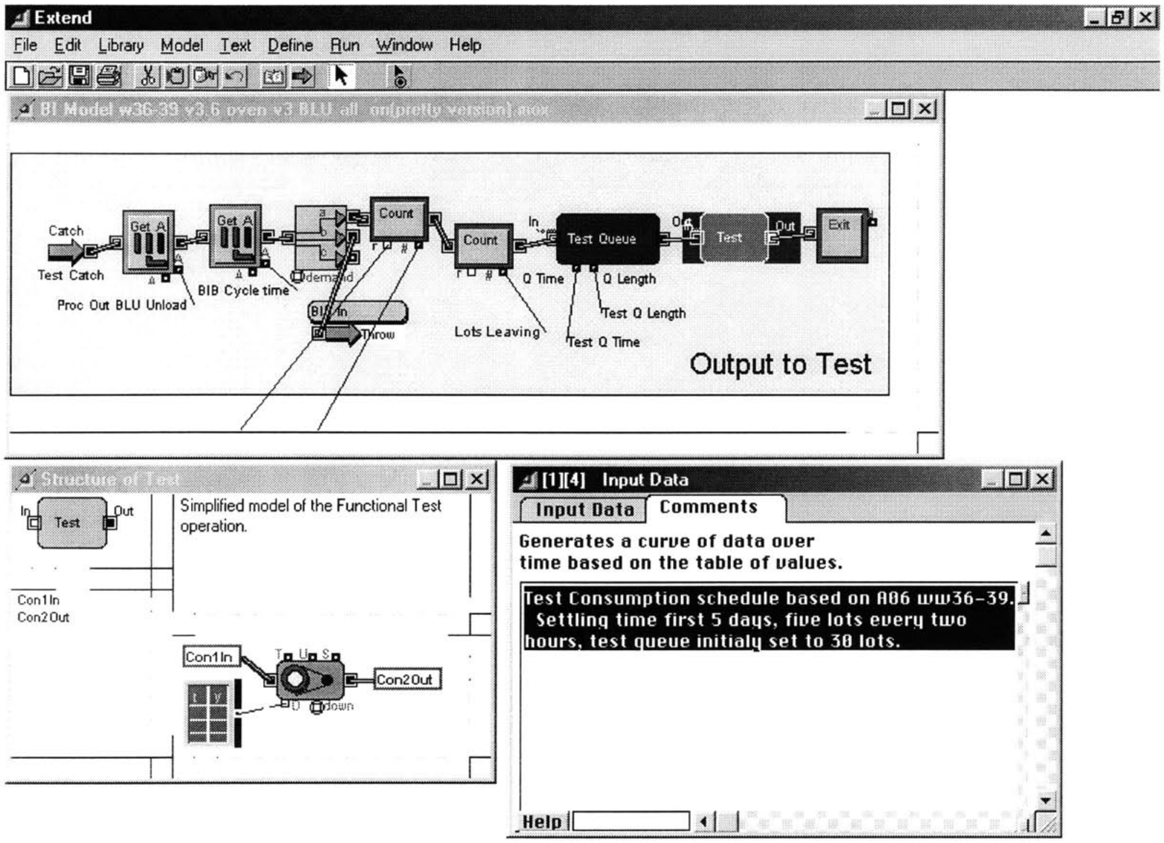


Figure B.13 – Data Output Handler

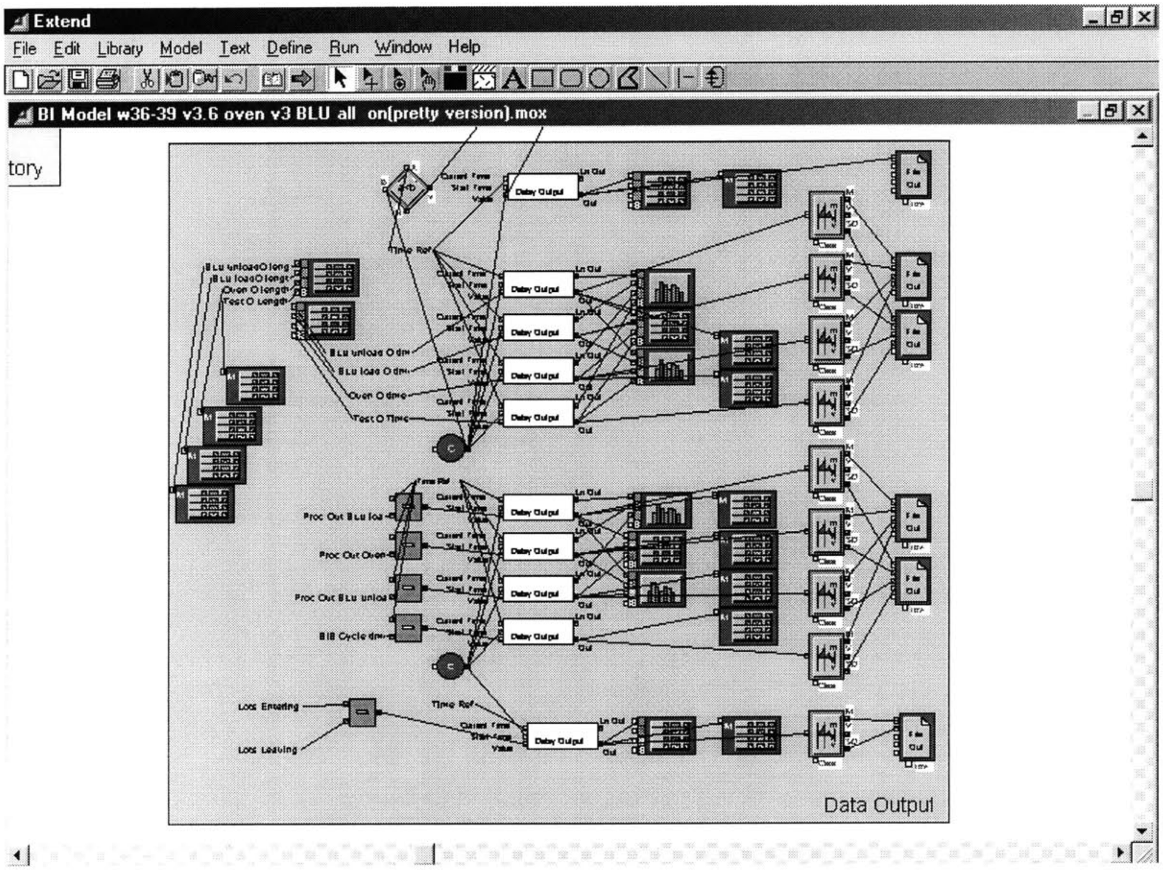
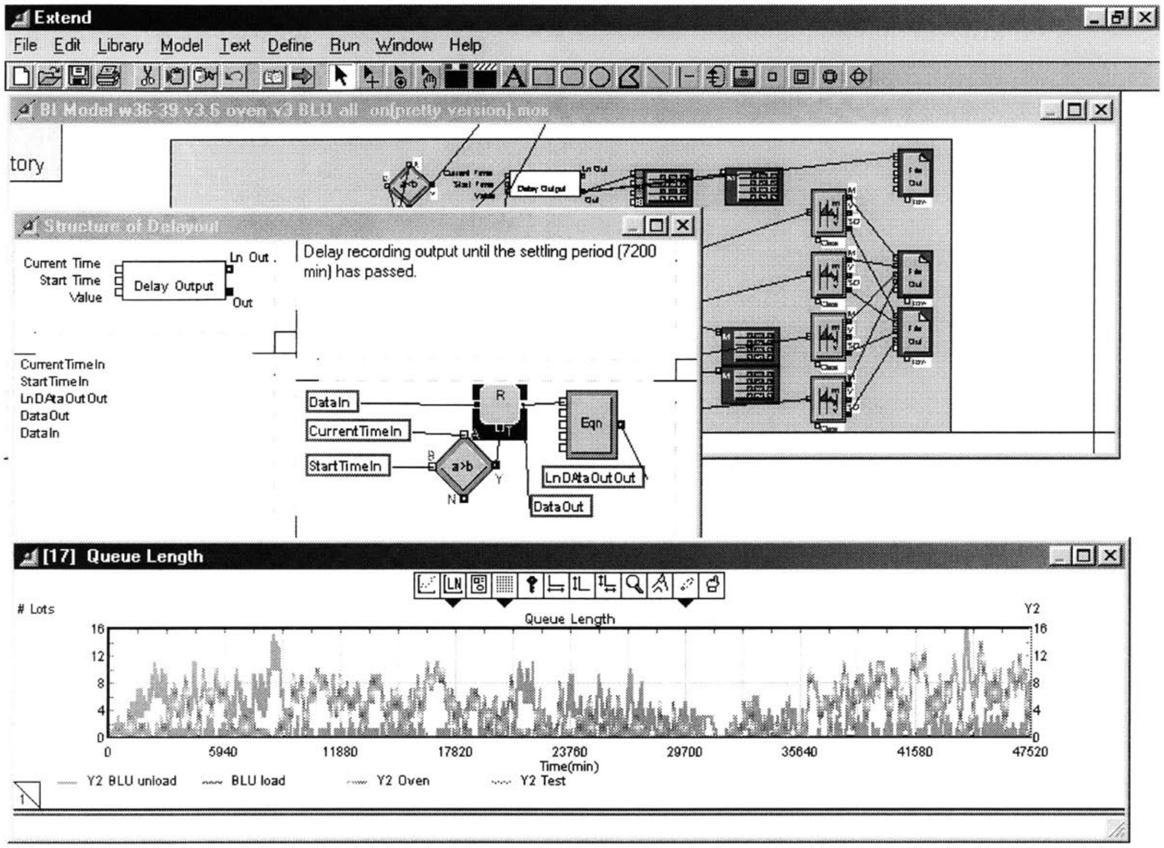


Figure B.14 – Delayed Output Block and Output Sample



Appendix C

This appendix catalogues the results of the simulation experiments. The experiments were conducted based on scenarios that were then compared with the baseline simulation. Figure C.1 shows the results of a series of experiments that were conducted in order to understand the impact of variation reduction on the maximum output of the BI area. Figure C.2 shows the results of a series of experiments while running at the baseline arrival rate. This series of experiments were designed to understand the impact of variation reduction on cost, given that ATM would like to keep extra capacity to buffer against demand variability. Accompanying both figures are descriptions of the scenarios.

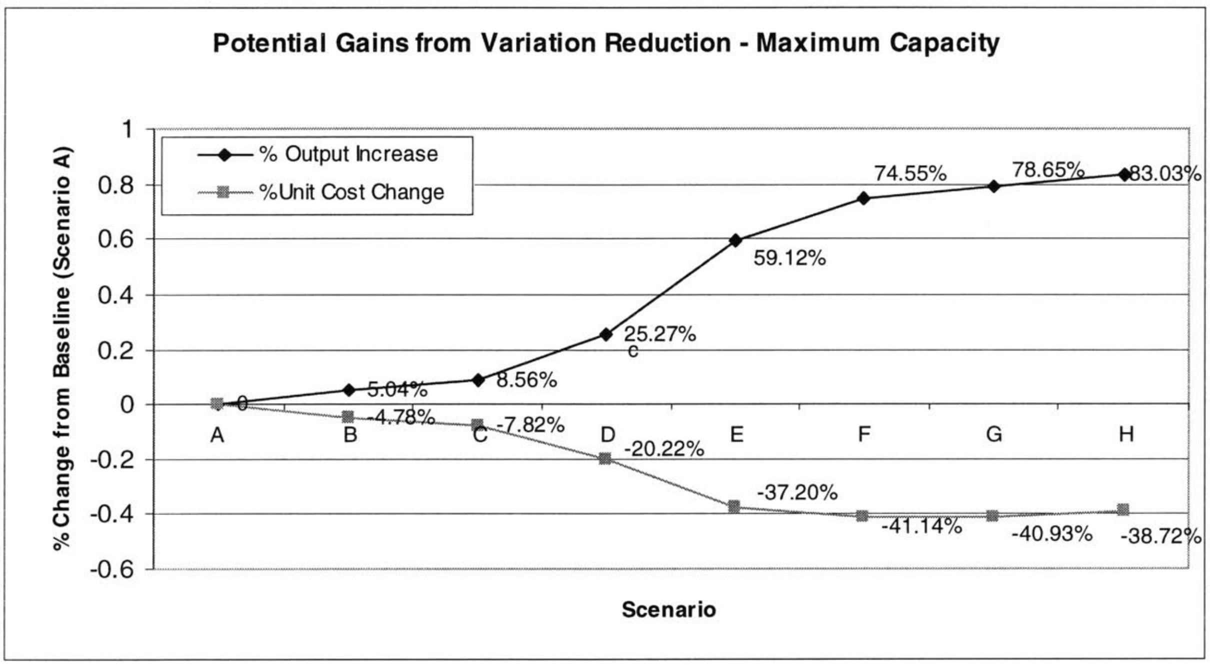


Figure C.1 – Maximum Capacity Experimental Scenarios

Scenario Description:

A. <u>Baseline WW36-39</u>: This is a model of the Costa Rica Factory during WW36-39 (September 2000). Only the FCPGA package is modeled. Lots are assumed to consist of 1000 units. The goal of this model was to capture the impact of variation on the BI process so validation focused on process and queue time distributions.

B. Baseline Potential Maximum: In studying the Baseline WW36-39 model it was found that the BI area was starved for WIP on occasion so to determine the potential maximum output of the existing system, the arrival rate from the front of line operations was increased so that the BI load queue was always full.

C. BLU Variation Reduction Potential Maximum: In this study the distribution used in the BLU model was modified so that in no instance did it take longer than 100 minutes to process a lot. The average processing time was 64.1 minutes with a three standard deviation range of 55.5 to 72.8 minutes. This processing time included the Vcc-Ground check as part of the BLU load sequence.

D. BIO Variation Reduction Potential Maximum: In this study the distribution used to model the manual load and troubleshooting operation prior to running the lot through the BI oven was modified so that in no instance did it take longer that 75 minutes. The average load and troubleshooting time was 27.7 minutes with a three standard deviation range of 23.5 to 31.8 minutes.

E. BLU and BIO Variation Reduction Potential Maximum: This study combined scenarios C and D.

F. BLU and BIO Variation Reduction with additional BLU with modified logic added: The result of scenario E showed that WIP levels in the BIO load queue were empty frequently and the WIP levels in the BLU Unload queue were consistently high indicating that there is insufficient BLU capacity to match the output from the BI ovens. An additional BLU and operator were added in this scenario with the operators decision making criteria set so that anytime the WIP in the BLU Unload queue was greater that 3, the operator would draw from the unload queue. If the unload queue was empty the BLU operator would load to feed the BIO queue.

G. Scenario F with an additional unit of BLU Capacity: An additional pair of BLUs and another operator are added.

H. BIB inventory is increased by the equivalent of two lots: Noting that BIB utilization was very high in Scenario G, BIB inventory was increase to keep BIBs from constraining output.

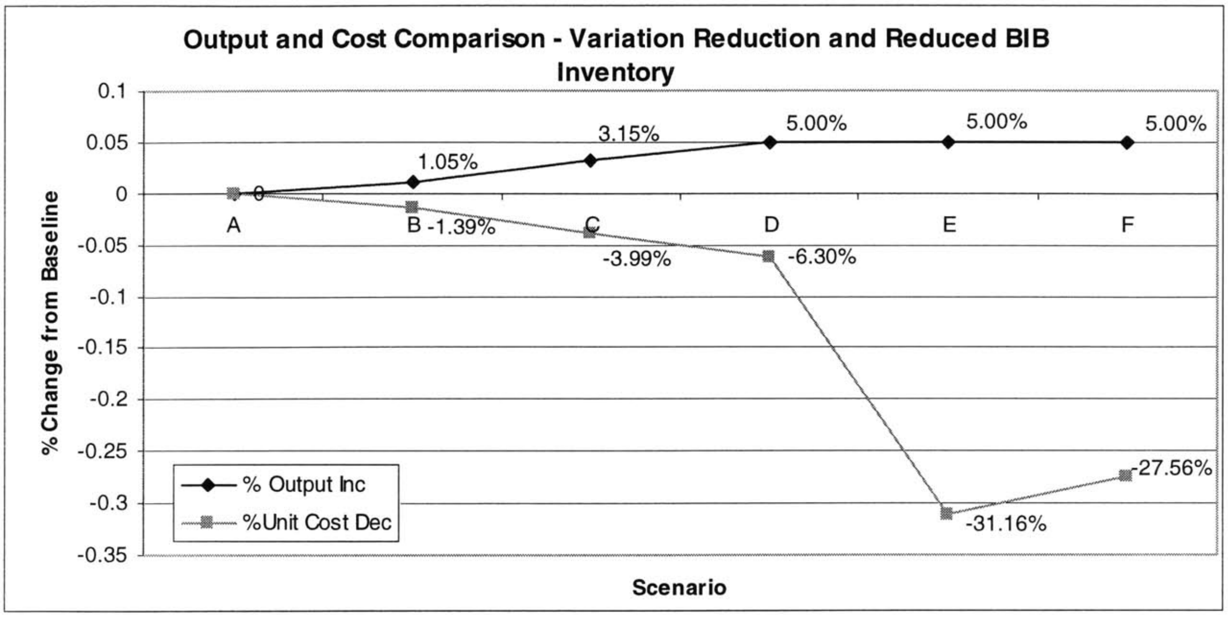


Figure C.2 – Baseline Arrival Rate Experimental Scenarios

Scenario Descriptions:

A. Baseline WW36-39: This is a model of the Costa Rica Factory during WW36-39 (September 2000). Only the FCPGA package is modeled. Lots are assumed to consist of 1000 units. The goal of this model was to capture the impact of variation on the BI process so validation focused on process and queue time distributions.

B. BLU Variation Reduction at Baseline Arrival Rate: In this study the distribution used in the BLU model was modified so that in no instance did it take longer than 100 minutes to process a lot. The average processing time was 64.1 minutes with a three standard deviation range of 55.5 to 72.8 minutes. This processing time included the Vcc-to-Ground check as part of the BLU load sequence.

C. BIO Variation Reduction at Baseline Arrival Rate: In this study the distribution used to model the manual load and troubleshooting operation prior to running the lot through the BI oven was modified so that in no instance did it take longer that 75 minutes. The average load and troubleshooting time was 27.7 minutes with a three standard deviation range of 23.5 to 31.8 minutes.

D. <u>BLU and BIO Variation at Baseline Arrival Rate</u>: This study combined scenarios B and C.

E. <u>BLU and BIO Variation Reduction with BIB Inventory adjusted</u>: The resultant WIP level in Scenario D dropped to 16.3 from the baseline case of 23.9. A BIB sensitivity study showed that the BIB inventory could be adjusted to the equivalent of 16 lots. In this scenario, the BIB inventory was set at 16 lots.

F. <u>BLU and BIO Variation Reduction with BIB Inventory adjusted, labor added</u>: This scenario looks at the impact of adding two operators to ensure that the variation reduction effort sustainable and reduced BIB inventory is managed effectively.

List of Reference

Brandimarte, P., and Villa, A., Modeling Manufacturing Systems. Springer, 1999.

Centeno, M. A. and Reyes, M. F., "So You Have Your Model: What to do Next, A Tutorial on Simulation Output Analysis". Proceedings of the 1998 Winter Simulation Conference, Washington, D.C.

Dieterich, C. A., "Using Simulation to Analyze Operational Policies in a Combined Development and Production Fabrication Facility". Leaders for Manufacturing Thesis (MIT), 1998.

Domaschke, J., et. al., "Effective Implementation Cycle Time Reduction Strategies for Semiconductor Back-End Manufacturing". Proceedings of the 1998 Winter Simulation Conference, Washington, D.C.

Farrelly, K. L., "Managing Variation in Near-Constraint Systems", Leaders for Manufacturing Thesis (MIT), 1998.

Goldratt, E. M., The Goal. North River Press, 1992.

Hopp, W. J., and Spearman, M. L., Factory Physics: Foundations of Manufacturing Management. McGraw-Hill, 1996.

Ishiwata, J., IE for the Shop Floor. Productivity Press, Inc, 1984

Ku, J., "Microprocessor Manufacturing Throughput Time Variability". Leaders for Manufacturing Thesis (MIT), 1994.

Newlin, A. W., Equipment Protective Capacity Optimization Using Discrete Event Simulation. Leaders for Manufacturing Thesis (MIT), 2000.

Nikoukaran, J., "Criteria for Simulation Software Evaluation". Proceedings of the 1998 Winter Simulation Conference, Washington D.C.

Perrin, S. E., "Production Management Policies in a Multiple Product Semiconductor Fabrication Facility". Leaders for Manufacturing Thesis (MIT), 1997.

Robinson, S., "Simulation Model Verification and Validation: Increasing the Users' Confidence". Proceedings of the 1997 Winter Simulation Conference,

Rodrigues Lopez, Y. A., Proyecto Para Aumentar la Capacidad Productiva en el Area de Burn in en la Empresa Intel de Costa Rica. Universidad Latino Americana de Ciencia y Technologia, 2000.

Sargent, R.G., "Verification and Validation of Simulation Models". Proceedings of the 1998 Winter Simulation Conference, Washington D.C.

Shannon, R. E., "Introduction to the Art and Science of Simulation". Proceedings of the 1998 Winter Simulation Conference, Washington D.C.