ABSTRACT

Title of dissertation: EVALUATION OF SETUP ECONOMIES IN CELLULAR MANUFACTURING

Steven Boyd Kramer, Doctor of Philosophy, 2004

Dissertation directed by: Professor Arjang A. Assad
Decision and Information Technologies
The Robert H. Smith School of Business

This dissertation addresses two research questions relating to the role of setups in discrete parts manufacturing. The first research topic uses a carefully designed simulation study to investigate the role of setup economies in the factory-wide conversion of functional layouts (job shops) to cellular manufacturing. The model-based literature shows a wide dispersion in the relative performance of cellular manufacturing systems as compared to the original job-shop configurations, even when the key performance measure is flow time and the assessment tool used is simulation. Using a standardized framework for comparison, we show how this dispersion can be reduced and consistent results can be obtained as to when the conversion of the job shop is advantageous.

The proposed framework standardizes the parameters and operational rules to permit meaningful comparison across different manufacturing environments, while retaining differences in part mix and demand characteristics. We apply this framework to a test bed of six problems extracted from the literature and use the results to assess the effect of two key factors: setup reduction and the overall shop

load (demand placed on the available capacity). We also show that the use of transfer batches constitutes an independent improvement lever for reducing flow time across all data sets. Finally, we utilize the same simulation study framework to investigate the benefits of partial transformation, where only a portion of the job shop is converted to cells to work alongside a remainder shop.

The second research question examines the role of dispatching rules in the reduction of setups. We use queueing models to investigate the extent of setup reduction analytically. We single out the Alternating Priority (AP) rule since it is designed to minimize the incidence of setups for a two-class system. We investigate the extent of setup reductions by comparing AP with the First-Come-First-Served (FCFS) rule. New results are obtained analytically for the case of zero setup times and extended to the case of non-zero setup time through computational studies.

EVALUATION OF SETUP ECONOMIES IN CELLULAR MANUFACTURING

by

Steven Boyd Kramer

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland at College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2004

Advisory Committee:

Professor Arjang A. Assad, Chair/Advisor
Professor Shapour Azarm
Professor Michael C. Fu
Professor Bharat K. Kaku
Professor Gilvan C. Souza

ACKNOWLEDGEMENTS

I am deeply indebted to Dr. Arjang Assad for his patience, assistance, and concern for scholarly research. I also appreciate the guidance of Dr. Bharat Kaku especially in the formative stage of my research on cellular manufacturing systems. I would also like to acknowledge the balance of my committee, Professor Shapour Azarm, Professor Michael Fu and Professor Gilvan Souza who constructively guided my direction at my proposal toward the addition of analytic contributions and provided valuable feedback at my defense.

I could not have completed this dissertation without the loving support of my family and friends. I owe everything to my wife, Susan, daughters, Shannon and Larkin and son, Keenan. I dedicate this work to them.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1

INTRODUCTION

1.1 The Manufacturing Environment

Buzacott and Shantikumar (1993, p.1) describe a manufacturing system as a system consisting of “machines and work stations where operations such as machining, forming, assembly, inspection, and testing are carried out on parts, items, subassemblies, and assemblies, to create products that can be delivered to customers.” In *discrete part manufacturing systems*, each item processed is distinct, although the processing may take place in batches or distinct packets. The batches are then used as transfer units between manufacturing areas. This is in contrast to chemical industries where the processed material may be in the form of continuous fluid. Discrete parts manufacturing systems arise commonly in “mechanical, electrical, and electronics industries making products such as cars, refrigerators, electric generators, or computers .” (p.1).

As an example, we examine a process designed to create a hole through a block of metal as illustrated in Figure 1-1. The process may require a single operation (single-*stage*) as in a drill press drilling a hole, or may require multiple operations linked together (multi-stage) if the completed hole requires further finishing such as the addition of a champher and de-burring.

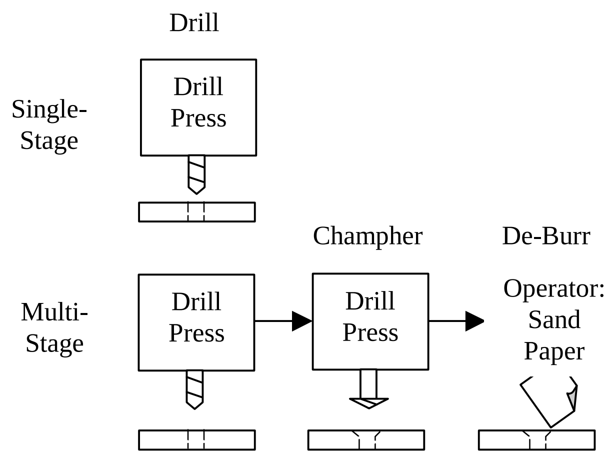

Figure 1-1. Single- versus Multi-Stage Processing.

Multi-stage processes may include internal buffer storage in order to account for variations in the time between successive outputs of product at a process of each process step allowing each to work more independently of the other. Single- or multi-stage processes may be linked together to provide a variety of processing capabilities. The time between successive outputs of a multi-stage process is usually regulated by the dynamics of the flow of parts under congestion and may depend crucially on *bottleneck* stages that limit the capacity of the overall process. The flow time of a job is the amount of time a job spends in the system. Specifically, it is the time from when a job consisting of demand for a certain batch size of a given product is introduced into the manufacturing facility at the location of its first operation to when the last operation required on the batch of product has been completed. It includes the time waiting for processing and material transfer between operations, setup times if required, as well as the time the batch is being serviced by machines. A bottleneck process adversely affects the flow time of all parts using that process.

One can characterize manufacturing processes based on the way the process flow is coordinated. A process can be synchronous or asynchronous. Synchronous processes have a fixed process rate where all work moves at the same rate through the processing steps in sequence. This is either done continuously, as in automobile painting operations using continuous conveyors moving at a fixed rate, or discretely, as in spot welding operations of chassis where automobiles move in and out of robot welding stations at regular intervals. Synchronous flows eliminate most of the need for storage between machines, but require tight coordination of customer orders, material supply and extremely high process quality. Asynchronous processes are much more common, where work is moved to its next process step when the current step is completed. Work, since not synchronized, is then staged in an "input" queue and waits as required for its turn at the next operation.

A key concern of the work presented in this dissertation is the few factory layout structures used to organize the material flow. The most common type and the one that naturally aligns with high part variety is the *job shop*. The term "job shop" (abbreviated as JS throughout this work) refers to a manufacturing facility comprised of general-purpose machines organized into a collection of machine centers or *departments* grouped on the basis of the operation performed (turning, drilling, milling, etc.). By providing the appropriate machine types, a small number of machine departments is sufficient in the factory to accomplish a high variety of part processing. These machine types can be applied in various sequences to produce a wide variety of parts. The job shop structure supports a high variety of jobs.

Typically, job shops are designed to handle small production batches of custom products requiring a variety of processing requirements. Accordingly, the equipment is organized by function as the same general type of operation may be performed by a number of machines in a wide variety of different ways. For example, when a hole is needed in a piece of metal, it is sent to the drilling department where a variety of machines from drill presses to mills to boring machines may reside. We will consider more details of operations of the job shop below.

Assembly lines (or flow lines) are structures where process equipment is organized in the order specified by their operations. This organizational principle is also known as a product focus. Assembly lines minimize material handling since the next machine needed is in immediate physical proximity. Material handling automation is commonly employed between process steps to retain part registration, minimizing setups and reducing labor. This type of structure is biased to the direction of part flow, so backtracking, where processes must travel opposite the direction of the standard flow in order to get access to a particular type of machine, is difficult and very disruptive.

Current industry trends encourage managers to focus their factories to provide products and services at high quality and low cost. A challenge in discrete parts manufacturing is to provide customized products to meet individual tastes while depending on the stability of common processes and equipment (Pine, 1993, p.7). Factories using general-purpose machines are capable of producing a large variety of parts by the nature of their process equipment. However, frequent tooling changeovers are required on general-purpose machines to account for part variety that can be time-consuming and expensive. Below, we outline some of the benefits of an alternative approach, which we call a *cell* shop.

1.1.1 Job Shop. In a job shop, a large fraction of the flow time of a given part is due to wait times. Parts often have to queue up to await their turn at a given machine or machine center due to limited capacity, wait for material handling devices for transport to or from a process or wait to join parts being processed in other parts of the factory. The machines typically require setups due to changeovers between operations in order to accommodate different part and processing requirements. The machines in each department share a common queue of incoming work and the length of this queue accounts for most of the delay at each machine center. If jobs are assigned at random, the larger the variety of parts types, the more likely it is for setups to be incurred. Increasing the frequency of setups increases the amount of time required to complete each job (expected setup plus run time). This increases the time spent at the machine for each job, and leads to longer queues and wait times. This relationship is apparent in the familiar M/M/1 queue, where the wait in the queue, W_q, is a function of the arrival rate and mean service time (λ and μ^{-1}, respectively) and machine utilization, which is represented in this case by $\rho = \lambda / \mu$: $W_q = \rho / \mu(1 - \rho)$. In this dissertation we will consistently associate the wait in the queue with the time from when a customer arrives in the system until service commences on that customer. We, therefore, imbed any required setup time in this queue wait. The batch flow time is measured from part batch introduction into the factory (from receiving) to part batch leaving the factory (sent to shipping).

Material handling also contributes to the flow time of parts in the job shop and wait times for material handling resources. Parts travel from department to department to complete their operation sequences traversing the factory. Factory and department size, part sensitivity, and sequence lengths all exacerbate move times.

1.1.2 Manufacturing Cells. A manufacturing *cell* is a collection of dissimilar machines positioned in proximity to work on products of similar shapes and processing like a production line (Chase, Jacobs and Aquilano, 2004, p.200). We assume that the nature of manufacturing demands and processing required is similar to what is found in a job shop. In cell-based production, otherwise know as a cellular manufacturing system (CMS), parts with similar features use common sequences of operations and similar tools or fixtures. A group of such related parts defines a part *family*. A CMS is therefore closely allied to the concept of group technology: the concept of grouping similar parts into part families to benefit design and manufacturing (see Askin and Vakharia, 1990).

In their recent comprehensive monograph on cellular manufacturing, Hyer and Wemmerlöv (2002, p.18) define a cell using the concept of families:

> A cell is a group of closely located workstations where multiple, sequential operations are performed on one or more families of similar raw materials, parts, components, products or information carriers.

Typically, a number of different part families occur in the product mix. One of the challenges in CMS is developing rules for cell formation to associate the part family data with the required machines (see for example Singh and Rajamani, 1996 for a review of the cell formation literature).

The two most basic benefits of cellular manufacturing according to Hyer and Wemmerlöv (2002, p. 48) are reductions in flow time (due to use of smaller batch sizes and use of shared tools and fixtures) and inventory (due to the proximity of equipment). Other benefits of cellular manufacturing according to Chase et al. (2004, p. 200) are better human relations due to small work clusters, and improved operator expertise due to learning through repetitions. Other advantages according to the literature are improved quality and easier control of operations. Physically moving both machines and associated product family to a cell enables the factory to focus on that product family. The part family in the cell enjoys unfettered access to a limited set of resources that are now in proximity to each other aiding quality control. Moreover, cell-based production makes it easier to incorporate other practices that improve efficiency such as job sequencing and the use of transfer batches.

The word "cell" is used quite liberally in practice to describe any association or grouping of machines. In this research, we define a cell as a grouping of machines used to process a family of one or more parts. We assume that the part families are pre-specified. In our factory representation, there are NC cells, indexed by $n = 1, \ldots, NC$. Each cell may include more than one of any machine type. Each cell has a certain number of machine types, with multiple machines of the same type organized into *machine centers*. We reserve this term for the cell shop and call the analogous machine cluster in a job shop a department. Of course, since cells do not contain duplicate machines very frequently, most machine centers just have a single machine of a given type.

The flow discipline for batches through the machine centers of each cell is identical to the rules governing the job shop as the batch visits several departments. Once the batch completes its processing within a given job shop department or cell machine center, it moves as an entire batch to its next operation or exits the factory if no further processing is required.

The preceding statement requires modification if a cell uses transfer batches. In this case, each batch is split up into the transfer batches that then queue up before the appropriate machine center. Note that because transfer batches constitute the only aggregation of units recognized within the cell, the identity of the original batch is not recovered until all of its constituent transfer batches have completed their processing within the cell. In fact, prior to leaving the cell and prior to being shipped, the work must be re-batched into its original batch size as required.

1.2 Factory Conversion

The conversion from process layout (job shop) to cellular configuration is a key question of both theoretical and practical importance in the field. As Cohen and Apte (1997) describe,

> In implementing cellular manufacturing an important task is to create a plan for smooth transition from process layout to manufacturing cells layout. Rearranging machines into cells based on part families is also a major undertaking requiring both considerable time and expense.

Once a machine is moved to a cell, it is removed from the general resource pool of the job shop and confined to processing within the cell. To avoid inter-cell moves as much as possible, cells are discouraged from accepting work required for parts that are not assigned to the cell, even if idle machine capacity exists. In this research, we assume that the cells are independent, so that each part family can be processed entirely within one cell. Inter-cell moves add to the complexity of flow and work control and can re-introduce setups. To avoid these drawbacks, we simply disallow them and assume that the cells formed are independent.

If the entire factory is partitioned as far as possible into cells we call this a *cellular manufacturing system* (CMS). This may include a remainder cell or residual job shop containing exceptional elements.

Example Factory

To illustrate the concept of cells, we present data from Morris and Tersine (1990) in Table 1-1. This table shows a part routing matrix for a factory with 30 machines falling into eight machine types. The factory produces 40 distinct parts that fall into five part families. For each part, the numbers listed along the row specify the order of the operations required, and the columns specify what machine type is needed for each such operation. For example, part 10 requires 3 operations (or processing steps), with the first performed by machine type 8, followed by type 1 for the second operation, and finally type 2 as the last operation. The path of the part through the departments is shown in Figure 1-2.

	Machine Type									Machine Type							
P/N	1	2	3	4	5	6	7	8	P/N	1	2	3	4	5	6	7	8
1					1	2			21			1		2	3	4	5
2			1	2	3				22		1	2	3	4	5		
3			1	2					23	1						2	3
4			1		2	3			24	1	2	3	4	5	6		
5					1	2			25					1	2	3	4
6			1	2		3			26		1	2	3	4	5		
7			1	2	3	4			27		2	3	4				1
8				1	2	3			28	1	2	3	4	5			
9	2	3				1			29	1			2	3			
10	2	3						1	30	2	3	4	5				1
11	2							1	31		2	3					1
12		2					1		32	2	3	4	5				1
13	2	3				1			33		1		2		3		
14	2							1	34	1	2	3	4	5	6		
15	3	4				1		2	35			1			2	3	
16		2					1		36			1	2	3	4	5	
17	4					1	2	3	37			1		2	3	4	
18		2					1		38	1	2		3	4			
19	1			2			3	4	39		1	2		3	4		
20	1	2	3	4	5	6			40	1	2		3	4		5	

Table 1-1. Part routing matrix: operation sequence linking part number with machine type.

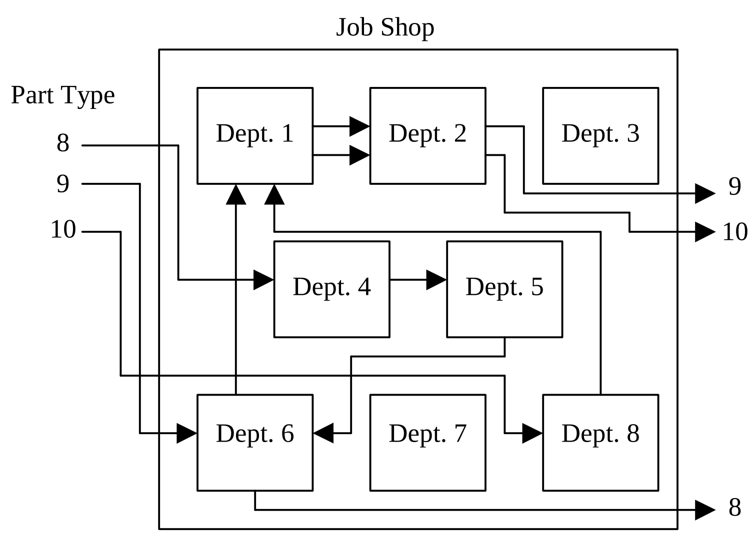

Figure 1-2. Illustrative part routings for parts 8, 9, and 10.

Morris and Tersine (1990) grouped the 40 parts listed above into the five families shown in Table 1-2. They formed the cells so that each family is assigned to a unique cell that is equipped with all the machine types required for the complete processing of the part family assigned to it.

Family	Part Types	Cell	Machine Types Required
1	33-40	1	1-7
2	19-26	2	1-8
3	27-32	3	1-5, 8
4	9-18	4	1-2, 6-8
5	1-8	5	3-6

Table 1-2. Summarized family and cell requirements.

The resulting cells are shown in Table 1-3. Five families and cells are identified in Table 1-4 where the block-diagonal form indicates the complete independence of cells. The numbers of machines of each type available in the original job shop were sufficient to equip all cells appropriately. If six cells had been formed then the addition of new machines would have been necessary (assuming the first five cells required the machine types shown in Table 1-3). In general, cell formation may augment or maintain the number of machines in the original job shop.

Machine Type	Number of Machines per Type in the Job Shop	Cells 1	2	3	4	5
1	4	1	1	1	1	
2	4	1	1	1	1	
3	4	1	1	1		1
4	4	1	1	1		1
5	4	1	1	1		1
6	4	1	1		1	1
7	3	1	1		1	
8	3		1	1	1	

Table 1-3. Machine distribution.

		Machine Type																													
		1	2	3	4	5	6	7	1	2	3	4	5	6	7	8	1	2	3	4	5	8	1	2	6	7	8	3	4	5	6
Family	P/N	Cell 1							Cell 2								Cell 3						Cell 4					Cell 5			
1	33		1		2		3																								
	34	1	2	3	4	5	6																								
	35			1			2	3																							
	36			1	2	3	4	5																							
	37			1		2	3	4																							
	38	1	2		3	4																									
	39		1	2		3	4																								
	40	1	2		3	4		5																							
2	19								1			2			3	4															
	20								1	2	3	4	5	6																	
	21										1		2	3	4	5															
	22									1	2	3	4	5																	
	23								1						2	3															
	24								1	2	3	4	5	6																	
	25												1	2	3	4															
	26									1	2	3	4	5																	
3	27																	2	3	4		1									
	28																1	2	3	4	5										
	29																1			2	3										
	30																2	3	4	5		1									
	31																	2	3			1									
	32																2	3	4	5		1									
4	9																						2	3	1						
	10																						2	3			1				
	11																						2				1				
	12																							2		1					
	13																						2	3	1						
	14																						2				1				
	15																						3	4	1		2				
	16																							2		1					
	17																						4		1	2	3				
	18																							2		1					
5	1																													1	2
	2																											1	2	3	
	3																											1	2		
	4																											1		2	3
	5																													1	2
	6																											1	2		3
	7																											1	2	3	4
	8																												1	2	3

Table 1-4. Partitioned part routing matrix indicating part operation sequences, part families, cells and machine types per cell.

An alternative to a completely converted CMS is what we call a *partial cellular manufacturing system* (PCMS). This is a hybrid layout where a number of cells are formed to work alongside a remainder job shop. In other words, the formation stops short of full conversion. The parts are therefore manufactured in the cells or in the residual shop; however each cell is dedicated to the manufacture of a unique part family. Naturally, machines not used in the cells implemented remain in their residual job shop departments.

The information gathered from industry practice shows that partial implementation is often the preferred path for implementation. Surveys show that firms create cells one by one (Wemmerlöv and Hyer 1989, Wemmerlöv and Johnson 1997). In fact, a study by Ahmed, Nandkeolyar and Mahmood (1997) indicates that practitioners do not exercise full conversions and that successful implementation is linked to long-term, step-by-step installations.

1.3 Key Trade-offs

A consistent feature of all conversions to a CMS environment is the segregation of machines of each type from the pooled arrangement of a department to smaller subsets assigned to the cells. Wolff (1989, p.260) uses the term pooling to refer to the aggregation of the arrival streams of c separate queues into a single queue where the server is equipped with the pooled resources of the original queues. He notes that the pooled queue performs better and goes on to state that "the superiority of pooling can be shown to be a very general result independent of the nature of the arrival process and the distribution of service." Accordingly, we refer to the diseconomies of segregating a given machine type by assigning them to independent cells as the *pooling loss*. This pooling loss always causes an increase in flow time. Therefore, for the cellular system to outperform the functional layout with respect to flow time, this pooling loss must be compensated by improvements in such other factors as setup times or move times. In summary, when flow time is the performance measure of primary interest, the superiority of cellular layout over functional layout is tantamount to finding the means for overcoming pooling loss.

A simple queueing model based on the well-known M/M/*c* formulas has been used to illustrate the nature of the pooling loss as in studies by Suresh (1991, 1992), Shafer and Charnes (1993, 1995,1997), and Suresh and Meredith (1994). A simple example will illustrate this modeling approach.

In Figure 1-3, we compare the flow times of two systems -- a pool of four machines corresponding to a job shop department (solid line), and a system of four cells, each consisting of a single machine performing a single operation (dashed line). We model the job shop as an M/M/4 system with $\mu = 1$ for the JS and equate the arrival rates to both systems. For point *A*, the flow time for the M/M/4 system equals 1.25 when $\rho = .65$ ($\rho = \lambda/4\mu$) corresponding to an arrival rate of $\lambda = 2.6$. When we segregate the shop into four equal demand streams of $\lambda/4$, the flow time for each cell equals 2.86 (point *B*), which is 2.28 times the M/M/4 flow time. In order for the flow time in the M/M/1 system to be the same as the pooled system, so that $W_{M/M/1} = \rho/(\lambda(1-\rho)) = 1.25$, the processing rate must be increased such that the resultant utilization is $\rho = .448$ or roughly one and a half times as efficient, $\mu_{\mathrm{CM}} = 1.45\mu_{\mathrm{JS}}$, as the same process in the JS.

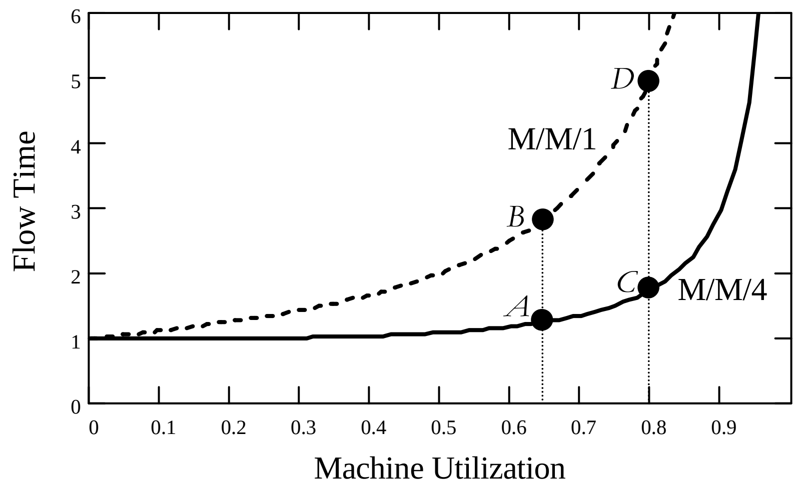

Figure 1-3. Pooling loss.

As Figure 1-3 suggests, flow time increases without bounds with the linear increase in machine utilization. If the M/M/4 is run at $\rho = .80$, the flow time is 1.75 (point *C*). After conversion, the M/M/1 flow time is 5.00 (point *D*) per cell or 2.86 times the M/M/4 flow. Comparing the pairs *A*-*B* and *C*-*D*, when ρ increases from .65 to .80, the ratio of the flow time increases from 2.28 to 2.86.

The last point is of particular importance since it shows how increased utilization magnifies the pooling loss. This effect occurs where bottlenecks arise as a result of conversion to cells, limiting the capacity of the process. In general, in Chapters 3 and 4 of this work, we will see how conversions from JS to CMS are especially sensitive to the loading of machines in both the cells and the remainder shop. Suresh (1991, 1992) has also alerted readers to “adverse effects in the remainder cell” that are typically due to loading imbalances.

We already mentioned that reductions in setup constitute one of the key factors for overcoming pooling loss. Major setups are typically incurred when the same machine switches from one family of parts to another. The frequency with which setup occurs depend on the demand and service rates as well as the dispatching rule. A dispatching rule is a priority rule or set of rules used in determining the order of service for customers waiting in line. In this dissertation, we focus on a dispatching rule that is designed to minimize the incidence of setups. We do not consider preemptive dispatching rules because job interruptions will markedly increase the complexity of workflow control.

1.4 Research Objectives

The research we present has two objectives. First, we investigate the role of setup economies in the factory-wide conversion of functional layouts (job shops) to cellular manufacturing. While the literature has chiefly focused on full job shop to cell shop conversions, we include both complete and *partial* factory conversion options (where a sizable residual shop is left in conjunction to the cells). Our second research objective is to examine the role of dispatching rules in the reduction of setups.

1.4.1 Research Issues and Methodology for Factory Conversions to CMS. Our research seeks to answer the following questions regarding the results of setup economies in the cases of factory conversions:

- Can consistent results be obtained as to when the conversion of the job shop to a cell shop is advantageous?

- What are the measured setup economies? When are setup economies large enough to overcome pooling losses?
- How do other cell factors, including reduced batch sizes and use of transfer batches, affect flow times achieved in cells?
- Can a partial implementation of CMS provide all or most of the benefits of full conversion to CMS?

The approach taken to answer these questions is to use a single simulation model to compare functional and cellular layouts across a test bed of factory environments extracted from the literature. In our attempt to perform such a comparison, we follow the established practice of most analytical or simulation conversion studies in using flow time as the primary performance measure for comparing JS and CMS layouts. Little's law then can be used to relate the flow time to inventory measures such as length and wait time in queue and number of customers in the system. We realize that the average batch flow time may not directly relate to the total product cost. We actually capture the flow time of each and then calculate the weighted average, using the part type demands as the weights. This will be a reasonable surrogate for cost if there is a linear relationship of cost to piece part flow time. For example, inventory-related costs are often modeled to be linear in the amount of time each part spends in the system. In this case our measure would be a surrogate for part costs if all part types have the same monetary value. Alternatively, we can use a weighted average in which we weight part types by their contributions to the total cost of goods sold (GOGS). Our contribution is to control the parameter choices in the data sets in such a way as to make them comparable. We call this approach standardization. Table 1-5 lists our assumptions in the factory conversion part of this research.

Primary performance measure	Average batch flow time
Process flow coordination	Asynchronous
Machine selection	One machine type specified per operation per part (no alternates)
Machine input queues	Infinite capacity, shared by machine type within job shop department or cell machine center
Machine operation	Sequential processing on the same machine type is combined within one operation sequence
Machine output queues	None: sufficient material handling capacity exists to move output immediately to next operation
Use of transfer batches	Only allowed in cells
Cell resources	No inter-cells moves allowed or job shop to cell moves allowed: all cells assumed to be independent and capable of processing part family in entirety

Table 1-5. Assumptions for factory conversion research.

To our knowledge, this is the first study where conversion benefits are studied across data sets selected from different sources in the literature. Our results show that for a given region of the parameter space, the conversion to cellular layouts consistently produces an advantage even in the absence of the gains resulting from lot size redefinition and lower movement or transport times. In addition, we are able to generate caveats for the implementation process from our PCMS results.

1.4.2 Research Issue and Methodology for Analytic Modeling of a Simple System with Setup. Our research seeks to answer the following intuitive question regarding setup economies using models of a simple system:

- What is the impact of the dispatching rule used in the reduction of setups?

The approach taken to answer this question is to apply analytic queueing models to a system that is simple enough to make exact analysis tractable. The single-stage, single-server system involving two customer classes is the simplest case where setups occur due to part changeovers. Our choice of this simple system is driven by the existence of exact results on flow times and the fact that modeling of setups best matches the manufacturing environment studied in this dissertation. We start by establishing a baseline using zero setup, evaluating flow times under FCFS versus a dispatching rule that minimizes the incidence of changeovers. We then extend the results to the case of non-zero setup. Table 1-6 lists our assumptions for the analytic section of this research.

Primary performance measure	Average batch flow time (batch size = 1)
Setup incidence	Incurred when switching from one class of part to the other (setup magnitude → 0)

Table 1-6. Assumptions for analytic modeling research.

New flow time results are provided using different dispatching rules. These results are obtained analytically for the case of zero setup times and extended to the case of non-zero setup time through computational studies.

1.5 Plan of the Dissertation

In Chapter 2, we review the literature relevant to the two distinct parts of this dissertation. We first review the literature on the conversion to cellular manufacturing using simulation modeling (including both complete transformations and partial transformations). Next, we review the key sources in the queueing literature that consider single-machine processing in the presence of setups. In Chapter 3, we present our study of the full conversion of job shops to cells shops. The first part of Chapter 3 outlines the factory production environment. Here we describe the choice of data sets included in the test bed, identify the manufacturing characteristics of each data set, introduce the standardization scheme for the simulation study, and describe the simulation model. Section 3.6 describes the results of the simulation runs comparing functional and cellular layouts. Of special importance are sensitivity runs included to study the effect of batch sizes, transfer batches, factory loads, setup parameters, and dispatching rules. Chapter 4 provides a brief account of our investigation of partial cellular implementation.

Chapter 5 is devoted to the analysis of a single-server system with two classes and switching (setup) costs. Section 5.1 is dedicated to the zero-setup baseline and 5.2 to the non-zero setup extension. Chapter 6 contains summaries of the key findings of our research and outlines several directions along which future research can be conducted. We have also included a short glossary of key terms used for the reader's convenience.

Chapter 2

LITERATURE REVIEW

This chapter reviews the literature relevant to the two segments of this dissertation. First, in Section 2.1, we review simulation studies that have dealt with the conversion of job shops to cellular layouts for both full and partial conversion (in a partial conversion, a sizable residual shop processes parts along with the cells). In this chapter, we reserve the term *factory conversion* for a change in the layout.

The second section, 2.2, reviews the modeling literature for the multi-class, single-stage processing facilities modeled as queueing systems. Our focus is on analytic models that can handle setup times.

2.1 Conversion Analysis Using Simulation

The comparison of functional and cellular layouts in the manufacturing of discrete parts is a topic that has received much research attention over the last decade. This comparison is often performed when a job shop (JS) is converted to a cellular manufacturing system (CMS) experiencing the same demand. On the one hand, reports from industry continue to claim superior performance for cellular layouts, although the measured improvement seems to vary substantially. For example, Wemmerlöv and Hyer (1989) reported average flow time reductions of 24% for cellular layouts, whereas Wemmerlöv and Johnson (1997) reported an average reduction of 61% in throughput times for 27 respondents. On the other hand, simulation modeling studies in the research literature have produced divergent and at times contradictory results in evaluating the effect of conversion on flow times. Nor is the literature of one voice in providing a clear basis or a consistent list of quantifiable factors that would ensure the benefits of conversion.

The empirical data also shows that partial conversion is also used in practice. A study by Ahmed, Nandkeolyar and Mahmood (1997) indicates that practitioners do not opt for full conversions and that successful implementation is linked to long-term, step-by-step conversion to cellular manufacturing.

To facilitate our review of the literature, we introduce our performance measure now. Since CM is used to improve the efficiency of a job shop, a job shop will be the basis for our performance comparisons. For comparative purposes, the *flow ratio* (*FR*) is defined as the ratio of the average batch flow time after cellular conversion to the average batch flow time of the job shop with the same factory operational parameters of load, machines and batch size. This definition is slightly different than that used by Suresh (1992) where the flow ratio related the cellular transformed flow time to the best job shop flow time which may be measured at a different batch size.

2.1.1 Complete Factory Conversions

In their paper on this subject, Johnson and Wemmerlöv (1996) performed a meta-analysis of the results of 24 simulation studies designed to investigate the performance characteristics of conversions from JS to CMS. These authors conclude, “universal evidence regarding the superiority of cellular versus functional systems can never be provided due to the data dependency involved.” However, they also remark that whether cellular layouts outperform their functional counterparts depend on a complex interaction among several key factors including the utilization level, the degree of resource pooling, setup and move time reductions, and batch sizes used.

To aid in our review of the simulation-based literature on factory conversion, it is useful to compile a list of factors that can be expected to influence the performance of job shops as compared to cell shops. We then look at the comparisons provided in Johnson and Wemmerlöv's 1996 meta-analysis and examine the different factory conditions tested. In this chapter, our focus is on the setup reduction as the key advantage of cells, rather than material handling gains.

We define our terms used in this review in Table 2-1. We then compare the range of factors and factor settings in five simulation studies in Table 2-2. We follow with reviews of key studies in the literature (the five in Table 2-2 with others) that use simulation to investigate factory conversion.

Following the review of the studies, Table 2-4 lists the studies in chronological order and the overall conclusions drawn for each paper.

Operations/part	Range in the number of operations per part across all parts
Machine Types	Number of distinct machine types
Machines	Total machines
Machines/type	Range in the number of machines per distinct machine type
Cells	Number of cells the JS is converted into (one cell may be a "remainder" and process unrelated parts)
Batch Size	Batch size used in the JS layout and CMS unless stated otherwise. A list of batch sizes means denotes experimental factor settings
Major Setup	A major setup is incurred if two parts belonging to distinct families are processed consecutively on the same machine.
Minor Setup	Switching between two different part types in the same family. Typically less than a major setup.
Setup Ratio: s/br	Ratio of major setup to mean batch run time per part.
Setup Fraction	Ratio of minor to major setup per part.
Dispatching Rule	FCFS: First come, first served; RL: Repetitive Lot (from Jacobs and Bragg, 1988): (1) A single (pooled) queue is formed for all batches arriving to be processed at a machine center. (2) Any arriving batch encountering an available machine upon entry is immediately routed to the available machine where it would require the least setup time. If no machines are available, the batch joins (or forms) a queue to wait for a machine. (3) When a machine becomes available, the next job assigned to it is selected based on the minimum setup among all jobs in queue. If multiple jobs tie at this minimum setup value, the FCFS discipline is used to break the tie.
JS Utilization	Source JS average machine utilization as measured by simulation
Cell Transfer Batch Size	Transfer batches used only in cells. No transfer batches is designated by *b*, the JS batch size, otherwise a value is listed
Arrival Rate Distribution	Distribution of arrivals with its coefficient of variation in parentheses.
Setup Time Distribution	Like arrival distribution above, but for setup
Material Handling Times	Material handling as a fraction of part run time

Table 2-1. Study definitions.

Source	Morris and Tersine, 1990	Suresh, 1991	Shambu and Suresh, 2000	Suresh and Meredith, 1994	Shafer and Charnes, 1997
Parts	40	40	75	50	4
Operations/Part	2-6	1	2-7	3-7	4
Machine Types	8	1	12	10	4
Machines	30	4	63	31	16
Machines/Type	3-4	1, 1, 2	3-4	3-4	4
Cells	5	2 or 3	1-5	5	1-4
Batch Size	50	5-100	2-80	JS: 32-100 CM: 5-100	
Setup Ratio: s/br	0.06 or .1, .5, 1.0	6-0.3		0.3-6.0	1.3-0.2
Setup Fraction: Min/Maj Setup	0.5		0.1-0.5	0.1-1.0	0, 0.5, 1.0
Dispatching Rule	RL	FCFS, SPT	FCFS, FSP, RL	FCFS or RL-F in JS, FCFS in cells	RL
JS Utilization	60-70%			70% with b=50	62%, 75%
Cell Transfer Batch Size	b	b	b	b	b, $b/2$, 1
Arrival Rate Distribution (CV)	Poisson (1.0) deterministic (0)	Poisson (1.0)		Poisson (1.0) 3-Erlang (.58)	
Setup Time Distribution (CV)	Normal (0.08)			Poisson (1.0) 3-Erlang (.58)	
Part Run Time Distribution (CV)	Normal (0.36)	Gamma (0.7-1.2)		Poisson (1.0) 3-Erlang (.58)	
Material Handling Times	~.15r between depts., 0 within cells			3r - 120r between depts., .75r within cells	JS only: 0 or 0.6br

Table 2-2. Comparison of factor levels within simulation studies

Morris and Tersine (1990) studied the full conversion of a five-cell CMS. They examined the impact of changes of setup ratio, move time, demand stability and flow of work within the cells on the conversion. The "demand stability" factor regulated the sequence of part batch arrivals such that there was a maximum interval between like part types. The work within cells was random and allowed backtracks or part sequences were altered to provide unidirectional flow. Morris and Tersine (1990) considered their shop configuration "supportive" of CM due to the independence of their cells, use of identical lot sizes in both layouts, and use of RL dispatching.

Their results showed that the setup ratio factor could bring the flow time within 5% of the job shop value. In contrast, their base case resulted in an all-cell shop with flow time 50% greater than the job shop value. When high setup level was compounded with other factors such as slow JS move times, and unidirectional flow), the all-cell flow time was 10% better than that of the job shop. Overall, the authors concluded limited promise for CM. Looking closely at their experimental setup it is evident that simply increasing the setup time magnitude for each operation created the high setup level. Using the same run times, this increase in setup burden added to the machine utilization of both the job shop and all-cell shop and raised all flow times as reflected in their mean throughput times (see their Table 4). Operating the cell shop in this high machine utilization region, as noted in the conclusions of Morris and Tersine (1989), can distort the apparent impact of setup due to the sensitivity of the flow time to machine utilization.

Suresh (1991) used a single-operation simulation model with parts from three families. One of the three families represented 50% of the total parts in the factory and roughly 50% of the total demand. Although deemed a “family” by the author, there was essentially no similarity between parts. Setup discounting was handled differently than in Morris and Tersine (1990) – setup was not discounted in the job shop or in the family of unrelated parts and was discounted by a flat rate of 70% or 90% in the cells (independent of processing sequence). The dispatching rules included a truncated shortest-setup-plus-run-time (SPTT). The SPTT rule calculated a due date and gave priority to late jobs followed by shortest discounted setup plus run time. As each family was moved to a cell, a new batch size for that family was determined from a pre-selected range (approximately 10% of that originally in the JS).

Even with a 90% setup discount in the cells and at a setup ratio of 0.6, the all-cell flow time was 25% greater than the job shop value using FCFS in both job shop and cells and 9% greater using SPTT in the job shop and cell shop. The study also showed that SPTT performed better (14%) than FCFS in the job shop using the same batch size. We observe, therefore, that if SPTT was combined with cell conversion then it would have resulted in a 6% *improvement* over the job shop using FCFS. The authors noted that the flow time of parts in the cells improved even though the overall factory flow time was inferior to the job shop. The authors attributed this to adverse effects in the remainder. We understand these “adverse effects” to be pooling losses: machines were removed from the job shop pool, but the relative load per machine did not change. In Suresh (1991), only when the setup discount was coupled with a reduction in cell batch size (made feasible for the cells from setup reductions) was the transformed shop capable of improved factory flow times over the JS.

The results of Suresh (1991) appear to corroborate the conclusion of Morris and Tersine (1990) that large amounts of setup reduction alone are not sufficient for the cells shop to overcome the pooling losses and outperform the flow time of the job shop. Although Suresh included similar factors and levels as in Morris and Tersine, we note in Table 2-3 that they were handled differently.

Factor	Morris and Tersine (1990)	Suresh (1991)
Setup discount	family-based throughout the shop	flat-rate setup discount applied to two of three cells only
Dispatching Rule(s)	Repetitive Lot	FCFS or SPT
Remainder cell	none	50% of parts in remainder cell and did not receive setup discounts

Table 2-3. Difference in operating scenarios may confuse comparative results.

Shambu and Suresh (2000) confirmed Jacobs and Bragg (1988) in showing that RL is superior to FCFS and SPT dispatching rules. Shambu and Suresh (2000) report similar results as those in Shambu's 1993 dissertation. They found that in the cells RL/SPT (part batch with shortest expected processing time picked from queue) is only marginally better than SPT (without using RL), but both outperform FCFS. The authors note that the likelihood of identical parts being processed in succession in a cell is small so RL rarely impacted the queue. In addition, if the setup fraction is small then the savings potential due to eliminating the minor setup is minimal.

As in Suresh (1991), the flow time of parts in the remainder shop of Shambu and Suresh (2000) was found to deteriorate with increasing number of cells, even when the flow time of the cell parts improved over their flow times when in the job shop. The authors used family-based setup in the residual shop like Morris and Tersine (1990) and still found increasing flow time. They attribute this decline in performance of the residual to pooling losses that were not overcome by any residual shop setup improvements.

The choice of the batch size as a factor in conversion to cells is central to Suresh and Meredith (1994), who set batch sizes (one size used for all parts) across a range for the job shop and then reset them the cell shop configurations. Setups were family-based with the setup fraction ranging from 10% to 100% (no discount).

Their results with both the job shop and all-cell shop using family-based setup showed up to a 54% batch flow time reduction from a job shop to a cell shop (both with 10% setup fraction). This was assuming cells used use the same batch size as the job shop. They report improvements of 58% with batch sizes half that of the job shop. This was their most extreme result using equal batch sizes, but it was based on using a job shop with average machine utilization over 95%. At another setting, the job shop was loaded at approximately 75% machine utilization. The resulting reduction in batch flow time for the same setup fraction in the cell shop was 16% at the same batch size used in the job shop and 67% at a batch size $1/6^{th}$ that of the job shop. As expected, the job shop flow times were best with the lowest move time setting.

Suresh and Meredith (1994) concluded that of the factors they studied influencing the shop performance, setup and run time reduction had the greatest impact as opposed to batch size and variability reduction. We note that batch size of the cells did not have to be reduced from that of the job shop for the factory to realize savings in flow time (as long as setup fraction was less than 0.5).

Shafer and Charnes (1997) results show that the overall flow time increased with increasing setup ratio, but decreased with decreasing setup fraction. The flow time also decreased with transfer batch size. The job shop flow time increased with move delay. The authors concluded that each of the factors they tested, if set at the appropriate level, may be sufficient to overcome pooling loss resulting in improved flow time performance over the job shop. The authors concluded that an all-cell shop (using transfer batches of size one) can generally reduce job flow time by 45%-65% over a comparable job shop and showed that without transfer batches less than the original batch size the flow time could be reduced 11% (assuming 50% setup fraction).

Table 2-4 summarizes each of studies above in chronological order. The column labeled “factor” specifies the key factory investigated in the paper. For example, the first paper listed investigated the effect of move times and the demand distribution on the conversion. The last column, entitled “limitations,” summarizes our observations on the study from the perspective of the research questions addressed in this dissertation.

Source	Factor(s)	Conclusions	Limitations
Flynn and Jacobs, 1987	Move times	Less move times when using cells. Congestion can more than offset setup and move advantages of cells.	Used average of >3 inter-cell moves per part style. Only tested FCFS.
	Distribution of demand based on part processing times	Very little impact	
Jacobs and Bragg, 1988	Dispatching rules: changeover minimizing (repetitive-lot, *RL*), FCFS	RL dispatching superior to FCFS in environments with setup. RL also worked well in presence of transfer batches	
Burgess, 1989; Burgess, Morgan and Vollmann, 1993	Labor constraint	Reduced with increasing setup reduction	No setup discounting in residual JS, fixed setup discount applied to cell
	Magnitude of cell-based setup	Cell machines increase efficiency with increasing setup discount.	
	Fraction of factory load sent to cell	Cells machines more efficient with setup reductions so optimum load is greater than JS fractional load	

Table 2-4. JS to CMS Conversion Literature Summary

Source	Factor(s)	Conclusions	Limitations
Morris and Tersine, 1990	Setup	Ratio of setup to batch run time must be ≥1.0 for cells to improve on job shop flow time. Setup is dominant factor in cell success	High and low setup conditions run at different factory settings of machine utilization and therefore may have affected the reporting of the setup impact
	JS move time	Move time resulting in ≤0.5% of batch run time is not enough to overcome cell pooling loss	
	Constant arrival distribution, period batch control	Reduces flow time of cells, but not enough to outperform the job shop	
	Flow of work in cells combined with demand stability	Uni-directional flow with stable demand does not improve flow in cells over demand stability alone	
	Setup, move, demand stability, flow of work	Cells combining four factors yield 10% flow time improvement over job shop, thus cells may have limited promise	

Table 2.4 (cont.). JS to CMS Conversion Literature Summary

Source	Factor(s)	Conclusions	Limitations
Suresh, 1991	Dispatching rule	Truncated Shortest Processing Time (SPTT) better than FCFS	No setup discounting in JS, fixed setup discount applied to all parts in cell
	Setup discount in cells	Large setup discounts not sufficient to lower factory flow time below job shop	
	Batch size	Reduced batch sizes feasible in cells due to much smaller setups	
	Number of cells	Residual flow time deteriorates with increasing number of cells	
Shafer and Charnes, 1993	Transfer batches	Capable of overcoming pooling losses even coupled with long process routes, small batch sizes, short processing times and the absence of natural part families	
Shafer and Meredith, 1993	Process route length	Long routes, small batch sizes, short processing times and the absence of natural part families limit the effect of transfer batches, a major advantage of CM	No setup discounting in JS, fixed setup discount applied to all parts in cell
	Batch sizes		
	Processing time per part		
	Part families		

Table 2.4 (cont.). JS to CMS Conversion Literature Summary

Source	Factor(s)	Conclusions	Limitations
Shambu, 1993; Shambu and Suresh, 2000	Dispatching rule	RL is better than SPT that is better than FCFS in job shop, remainder and cells.	
	Number of cells	Flow time for the remainder shop deteriorates as cells are added	
	Batch size	Setup and lot size reductions are most important for cell performance. Setup savings in cells is primarily from elimination of major setups – setup reduction once in a cell has minimal impact	
Suresh and Meredith, 1994	Minor/major setup	Must be under ½ for cell flow to be better than JS	
	Run Time	Setup and run time have greater impact on flow time than batch size and process variability	
	Arrival and processing variability		
	Batch size	Cell batch sizes don’t have to be less than JS for improved flow. Lower cell batch sizes leads to lower factory flow time	

Table 2.4 (cont.). JS to CMS Conversion Literature Summary

Source	Factor(s)	Conclusions	Limitations
Seifoddini and Djassemi, 1997	Part changes	Cells more sensitive than JS. CMS sensitive to product mix changes	
Shafer and Charnes, 1997	Move time	Each capable of overcoming pooling loss due in cells due to CMS conversion	only 4 parts: one part per cell
	Minor/major setup		
	Transfer batch		
	Setup/batch run time		

Table 2.4 (cont.). JS to CMS Conversion Literature Summary

Other Relevant Full-Conversion Studies

The following studies provide insights on other factors of a more subtle nature to the factory conversion literature.

Shafer and Meredith (1993) were mainly interested in transfer batches in a study of data from industries. Transfer batches were used exclusively in the cells. They reported improvement in performance largely due to transfer batches. Recognizing this, looking across their plant-specific results they determined a number of factors that *limit* the advantages of cellular manufacturing by limiting the effect of transfer batches:

1. Short process routes
2. Small batch sizes
3. Short processing times per part
4. Absence of natural part families (reduces ability to form cells, and therefore the use of transfer batches or cell-based setup reductions)
5. Existence of bottleneck machines (in general cause large queues, also reduces benefit of transfer batches)

Finally, Seifoddini and Djassemi (1997) compared the effect of part mix changes to a shop configured as a job shop or all-cell shop. For a fraction of parts, part operations were changed and then the resulting changed parts were re-assigned to different cells. For example, one part type was eliminated from the first cell family at the same time one part type was added to the third cell family. Each part added to a part family contained machine requirements consistent with its cell machine types (no inter-cell moves required). Following this example, the first cell experienced a reduction in demand and the third cell experienced an increase in demand. As we would expect, the smaller cell machine pools were more sensitive to part changes than the job shop experiencing the same part changes. We conclude from Seifoddini and Djassemi that CMS sensitivity to changes is not reflected in the factory flow time measure.

Full Conversion Summary

The literature provides sufficient evidence that given enough potential, move time, setup or transfer batches are capable of overcoming pooling losses independently of the other in cellular conversion. We also see the indication that the use of transfer batches and machine loading may be key factors in cellular conversion.

Review of Meta-Analysis

We now look at the results and then the factor settings compiled in Johnson and Wemmerlöv (1996) more closely to capture their variety. Figure 2-1 plots the range of observed flow ratios for 24 studies in the literature summarized by Johnson and Wemmerlöv sorted by the lowest reported flow ratio. We simply converted the measure called RAT (reduction in average flow time) reported by Johnson and Wemmerlöv into flow ratios and used the lowest and highest flow ratios observed by the authors of each study in the course of their experiments. Consider the vertical line indicating a flow ratio of 1.0. Any study for which the bar intersects this line includes results where the CMS and job shop have the same flow times. Similarly, if we draw two additional lines to mark the boundaries of a 20% band about the 1.0 line, we can highlight the regions where a given study shows a clear advantage for either the job shop or CMS. We see the results are mixed. Only one study, namely Shafer and Charnes (1993), reports flow ratios that lie consistently below 1.0, a majority shows their range of results entering this region, but with the range extending into region that show a clear advantage for the job shop. While we wouldn't expect the ranges to be the same, we find that some studies have *no* common cell conversion performance.

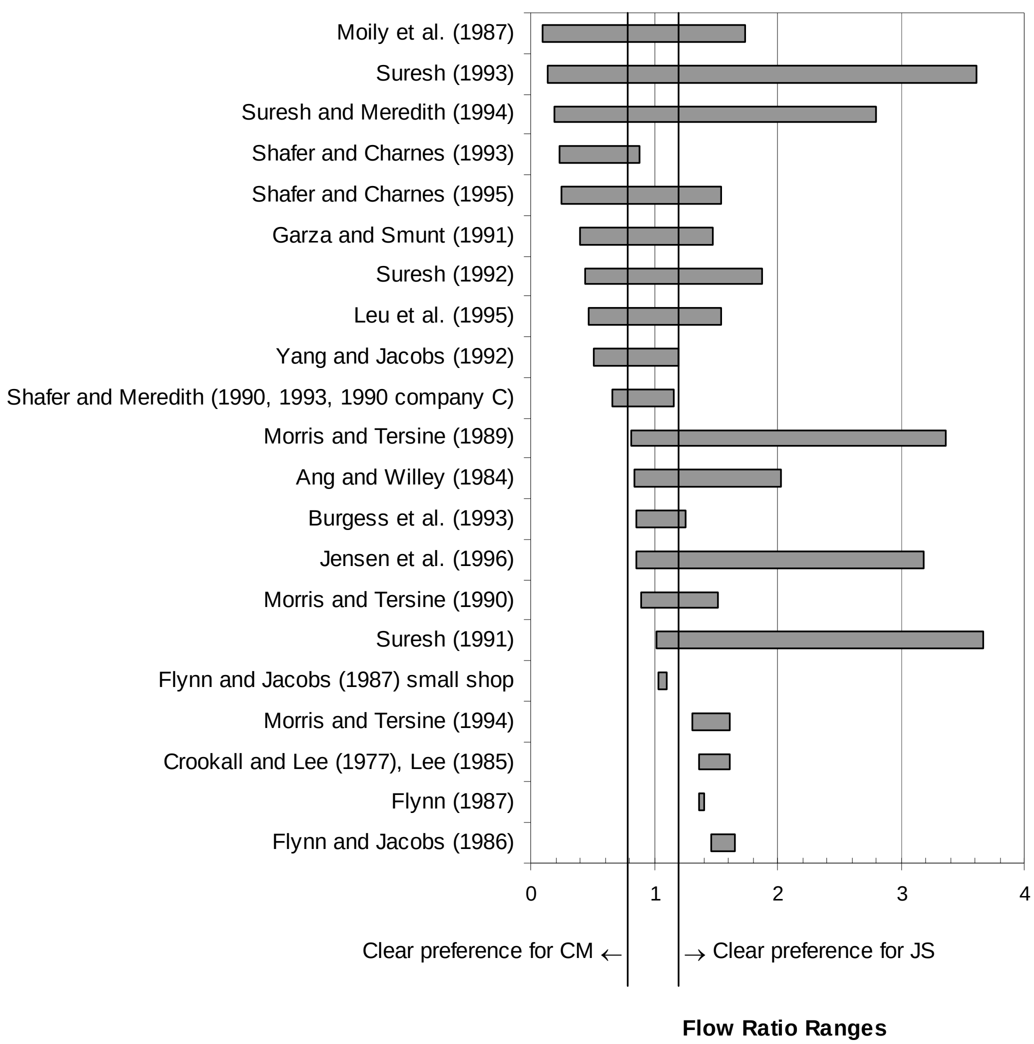

Figure 2-1. Disparity of results reported in Johnson and Wemmerlöv (1996).

There may be a number of reasons underlying the mixed results on the comparative performance of functional and cellular manufacturing layouts. The simulation modeling literature uses flow time to determine the success of the CMS implementation. Within industry, however, the implementation of cellular manufacturing may be driven by benefits that are not easily measured by traditional metrics in computational studies. For example, several key products may be segregated into cells to provide better control of operations or quality.

Interestingly, comparative results reported in the literature vary widely even when flow time is taken as the primary performance measure as measured by a simulation model. Closer examination shows that the studies reflect different values of key input parameters and use disparate operational rules as seen in Table 2-3 using the definition of terms in Table 2-2. Given the wide range of manufacturing settings investigated, it is not surprising that the results of conversion studies are not consistent.

2.1.2 Partial Implementation of Cells. We now review the literature on partial conversion where only part of the original JS factory is organized into cells. As mentioned before, this means that a significant part of the factory continues to operate as a JS, we call this the remainder shop. The overall hybrid system is also denoted by PCMS (for Partial CMS). We review the studies that specially focused on partial conversion and follow with a summary in Table 2-6.

Shunk (1976) was one of the first authors to use simulation for comparing CMS to JS. He identified experimental results where the flow time for PCMS was superior to both the JS and all-cell settings. However, the study did not offer any insights as to what lead to this phenomenon. When comparing flow time across the JS to eight- or nine-cell shop, the minimum flow time generally occurred with three to five cells, although it ranged from the two-cell to the nine-cell. In some cases the PCMS was better than the job shop configuration with respect to flow time, while the all-cell configuration was worse. Curiously, the job shop never exhibited the best flow time.

Burgess, Morgan and Vollmann (1993) compared the configuration of a single cell with a remainder shop to a job shop, without evaluating the all-cell alternative. These results are similar to those found in Burgess (1989). The research of Burgess et al. (1993) focused on the inclusion of labor constraints and we will not be considering labor constraints in our research. However, the converted shop in their research was not labor constrained so their insights on cell loading effects are relative to a machine-constrained shop.

Burgess et al. (1993) varied the fraction of parts sent to the cell. Since the work in the cell was discounted, the resources (machines and labor pool) appeared to become more efficient as compared to the job shop. In fact, the machine and labor pool capacity did not change in the cell, rather the setup requirement for each part entering the cell was reduced. The resultant factory-wide flow time was reduced even though the un-discounted part loading sent to the cell increased from 80% to 120% of the machine capacity. Of course the 120% loading is misleading because it assumes that the cell parts are paying a full setup which they are not.

As shown in Figure 2-2, it took only a 25% setup reduction in the cell to overcome the pooling losses as long as at least 40% of the parts were routed to the cell. We would expect the flow curves of Figure 2-2 to rise again when *too* many parts were sent to the cell suggesting an optimal loading exists.

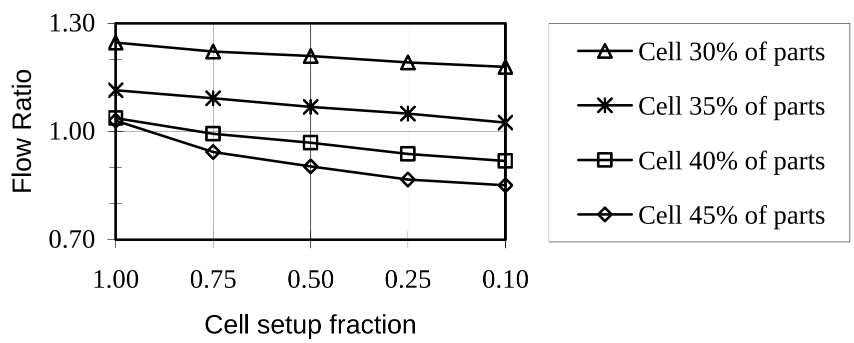

Figure 2-2. Optimal flow time improvements require controlled cell loading.

Burgess et al. (1993) concluded that prorating loads to cells in a manner simply commensurate with the resource fraction found in the job shop results in flow times that are inferior to the job shop configuration. In other words, prorating underestimates the load that should go to the cell. They suggest machine loading in general may be more critical to cellular success than advantages gained through shorter setup times. For our research, if we were to pick a single cell, favorable machine loading is something we would look for.

Suresh (1991) included an analysis of a hybrid shop transformation along with the complete conversion reviewed above. Suresh transformed a job shop into a hybrid configuration using either a single cell or two cells (operating alongside a residual job shop). Parts in the job shop and residual were not discounted; parts sent to cells were discounted at a fixed rate of 0.3 or 0.1. We see by the flow ratio results of Suresh listed in Table 2-5 that using a similar setup discounting scheme as Burgess et al. (1993), but sending loads commensurate with the machine fraction in the cell, that pooling loss is not overcome.

(FCFS in factory)	Cell parts setup fraction	Batch size: 30	40	50	100
		JS Flow			
Job Shop with four machines for 100% of parts	n/a	36.9	12.4	12.1	15.1
		Flow Ratios			
One cell with one machine for 20% of parts + residual	0.3	1.21	1.19	1.11	1.10
	0.1	1.20	1.16	1.08	1.08
One cell with two machines for 50% of parts + residual	0.3	1.69	1.50	1.22	1.16
	0.1	1.68	1.46	1.17	1.12
	setup/batch run time	1.0	0.8	0.6	0.3

Table 2-5. Deep setup discounts may not be sufficient to guarantee PCMS success.

Suresh (1991) concluded that partial (hybrid) situations are clearly unfavorable when compared to the JS even with high degree of setup reduction. He noted that the flow time of the cell parts improved over the job shop, yet the overall factory flow time did not. This indicates that the residual job shop is adversely affected. As we discussed earlier in this literature review, we expect the effects in the residual from its own pooling loss.

Shambu and Suresh (2000) compared a job shop to a PCMS with a remainder shop. Shambu (1993) presents similar results. They showed flow time results throughout the transition from JS to single cell all the way to five cells (with a remainder). Unlike the PCMS studies of Burgess et al. (1993) or Suresh (1991), the study by Shambu and Suresh used family-based setups throughout the factory. Setup discounts were not, therefore, strictly found in cells. This translated into a more even-handed comparison of factory environments. They showed that job shops using family-based setups could use smaller batch sizes than those that did not allow discounts in the job shop, confirming Suresh and Meredith's results (1994) for total conversion.

In their environment, it was shown that the a single cell shop (with residual) could be better than the job shop using the same batch size which was counter to the results of Suresh (1991). Looking carefully at the flow times, however, the residual flow was 4% worse than the job shop but the single cell flow was low enough to compensate (45% improvement) weighted by its demand. At five cells, the cells logged an improvement of 38% over the JS flow time and were paired with a residual that was 84% worse than the JS flow time. The net result was still a 12% improvement for flow times over the JS. This supports the previous research of Burgess et al. (2000), and Suresh (1991) suggesting that managing both cell loading and residual loading are important to optimize factory flow time of the PCMS. Finally, the authors sequentially picked cells for implementation based on an arbitrary cell numbering scheme even though they noted that each cell was not equally loaded and therefore not equal performers with respect to flow time. They concluded from their results that there were decreasing marginal cell gains as the number of cells formed increases. We do see differences in the marginal gains in their results, but (and by their own admission), it is due to loading differences and thus coincidence in cell implementation sequence. This helps motivate our research into the impact of picking cells to optimize factory flow time.

More recently, Kher and Jensen (2002) presented a study of PCMS based on a single data set they modified from Vakharia and Wemmerlöv (1990). The authors measured flow time while serially moving machines (in order of machine number) from the original job shop to complete pre-defined cells. The significance of the order of their implementation was not tested. Each machine level of implementation was run assuming that the cell the machine created or joined enjoyed a level of both setup and run time reduction. This reduction level was controlled from 5% to 17.5% in equal 2.5% intervals. These "processing time reductions" were apparently applied as flat rates to all work within the cells and never to work completed within either the original job shop or any machines within the residual job shop. The processing improvements from Morris and Tersine (1989 and 1990) they cite do not include setup reductions due to family-based processing. The authors used a dispatching rule that minimized setup incidence (RL), yet did not disclose whether they followed a family-based setup structure. Another important detail left unspecified was the amount of setup relative to the run time of work within the factory. In Chapter 3, we relate these two by introducing the notion of a "setup potential" and show it to be a key factor in the total factory transformations. Kher and Jensen's (2002) results support those in Suresh (1991) that the cell flow improved, but non-cell residual worsened and the conclusion in Shambu and Suresh (2000) that the remainder shop flow time deteriorates as cells are added. By sending a machine at a time they also recognized the conclusion of Burgess et al. (1993) that the fraction of the factory load sent to the cells can be more than the load when in the JS to improve the performance of the residual job shop.

Table 2-6 mimics Table 2-4 in its structure and summarizes the key studies that considered partial cellular conversions.

Source	Factor(s)	Conclusions	Limitations
Shunk 1976	Cell formation technique	Optimal shop configuration may be PCMS hybrid, not pure JS or pure CMS	Number of machines in JS versus CMS not constant; no insights on PCMS
	Move times	Move time decreased with increasing number of cells	
Burgess, Morgan and Vollmann, 1993	Labor constraint	Reduced with increasing setup reduction	No setup discounting in residual JS, fixed setup discount applied to cell
	Magnitude of cell-based setup	Cell machines increase efficiency with increasing setup discount	
	Fraction of factory load sent to cell	Cell machines more efficient with setup reductions so optimum load is greater than JS fractional load	
Suresh, 1991	Number of cells	PCMS situations clearly unfavorable when compared to JS even with high degree of setup reduction: cell flow improved, but non-cell residual worsened	No setup discounting in residual JS, fixed setup discount applied to cell
	Setup discount in cells		

Table 2-6. PCMS Studies

Source	Factor(s)	Conclusions	Limitations
Shambu and Suresh (2000)	Dispatching rule	RL is better than SPT which is better than FCFS in job shop, remainder and cells	
	Batch size	Optimum remainder batch size minimized with job shop and increases with number of cells	
	Number of cells	Remainder shop flow time deteriorates as cells are added	Cell picks not optimized
Kher and Jensen, 2002	Number of cells (or partial cells)	Use of partial cells avoids formation of bottlenecks. Remainder shop flow time deteriorates as cells are added. Offloading demand from residual job shop to cells improves performance within residual	Single data set. Order of machines put into cells was not investigated.
	Processing time savings	Cells improve flow time and tardiness performance of parts they process (not necessarily the entire factory). Cells benefits increase with cell processing time savings	Setup to run time ratio not mentioned. Discounting in residual JS, fixed setup discount applied to cells

Table 2.6 (cont.). PCMS Studies

In summary, the PCMS literature suggests that cell and residual loading are both important to obtaining good overall factory flow times. An optimum load ratio between the cell(s) and residual has not been established with the objective of optimizing the factory flow time. Similarly, even though it has been acknowledged that cells are not always loaded consistently, the selection of cells to obtain the best flow time performance has not been pursued systematically in any of these studies. We address this issue in Chapter 4.

2.2 Two-Class, Single-Stage M/G/1

We present a summary in Table 2-7 of the contributions to flow time statistics of two-class models followed by details of the models. We list the arrival, setup and service distributions using “0” for zero, “M” for Markov, and “G” for general. The models use either first-come-first-served (FCFS) or an alternative dispatching rule, called *alternating priority* (AP) defined by Maxwell (1961) and others. The FCFS rule suffers from the drawback that setups are incurred based entirely on the random pattern of arrivals. In other words, no attempt is made to avoid setups. Maxwell (1961) and others have defined an alternative dispatching rule, called alternating priority (AP). Under this rule, all jobs in queue of a given class are served before switching to the other class. The server thus alternates between strings of jobs of either class 1 or class 2 and the idle state, but never switches from class i to class j $(j \neq i)$ if there are jobs of class i still in queue. Clearly, the AP rule is designed to minimize the incidence of setups. Finally, we list the system performance results from each model.

Source	Input Distributions Arrival	Setup	Service	Dispatch rule	Available Results
Maxwell, 1961	M	0	M	AP	Solution for mean number in system
Gaver, 1963	M	G	G	FCFS	Moment generating function for flow time
Avi-Itzhak, Maxwell and Miller., 1965	M	0	G	AP	Exact solution for flow time
Miller, 1964	M	G[1]	G	AP	Moment generating function for flow time
Eisenberg, 1967	M	G	G	AP	Moment generating function for flow time

[1]Setup forced at the conclusion of each machine idle period

Table 2-7. Single-server modeling contributions.

2.2.1 Single Queue. To analyze the impact of setups, we begin with one of the simplest exact models: a single-server queueing system with two classes of customers. Gaver (1963) provides results for this system under the FCFS rule. For the symmetric cases with equal arrival rates, setup times and service times Gaver provides a closed-form solution. We assume arrivals follow a Poisson process with rate λ_i for class i arrivals ($i = 1,2$), $\lambda = \lambda_1 + \lambda_2$ and with symmetry $\lambda_1 = \lambda_2$. The expected setup paid on switchover to the other class is $E(U)$. The expected service time paid per part is $E(S)$.

To obtain $E(U)$ consider a pair of successive arrivals and note that the occurrence of setup depends solely on whether these are of the same class or not. Let (i, j) describe the event that the first arrival is of type i and the second arrival of type j

$(i, j = 1,2)$. Then $E(U) = \sum_{i=1}^{2}\sum_{j=1}^{2} E[U|(i, j)]P(i, j)$ where $P(i, j) = \frac{\lambda_i}{\lambda}\frac{\lambda_j}{\lambda}$. Clearly, no setup is required if $i = j$ so $E[U|(1,1)] = E[U|(2,2)] = 0$. We then obtain $E(U) = \frac{\lambda_1\lambda_2}{\lambda^2}[E(U_1) + E(U_2)]$. The utilization to include expected setup is therefore $\overline{U} = \rho + \lambda\left[\frac{\lambda_1\lambda_2}{\lambda^2}\{E(U_1) + E(U_2)\}\right]$ where $\rho = \lambda_1 E(S_1) + \lambda_2 E(S_2)$ and for system stability $0 \le \overline{U} < 1$. Gaver's equation for the expected flow time assuming symmetry is $\overline{F} = \frac{\lambda}{4(1-\overline{U})}\left[E(S^2) + E\{(U+S)^2\}\right] + E(S) + \frac{E(U)}{2}$. To solve for the general flow time using the method of Gaver, we must use numerical methods to solve for a parameter that is a function of the λ_i's, $E(S_i)$'s, and $E(U_i)$'s.

2.2.2 Two Queues, One Server: Two Classes with Alternating Priority. This system can be modeled as a semi-Markov process (SMP) (see Wolff, 1989 p.220) and analyzed using fundamental results from renewal theory. It is customary to assume that the SMP is *regular* which it obtains if the state of the system at any time t is determined by a finite number of state transitions (jumps).

This type of problem is solved with renewal theory. If we define the states of a system such that their selection is Markovian, but allow the sojourn time in each state to be arbitrary then we have a semi-Markov process (SMP) with embedded, discrete-time Markov chain (EMC) transition probabilities (Wolff, p.221). For an EMC, the stationary probability of state j, p_j, represents the fraction of transitions that are visits to state j. The fraction of time spent in state j, π_j, is proportional to the transition fraction by $p_j \propto \pi_j / m_j$ where m_j is the sojourn time in state j. The time-average limit is $\lim_{t \to \infty} P_{ij}(t) = m_j / l_j = \pi_j$ where l_j is the mean recurrence time. As long as $\sum_{i \neq j} p_i m_i < \infty$ (noting that $1/m_j$ is the rate into or out of state j) then state j is *positive recurrent* enabling us to use: $\pi_j = \dfrac{p_j m_j}{\sum_j p_j m_j} > 0$ (Wolff, p.223) yielding the fraction of time the SMP spends in state j.

We start reviewing the two-class, single server model assuming zero setup and an alternating priority dispatch regime. Maxwell (1961) defines the states using a triple: the number of items of type-1 in the system, the number of items of type-2 in the system and an indication of the machine setup: 0 for idle, 1 for setup for type-1 and 2 for setup for type-2. This state definition loses the setup status of the machine upon entering the idle state, but this information is not required since setups are assumed to be zero. Maxwell then uses generating functions and relates the expected number of items of each type in the system to these generating functions. His resulting equation for mean number in the system is:

$$L = \frac{\rho}{1-\rho} + \frac{\rho_1 \rho_2 \left\{ \left[\dfrac{E(S_2)}{E(S_1)} - 1 \right](1-\rho_1) + \left[\dfrac{E(S_1)}{E(S_2)} - 1 \right](1-\rho_2) \right\}}{(1-\rho)\left[(1-\rho_1)(1-\rho_2) + \rho_1 \rho_2\right]}$$

where $\rho_i = \lambda_i E(S_i)$ and $\rho = \rho_1 + \rho_2$.

Avi-Itzhak, Maxwell and Miller (1965) computes wait times by conditioning on the job arrival class and the state of the system. A random arrival experiences a wait time based on the current class of work being processed. If the arrival is of the same class, then it must wait for the existing job to be completed as well as all jobs of its class ahead of it in line. If the arrival faces the server working on the other customer class, then it must wait for processing of all jobs of the other class to be completed as well as the jobs ahead of it of the same class. Flow times are calculated based on summing the conditional probabilities that the random jobs arrive within a specific block of time (a cycle). Fortunately, a closed-form solution is available for this infinite sum (number of potential cycles to consider). The type-1 mean flow time is:

$E(F_1) = E(S_1) + \dfrac{\lambda_1 E(S_1^2)}{2(1-\rho_1)} + \dfrac{\lambda_1 \rho_2^2 E(S_1^2) + \lambda_2 (1-\rho_1)^2 E(S_2^2)}{2(1-\rho_1)(1-\rho)((1-\rho_1)(1-\rho_2) + \rho_1\rho_2)}$. We note the similarity to the P-K formula: the first term is the service time, the second term is the wait due to FCFS within a cycle and the third term adds the expected wait for the other class of work to end its processing.

Miller (1964) modified the procedure of Avi-Itzhak et al. (1965) procedure to handle non-zero setups. Miller's model assumed setup at the beginning of every *busy period*, the unbroken work interval between idle periods, regardless of the type work ending the previous busy period. The mean flow time is computed by conditioning on the type of cycle a random arrival sees upon joining the system. The values of mean flow time are expressed analytically, but numerical methods are required to approximate the infinite sums encountered. Miller also showed that due to the reduced incidence of changeover in high traffic the system will not saturate if $\rho < 1$ where $\rho = \sum_i \lambda_i E(S_i)$, independent of the setup magnitude.

Miller (1964) uses a description of the system state that does not keep track of the last class served prior to an idle period for the machine. Since a setup is incurred at the start of each busy period, it is unnecessary to record this information in the state. Naturally if $\rho_1 >> \rho_2$, it may be that the job starting the next busy period matches the class of the last period before going idle. In such a case, Miller would assume that a setup occurs even though it is not required. In the case of equal Poisson arrival rates, the probability of two arrivals of the same type in succession (last of previous busy period and first of next busy period) is $2 \times \left(\frac{1}{2}\right)^2 = 0.50$. In the case that the busy period ends with equal probability of each type then we would expect that 50% of the subsequent busy periods would not need to start with a setup.

Eisenberg (1967) addressed the case of "setups as needed" by using a richer state description than Miller (1965). Eisenberg considers the embedded Markov process of queue lengths at the instant of service completion, and includes the class of service just completed. Thus, state ${}^{i}_{mn}$ denotes "server is at line i and m customers are waiting at line 1 and n customers are waiting at line 2." This state definition is event driven: it provides a snapshot of the system whenever a departure occurs. The idle states are exceptional in this regard: the probabilities of states ${}^{1}_{00}$ and ${}^{2}_{00}$ (the idle time) are the same for the imbedded and general-time probabilities. Solutions to his model also require numerical methods based on known values due to the existence of an infinite sum.

Eisenberg also provides three limiting cases. First, in the special case of zero setup times, he provides a wait equation that agrees with Avi-Itzhak et al. (1965). Next, when service times are assumed to be zero so that only setup remains, Eisenberg provides both the probability of idle with the machine ready to work on type-1, π_{00}^{1}, and the mean wait time. The last limiting case is for symmetry where the following are the same for both classes: $\lambda_i, E(S_i), E(S_i^2), E(U_i), E(U_i^2)$. The symmetric result is consistent with that of Avi-Itzhak et al., and the overall wait time is the same as for FCFS.

Sykes (1970), Eisenberg (1972) and Takagi (1990) investigate a different dispatching regime. They all assume that when a queue has been exhausted the server immediately switches over to the other customer type. Further, the server performs a setup upon switchover and this is done whether or not any jobs are present at the other queue. If there are no jobs waiting in that queue after the setup is complete then the server moves back to the other queue setting up again (again, whether or not there are jobs waiting). If a customer of class j arrives just as the server initiates a setup for class i and there are no class i present in the queue to be worked (and none arrive during the setup time) then he must wait yet another setup delay while the server is switched back to work on class j.

Cooper, Niu and Srinivasan (1999) show that some classes of state-independent setups (setting up whether or not work is waiting at that queue) yield equal or even less wait times than their state-dependent (setup only when there is work in the queue) setup classes. They consider a *switchover* time, the time required for the server to travel from queue $k-1$ to queue k, in addition to setup time (the time required to prepare for work at queue k) and processing time at queue k. If we assume in a manufacturing setting that the review time immediately after a service completion to consider if there are jobs immediately available for processing is zero then the analogous *switchover* time in Cooper et al. is zero. Left with only setup times and processing times, they concede that state-independent setup regimes are at best equal in expected wait time to their state-dependent counterparts and if any variability is present in the setup distribution then the state-dependent regime encounters less expected wait than its state-independent counterpart.

In summary, results for general setup and general service time typically require numerical methods due to the existence of an infinite sum term. Certain simplifications can be applied (as are done in cyclic models), but restrictions on setup variability quickly reduce the potential in suitability of such models in the o manufacturing environment. There still may exist rules between the extremes of state-dependent and state-independent that allow polling models to be adapted to manufacturing. For example, one can devise decision rules for setup incidence that consider the probability of customer arrival type within a given time interval that corresponds to idle time prior to committing to a setup.

Chapter 3

FACTORY CONVERSION TO CELLULAR MANUFACTURING SYSTEMS

The objective of this part of the research is to use a single simulation model to compare functional and cellular layouts across a test bed of factory environments extracted from the literature. In our attempt to perform such a comparison, we follow the established practice of most analytical or simulation conversion studies in using flow time as the primary performance measure. We use the *flow ratio* (*FR*), which we define as the ratio of the average batch flow time in the after cellular conversion to the average batch flow time of the job shop with the same factory operational parameters of load, machines and batch size. Therefore, measures below 1.0 indicate flow time superiority for the transformed shop.

It is well known that flow time deteriorates when the size of the machine pool is reduced, the *pooling loss*, as described in Chapter 1. Therefore, for the cellular layout to outperform the functional layout, the pooling loss must be compensated for by reductions in setup or move times. The key trade-off we consider is between pooling loss and setup reduction. While a number of well-known studies in the literature have studied this tradeoff, each has used its own data on demand, manufacturing capabilities, parts structures, and operating rules. This makes it difficult to compare the results across the disparate data sets. For this research, we have selected six studies from the literature that provide sufficiently specific information for our simulation model. We feel that these studies provide us with sufficient diversity in terms of the parts, machines, and operations, used in the manufacturing simulation. Having ensured that the same operating rules and measurement procedures apply to all data sets, we proceeded to choose a common range of key parameter values. We call this process standardization, although it may also be viewed as a focusing on a region of the parameter space where the six different data sets we selected can be compared. Of special importance in this standardization is the use of the same major-minor setup structure and identical operational rules across all data sets. This provides a level playing field for our simulation study.

3.1 Factory Environment and Notation

We now describe the main characteristics of the factory environment and introduce the notation used in our simulation study. Each data set specifies a set of available machines and a set of demands for parts. The demand is given as a set of parts, with associated operations sequences, part families, and demand levels. The set of parts is indexed by $i = 1, \ldots, I$. Each part i has a unique operations sequence consisting of $G(i)$ operations.

For each part i, the following information is available as input:

$V(i)$ = demand for part i in units/year

$G(i)$ = number of operations required by part i

k = operation index where $k = 1, \ldots, G(i)$

$O(i,k)$ = machine type required for the k^{th} operation of part i

$r(i,k)$ = expected run time for a single unit of i on its k^{th} operation

$s(i,k)$ = expected major setup time for i on its k^{th} operation

$b(i)$ = batch size for i

$f(i)$ = part family to which part i belongs

We assume that the demand for part i occurs in batches with mean $\lambda(i)$ defined as part demand divided by batch size, $\lambda(i) = V(i)/b(i)$. Sequential processing on the same machine type is combined within one operation sequence so that $O(i,k) \neq O(i,k+1)$ for all k.

In this research, we do not investigate the effect of move times on the conversion benefits in much detail. We argue that move times are negligible in cells due to the proximity of machines. In the job shop, move times may suffer due to congestion effects or limited transport resources. An investigation of this effect is beyond the scope of this research. However, we should note that if move times simply reflect known transport times, then their effect can be studied ex post as described later in this chapter.

3.2 Job Shop Operation

The job shop is configured in a functional layout with J departments, where department j houses the all the $NM(j)$ copies of machine type j. All machines are available 100% of the time at full capacity. Upon entry, each batch of part i immediately reports to the department required by the part's first operation $O(i,1)$. The batch then travels from one department to the next following its operations sequence, until all of its $G(i)$ operations are completed. The batch flow time is measured from part batch introduction into the factory (from receiving) to part batch leaving the factory (sent to shipping).

3.3 Standardization Scheme

An important theme of the present study is to pursue a dual objective. On the one hand, we wish to preserve the main characteristics of the various data sets as studied in the literature, since these do differ in such key inputs as the number of parts, number of part families, and the operations required by these parts. On the other hand, we wish to use uniform operating rules, and a comparable setup structure, batch size, and job shop load across all data sets. We believe that this is necessary to gain any general insights. For example, papers in the literature differ in how they account for setups in the job shop and the cells. We use the same setup structure and measure setups in the same way in both layouts. In what we call *operational standardization*, we ensure consistency in the flow control disciplines and adopt a common range of parameter as listed in Table 3-1. These values may be compared to Table 3-2, which lists the rules and parameters adopted by each of the sources used in our test bed. We now discuss and try to justify the choices for each of our baseline parameters.

3.2.1 Batch Size. We use a common batch size in the job shop for all parts. From the literature, we have noted that batch sizes are generally small for job shops. Batch sizes used for the job shops studied by Suresh (1991, 1992), Shambu (1993), Suresh and Meredith (1994), and Shambu and Suresh (2000) were 50 or less. We therefore used a range of 25-50 for our batch sizes. In this research, we do not use transfer batches within the job shop: Transfer batches make sense for cells where all machines are placed in close proximity of one another. This makes manual or automated machine-to-machine hand-offs reasonable. Job shop departments typically involve much longer distances and require material handling equipment to transfer goods. In the cell shop, we use a transfer batch size that is equal to b, $b/2$, or 1, where b is the original batch size used in the JS. Smaller values of batch sizes in the cells were used in the sensitivity runs.

3.2.2 Setup Structure. We use a major-minor setup structure whereby the setup is a major setup, a minor setup, or no setup at all. The same setup structure is used in both the job shop and the cell shop. The incidence of setups is tied to the family structure of parts types (recall that the I part types are partitioned into F families numbered $f = 1,\ldots,F$). A major setup is incurred if two parts belonging to distinct families are processed consecutively on the same machine. Switching between two different part types in the same family incurs a minor setup. Naturally, no setup is required if a machine processes two batches of the same part type consecutively.

3.2.3 Setup magnitudes. Past studies have shown that the relative magnitude of setups is an important factor in conversion studies [see Morris and Tersine (1990), Suresh (1991, 1992), Suresh and Meredith (1994), Shafer and Charnes (1997), Shambu and Suresh (2000).] We therefore control the setup potential, which refers to the amount of setup reduction that can be realized by cell conversion. *Setup potential* involves the choice of two parameters-- the setup ratio and the setup fraction. The *setup ratio* is the ratio of major setup, s, to batch run time, $b \cdot r$. The *setup fraction* is the ratio of minor to major setup. We standardize the setup ratio at 1.0. We selected 1.0 by considering the ranges used in earlier papers: Morris and Tersine (1989) use values ranging from 0.06 to1.0, while ranges of 0.4-2.3 and 0.3-6.0 are used in Yang and Jacobs (1992) and Suresh and Meredith (1994), respectively. We standardize the setup fraction at 0.20. This ratio is consistent with the simulation studies of Jensen et al. (1996) and within the range of setup fractions of 0.1-0.9 used in Garza and Smunt (1991) and Suresh and Meredith (1994).

3.2.4 Choice of Dispatching Rule. We use the *repetitive lot* (RL) dispatching rule across all departments. This rule is used to minimize the incidence of the setup paid and Jacobs and Bragg (1988) found this discipline superior to FCFS. Shambu and Suresh (2000) have confirmed its superiority to both FCFS and SPT in the job shop and cell environment with setups. It is also an appealing rule to use given our setup structure. The RL dispatching rule operates as follows:

1. A single (pooled) queue is formed for all batches arriving to be processed at a machine center.

2. Any arriving batch is immediately routed to the available machine where it encounters the least setup time. If no machines are available, the batch joins (or forms) a queue to wait for a machine.
3. When a machine becomes available, the next job assigned to it is selected based on the minimum setup among all jobs in queue. If multiple jobs tie at this minimum setup value, the FCFS discipline is used to break the tie.

3.2.5 Batch setup and run time

This choice specifies the magnitude of ($s + br$). While this value may depend on the part, the operation, and the machine type used, we standardize the part processing time by selecting distributions for the setup and run times. We use the $k-\mathrm{Erlang}(\beta)$ with $k = 2$ and β = mean of the setup or run time. We chose this distribution because it has less variability than the exponential (CV = 0.707 versus 1.0). Being non-symmetric (and skewed to the right), this distribution is more suitable for the time to complete a task (Law and Kelton, 1991, p.186; Pegden et al., 1995, p. 40). We provide results of other choices of distributions in Appendix A.

3.2.6 Factory Loading and Measurement. The overall level of utilization in the job shop has a major impact on the magnitude of pooling losses observed. Based on the studies used in our test bed, we use a target of 65% for the average machine utilization in the job shop. For examples, Morris and Tersine (1990) loaded their job shop at 60%-70%, Garza and Smunt (1991) used 60%, and Suresh and Meredith (1994) chose 70% for their job shops. Values of other studies appear in Table 3-2. We reach our target utilization by adjusting the overall factory demand (retaining relative product mix ratios) until the ex-post utilization value reported by the simulation lands within 2% of this target value. A summary of standardized parameters is in Table 3-1.

Factor	Proposed standard
Batch Size, *b*	25 to 50, fixed for all parts
Transfer Batch Size	*b*
Part Batch Arrival Rate Distribution	Poisson, CV=1.0
Setup Time Distribution	2-Erlang, CV=.7
Run Time Distribution	2-Erlang, CV=.7
Setup Ratio = s/br	1.0, fixed for all part operations
Setup Structure	identical = 0 setup distinct within same family = minor setup distinct families = major setup
Setup Fraction = Minor/Major Setup	0.2
Dispatching Rule	repetitive lot (RL)
Material Handling	unconstrained capacity, 0 move time
Labor	unconstrained
Job Shop Average Machine Utilization	65% ±2%
Machines	100% available at all times

Table 3-1. Choices and parameters values for operational standardization.

3.2.7 Formation Standardization. We expect conversion results to be sensitive to the particular choice of cells. The configuration of cells formed must therefore be closely monitored. In formation standardization, we ensure that all data sets use the same cell formation technique. While there is a vast literature on cell formation techniques (e.g., Singh and Rajamani, 1996), our interest is to choose a single algorithm that we can apply to all six data sets. We chose the cell formation procedure due to Vakharia and Wemmerlöv (1990) because it considers both sequences and capacities, factors that are left out in earlier cell formation techniques.

Vakharia and Wemmerlöv's method first groups parts by the commonality in their operations sequences and then proceeds to assign machines to such groups to provide sufficient capacity to meet demand.

In what follows, the *standardized cell configuration* refers to the design produced by the Vakharia and Wemmerlöv algorithm (V-W) when applied to each data set. This procedure generally results in cells that differ from the CMS configuration in the original data source. In fact, differences in the number of cells or number of machines of each type can both arise. In any case, for each data set, we run the simulation model twice, once for each cell configuration (source and V-W).

3.4 Choice of Data Sets

One of the objectives of this research on factory conversions is to use a single simulation model to run all the data sets in the test bed we selected. Since sources of these data sets (as published in the literature) refer to different factory environments and/or modeling assumptions, the uniformity required for the inputs to our simulation model is not easily obtained. Of the 24 data sets cited in the Johnson and Wemmerlöv (1996) overview of modeling studies, we used six in our simulation studies because they provided information specific enough for our model. We supplemented these with two data sets from Morris (1988).

We require four eligibility conditions in selecting data sets for our study.

1. The original data source must provide a cell configuration; the number of cells as well as the assignment of machine types and parts to each cell must be specified,

2. The cell configuration provided must not require inter-cell moves,
3. The number of machines of each type must be specified for both job shop departments and each cell, and
4. At least one machine type must have more than one copy in the original functional layout.

Thus, condition (3) excludes a number of data sets in the literature that form cells based on part-machine incidence, but do not unambiguously define the machine types used. A number of data sets were eliminated by condition (4).

In constructing our test bed, we sought data sets that provided some details on operations sequences, setup and run times, arrival and processing distributions, and available machines as in Table 3-1. Our final test bed therefore uses six data sets from eight sources in the literature (see details in Table 3-2) - all but two were used in prior simulation studies by their authors. None of the authors provided an explicit description of the cell formation technique they employed to configure their cell shop. The source for data sets 2 and 3 does not provide simulation results for these data sets. However, this source does supply the required part and machine structure along with a cell solution; we generated the balance of the operational data.

Table 3-2 lists the operational settings for all data sets as provided in the original papers. A glance at this table shows considerable differences among these settings, arguing the case for standardization. Table 3-3 shows the data sets after standardization.

Table 3-2. Data sets used in analysis as reported by source (blanks denote omissions by source).

Source	Garza and Smunt, 1991	Morris, 1988: data set #2	Morris, 1988: data set #5	Suresh and Meredith, 1994	Yang and Jacobs, 1992	Morris and Tersine, 1989	Morris and Tersine, 1990	Shafer and Charnes, 1995
Data Set ID	1	2	3	4	5	6a	6b	6c
Parts	60	24	45	50	18	40	40	40
Operations/part	4	2-4	2-6	3-7	1-4	2-6	2-6	2-6
Machine Types	8	6	14	10	4	8	8	8
Machines	24	20	35	31	10	30	30	30
Machines/type	3	3-4	2-3	3-4	2-3	3-4	3-4	3-4
Cells	6	4	4	5	3	5	5	5
Batch size, b	10, 15, 20, 25			JS: 32-100 CM: 5-100	1	50	50	50
Setup ratio: s/br	.05-.35			0.3-6.0	0.4-2.3	0.06	0.06 or .1, .5, 1.0	0.06
Setup fraction: min/maj setup	0.1-0.9			0.1-1.0	0.5	0.5	0.5	0.5
Dispatching rule	FCFS			FCFS or RL-F in JS, FCFS in cells	RL	RL	RL	FCFS
Source JS avg. mach utilization as measured by simulation	60%			70% with b=50	56-86%	44%	60-70%	
Cell transfer batch size	b			b	b	b, 1	b	1
Arrival rate distribution (CV)	deterministic (0)			Poisson (1.0) 3-Erlang (.58)	Poisson (1.0)	Poisson (1.0)	Poisson (1.0) deterministic (0)	Poisson (1.0)
Setup time distribution (CV)	deterministic (0)			Poisson (1.0) 3-Erlang (.58)	Poisson (1.0)	Normal (0.08)	Normal (0.08)	Normal (0.08)
Part run time distribution (CV)	0, .33, .66, 1.0			Poisson (1.0) 3-Erlang (.58)	Poisson (1.0)	Normal (0.36)	Normal (0.36)	Normal (.01 per batch)
Material handling times	$2r$ or $10r$ between depts., 0 within cells			$3r$ - $120r$ between depts., $.75r$ within cells	$.7r$ between depts., 0 within cells	5mph + 3 min load/unload between depts., 0 within cells	~$.15r$ between depts., 0 within cells	r between depts., 0 within cells
Unique features of data set	(1)	(2)	(2)	(3)	(4)	(5)		

b, batch size; r, run time per part; CV, coefficient of variation

(1) no minor setup in JS, assumes minor setups in cells due to tooling

(2) not simulated by author

(3) part to same part type required minor setup, included run time productivity improvement factor

(4) designed to test MRP vs. Period Batch Control order-release-and-due-date-assignment systems

(5) cell must be empty before setup changeover

In examining Table 3-2, we particularly focus on five factors that are important to us in this study: batch size, setup ratio, setup fraction, dispatching rule and job shop loading. There was a wide range of setup ratio. Some studies (1, 5 and 6b) evaluated the same shop over a range of setup ratios. In the case of study 4, the setup and run time per part were fixed so when the authors varied the batch size the setup ratio changed, too. The setup fraction reflects the setup discounting for similar batches processed in sequence. Studies 1 and 4 tested for this factor explicitly, while the others used a midpoint value of 0.5.

When using simulation to evaluate their factory performance, each source selected a certain load for the job shop, and then replicated the same demand for the cell shop. The average job shop machine utilization varied from 44% to 86% from data sets 6a and 5, respectively. The authors in study 4 chose the $b = 50$ case for their job shop standard for comparison, which resulted in an average machine utilization of 70%.

Our experiments focus on machine-constrained environments; we do not consider labor constraints. When labor and machine are both limited, then the conversion study must study the interaction between these two factors as illustrated in Suresh (1993) and Morris and Tersine (1994). In fact, labor constraints were absent from all studies in Table 3-2, except for study 6c.

For convenience, we report the material handling time included by some of the studies for travel between departments. When included, the time varied from 15% of a single part run time, or 0.15*r* (data set 6b) to 120*r* (data set 4), with the average being approximately *r*. Material handling was always assumed unconstrained so travel time and not time due to material handling congestion was included.

Upon standardization, data sets 6a-6c collapse into a single data set in our test bed identified simply as data set 6. Table 3-3 contains the final standardized values of the parameters in our test bed.

	Data Set ID					
	1	2	3	4	5	6
Parts	60	24	45	50	18	40
Operations/Part	4	2-4	2-6	3-7	1-4	2-6
Machine Types	8	6	14	10	4	8
Machines	24	20	35	31	10	30
Cells	6	4	4	5	3	5
Batch Size (*b*)	25	25	25	32	25	50
Cell Transfer Batch Size	*b*					
Arrival Rate Distribution (CV)	Poisson (1.0)					
Setup Time Distribution (CV)	2-Erlang (.7)					
Part Run Time Distribution (CV)	2-Erlang (.7)					
Setup Ratio (*s*/*br*)	1.0					
Setup Fraction:min/maj setup	0.2					
Dispatching Rule	RL					
Material Handling Times	0					
JS Average Machine Utilization	65%					

Table 3-3. Data sets characteristics after operational standardization.

We did not expect the standardized formation technique to provide the same cell configurations as specified in the sources. Table 3-4 lists the differences between configurations in the source and standardized designs.

	Data Set ID					
	1	2	3	4	5	6
Number of cells from source	6	4	4	5	3	5
Number of cells using *standardized formation*	6	4	4	5	2	5
Machines from source	24	20	35	31	10	30
Machines using *standardized formation*	24	20	41	32	10	32

Table 3-4. Comparison of cell designs in source and standardized configurations.

3.5 Metrics and the Simulation Model

The primary metric for the simulation model is average batch flow time. The simulation also tracks key explanatory output measures including average batch setup and machine utilization. While the simulation model is capable of measuring move time, we do not do so here based on our standardized move time of zero. The expressions used to calculate these measures are listed in Appendix B.

We evaluate all six data sets with the same simulation model. Our model was designed to possess sufficient generality to apply to both job shop and cellular configurations. Each data set was first run in its job shop configuration using the operational standardization. We then evaluate the CMS layout following the cells designs provided by the data source and ensure that the CMS run uses the same relative part volumes as the job shop configuration. In keeping with recent industry survey results (Marsh et al., 1999), we allow for a remainder cell to process non-related parts.

Each experimental condition tested was first warmed-up from an empty factory for a period long enough for the WIP to stabilize via inspection of time series plots as developed by Welch (1983). The end state of the warm up period was saved and used for initial conditions for each of 100 replications starting with different random number seeds to avoid autocorrelation. Each replication was run long enough for each part type to have at least 250 completed batches in order for arrival and service distributions to be adequately represented in the results. For example, data set 2 containing 24 parts and 100-minute flow times was run for approximately 100,000 simulated minutes per replication. The same set of random number seeds used across replications was used across data sets to reduce variability. Typically, testing a single data set required 300 simulation runs (each data set run at three levels and replicated 100 times). The comparisons between job shop and CMS flow times under operationally standardized conditions as listed in Table 3-5 are all based on this run length and $n = 100$ replication scheme.

We list both the mean and standard deviation of each statistic in the tables that follow. The mean for each statistic is calculated as $\bar{x} = \sum_i x_i / n$ where each replication provides a data point and n is the number of replications. The standard deviation is then calculated as $s(\bar{x}) = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n(n-1)}}$. This data is sufficient to then calculate confidence intervals. The confidence interval using the t-test as outlined in Pegden, Shannon and Sadowski (1990, p.177) is calculated as $h = t_{n-1,1-\alpha/2}\, s(\bar{x})$.

The simulation model is written in GPSS/H (Schriber, 1974; Henriksen and Crain, 1989). The model was run on a 266 MHz AMD K6-based PC running Wolverine Software's GPSS/H Professional (Release 3: 1995). The execution time per replication per level for each data set was roughly two minutes and equal since each shop was loaded at the same level of congestion.

3.6 Simulation Results Comparing Functional and Cellular Layout

Our goal is to measure the results of conversion and to evaluate their consistency across data sets. Prior to showing overall flow time results, we examine the measured setup reduction resulting from the conversion to CMS. We then use this information as well as congestion effect to explain the overall flow time results.

3.6.1 Setup Reduction Effect. We expect a significant reduction in setups as we convert to CMS since major setups are eliminated whenever a part family is assigned to a single cell. Tables 3-5 and 3-6 list the average setup time per batch for both CMS and JS layouts as reported by the simulation output. In these tables, the setup is measured as a fraction of the JS flow time per batch (which is normalized to 1.0 for each data set). Each flow ratio data point is the ratio of the average batch flow time of the transformed shop to the original job shop for the same replication. The setup reduction is calculated as (1 – transformed shop setup/job shop setup)*100% for each replication. We observe in Table 3-5 that the setup reduction is very consistent across data sets and ranges from 69% to 77% with an average setup reduction of 73% per batch. The confidence interval using the t-test is calculated as $h = t_{n-1,1-\alpha/2}s(\bar{x})$ so for the setup reduction for data set 1, $h = (1.9842)(0.001) = 0.00198$. We therefore have 95% confidence that the true mean is within 0.00198 of 0.69 or roughly within 0.3% of our estimate of 69% (0.00198/0.69). Table 3-6 lists the results when formation standardization is used for each data set. We get similar results indicating that the standard cell configuration can also reduce setups significantly.

Operational Standardization

Data set	JS setup mean	JS setup stdev	CMS setup mean	CMS setup stdev	Setup reduction mean	Setup reduction stdev	Flow ratio mean	Flow ratio stdev
1	0.293	0.002	0.090	0.001	69%	0.001	0.72	0.003
2	0.286	0.003	0.066	0.001	77%	0.002	0.87	0.010
3	0.201	0.004	0.060	0.001	70%	0.002	0.89	0.013
4	0.299	0.002	0.085	0.001	72%	0.002	0.78	0.005
5	0.267	0.001	0.069	0.000	74%	0.001	0.80	0.004
6	0.322	0.002	0.078	0.001	76%	0.001	0.82	0.006
				average	73%		0.81	

Table 3-5. Setup reductions and associated flow ratios for Operational Standardization

Formation and Operational Standardization

Data set	JS setup mean	JS setup stdev	CMS setup mean	CMS setup stdev	Setup reduction mean	Setup reduction stdev	Flow ratio mean	Flow ratio stdev
1	0.295	0.002	0.105	0.001	64%	0.001	0.71	0.004
2	0.284	0.003	0.075	0.001	74%	0.002	0.86	0.008
3	0.250	0.004	0.079	0.001	68%	0.002	0.99	0.006
4	0.303	0.002	0.133	0.001	56%	0.002	0.93	0.007
5	0.149	0.004	0.069	0.003	60%	0.003	1.15	0.069
6	0.307	0.003	0.088	0.001	71%	0.003	0.92	0.014
				average	66%		0.93	

Table 3-6. Setup reductions and associated flow ratios for Formation and Operational Standardization

3.6.2 Overall Flow Time Improvements. To compare flow times, we ran each data set with the source and the standardized cell configurations. The results appear in Table 3-5 and Table 3-6, respectively. The setup reduction realized in the cells tended to overcome pooling losses to outperform job shops by an average of 19% corresponding to a flow ratio of 0.81. The confidence interval using the t-test is calculated as $h = t_{n-1,1-\alpha/2} s(\bar{x})$ so for the flow ratio for data set 1, $h = (1.9842)(0.003) = 0.00595$. We therefore have 95% confidence that the true mean is within 0.00595 of 0.72 or roughly within 1% of our estimate of 0.72 (0.00595/0.72). The formation standardization results show an average improvement of seven percent from conversion corresponding to a flow ratio of 0.93. This 7% average improvement increases to 13% if we exclude data sets 5 and 6 containing bottlenecks (see Figure 3-1 for high utilization levels for these data sets). We remind the reader that standardized formation results in changes to the number of cells and/or machines as shown in Table 3-4.

It is useful to compare our results with the findings of Suresh (1991) who investigated the level of setup required to overcome the pooling loss (Suresh calls this the *breakeven* δ). Using an analytical model, Suresh (1991) reduced the magnitude of the setup (s) in the cells by 80% to overcome the pooling loss. A CMS with this level of setup reduction will then have the same flow time as the job shop. The results of our tests are more favorable to CMS. We show an average improvement of 19% in flow time with a corresponding setup reduction of 72%. We should note that the 80% figure cited from Suresh (1991) corresponds to a simulation example using FCFS, *no* setup discounting in the job shop, and a flat-rate discount in the cells. If we look for operating assumptions closer to ours, we should consider

Suresh’s family-based setup configuration for the job shop. The conversion of this configuration to cells (using the same 80% setup reduction and a lot size of 20) indicated an improvement of 22%, which is more consistent with our simulation results.

To gain some insights into the flow times reported in Table 3-5 and Table 3-6, we can examine the changes in machine utilization in greater detail. Figure 3-1 shows the average overall utilization levels for JS and CMS for each of the six data sets (labeled on the horizontal axis). Also shown are the maximum and minimum average utilization levels realized across all machine types. As expected, the average utilization for the job shop stays close to the target line of 65%. This is because we adjust the load on the JS to attain this target utilization within two percent. The simulation output shows that the average utilization after conversion to CMS is 48% (this is the lower dashed line in Figure 3-1). Thus, on the average, conversion yields an overall reduction of 17% in the average machine utilization.

Next, we examine the utilization levels by machine type. Since conversion involves segregating pools of machines in departments into cells, imbalances may arise readily unless the cell formation technique takes capacity issues carefully into account. In fact, the range of machine utilization (computed as the difference between maximum and minimum levels) increases eight percent when the JS is converted to CMS using the source formation technique reflecting the machine loading imbalance. The standardized cell formation technique produces a wider range (25% as compared to eight percent for the source configuration).

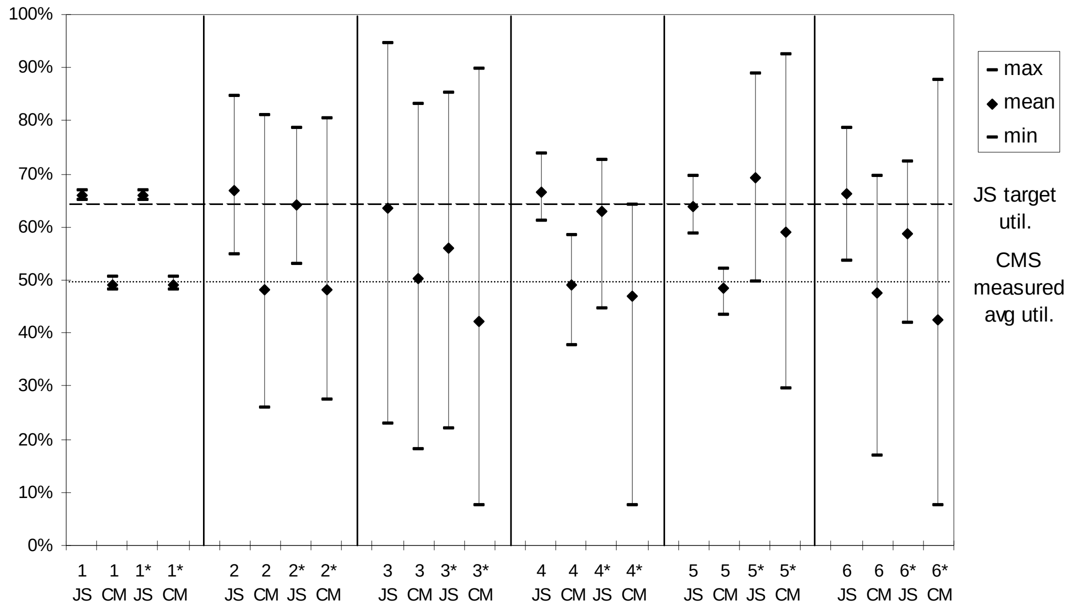

Figure 3-1. Comparison of machine utilization for JS and CM
(the asterisk refers to standardized formation).

In eight of the 12 results tabulated (5 out of 6 from source and 3 out of 6 for standardized formation), conversion succeeds in reducing both the average and the maximum utilization. These are the cases that show favorable flow time reductions in Table 3-5. It is worthwhile to examine the other four cases where the maximum utilization has not been eased: 2*, 3*, 5*, and 6*. First we note that machine types utilized less than 65% in the job shop did not have utilization levels exceeding 65% in any of the cells. We therefore provide additional utilization detail form those machine types that are utilized more than 65% in the job shop. As seen in Figure 3-2, each of the four cases where the maximum utilization is not reduced exhibits a bottleneck in at least one of the cells. Such bottlenecks arise simply because of the way machines may be distributed among the cells during cell formation.

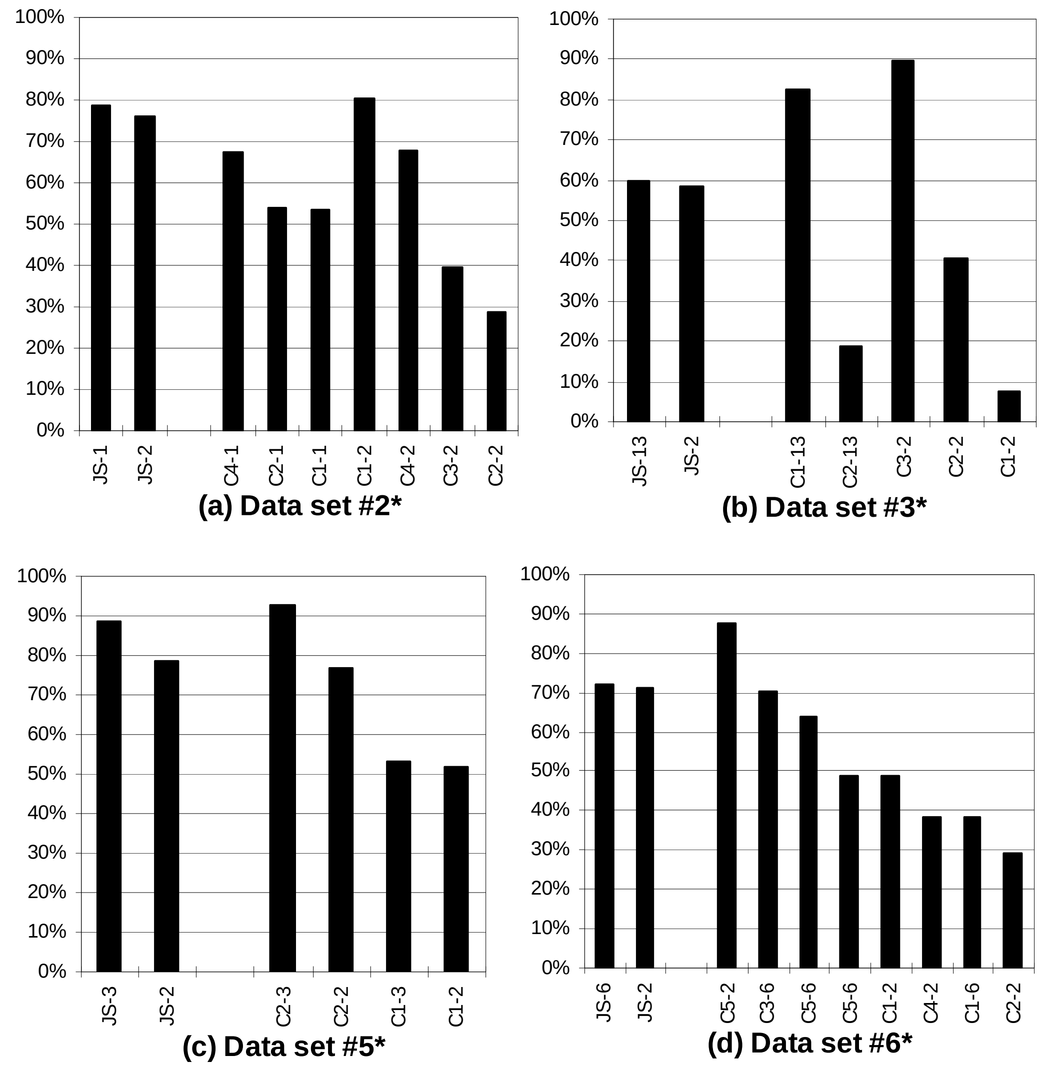

Figure 3-2. Simulation results for machine types with utilization above 65% in the JS layout. In the job shop, JS- j denotes machine type j. Within cells, Cc - j denotes machine type j in cell c.

For example, in data set 2 (using standardized formation), we single out machine types 1 and 2 in the JS since their utilization exceeds 65%. Additionally, we show the utilization for these two types wherever they occur in the cells. It is clear from Figure 3-2 (a) that the utilization of machine type 1 is reduced in cells 1, 2 and 4, but machine type 2's utilization has increased relative to the job shop to 81% in cell 1 and is reduced in cells 2, 3 and 4. The parts being processed in cell 1 requiring machine type 2 experience severe congestion resulting in a high flow ratio for the entire factory as seen in Table 3-5. Example (b) through (d) in Figure 3-2 show similar bottlenecks in data sets 3, 5 and 6, when standardized formation is used. In summary, these examples shows that bottleneck effects can dominate the results on flow time in a way that cannot be captured by system-wide average utilization alone.

3.7 Sensitivity to Key Operational Factors

In this section we investigate the sensitivity of flow time to four key factors. First, we evaluate the effect of using smaller batch sizes or transfer batches in the cells. Next we evaluate the effect of job shop loading. Then we study the sensitivity to the two key parameters of the setup structure. Finally, we compare the effects of the dispatching rule.

3.7.1 Batch Size Reduction and Transfer Batches. Our results of the last section matched the batch size in the cells with the original batch size used in the job shop. However, previous research (e.g., Suresh, 1991) shows that the setup reductions realized allow us to use smaller batch sizes in the cells than in the job shop and that this can have a profound effect on the flow time of cells. Moreover, cells can also make the use of smaller transfer batches possible, since machines are located in close proximity in cells. We therefore study two changes in the cells: (a) cutting the batch size to half its original value, and (b) use of transfer batches of size one. The first choice should provide a good idea of how a 50% reduction of batch sizes affects the CMS. The latter tests the extreme case of unit transfer batches to assess the maximum potential benefits small transfer batches are capable of producing (from Wagner and Ragatz, 1994, we know that moving to smaller transfer batch sizes within cells continues to produce benefits when no additional setup is incurred).

Table 3-7 compares the flow ratios for the job shop with batch size b and the CMS under four settings: the original batch size *b*, the reduced batch size $b/2$, and transfer batches of size one used with either *b* or $b/2$ as the batch size. In all cases, the flow time improves when a smaller batch size or a transfer batch of size one is used.

Data set	*b*	JS to CM baseline *b* mean	JS to CM baseline *b* stdev	JS to CM reduced *b* mean	JS to CM reduced *b* stdev	JS to CM baseline *b* with TB = 1 mean	JS to CM baseline *b* with TB = 1 stdev	JS to CM reduced *b* with TB = 1 mean	JS to CM reduced *b* with TB = 1 stdev
1	25	0.72	0.003	0.46	0.003	0.37	0.001	0.28	0.001
2	25	0.87	0.010	0.56	0.016	0.57	0.015	0.41	0.012
3	25	0.89	0.013	0.63	0.011	0.61	0.010	0.49	0.013
4	32	0.78	0.005	0.50	0.004	0.38	0.003	0.30	0.003
5	25	0.80	0.004	0.50	0.003	0.51	0.004	0.36	0.003
6	50	0.82	0.006	0.52	0.007	0.45	0.004	0.34	0.005
average		0.81		0.53		0.48		0.36	

Table 3-7. Flow times in cells with smaller batch size or transfer batches (JS flow time with batch size *b* provides baseline of 1.00).

For example, in data set 1, direct conversion to CMS reduces the flow time by 28% (flow ratio is 0.72) as compared to the job shop even when the same batch size is used. The use of batch size of $b/2$ provides an additional improvement of 26% (0.72-0.46 = 0.26), the use of unit transfer batches with the original batch size provides a 35% (0.72-0.37 = 0.35) improvement over the advantage of conversion alone.

Overall, the batch size reduction to $b/2$ improves upon the advantage of conversion alone by 28%. Using transfer batches of size one in the cells provides an average improvement of 33% over direct conversion (CMS with batch size b). However, if the batch size is already reduced, this improvement averages 17%. Interestingly, starting with a batch size of b in the job shop, the two alternatives of reducing the batch size to $b/2$ or using transfer batches of size one but retaining b in the cells produce comparable benefits (0.53 or 0.48). These results are of the same magnitude as those reported by Smunt et al. (1996) where transfer batches of size one were used in the first of four stages.

We also expect the improvement from using transfer batches to increase with the number of operations per part. Figure 3-3 illustrates this relationship for data set 6. The vertical axis of Figure 3-3 shows the additional improvement in flow ratio due to transfer batches, as compared to CMS without transfer batches.

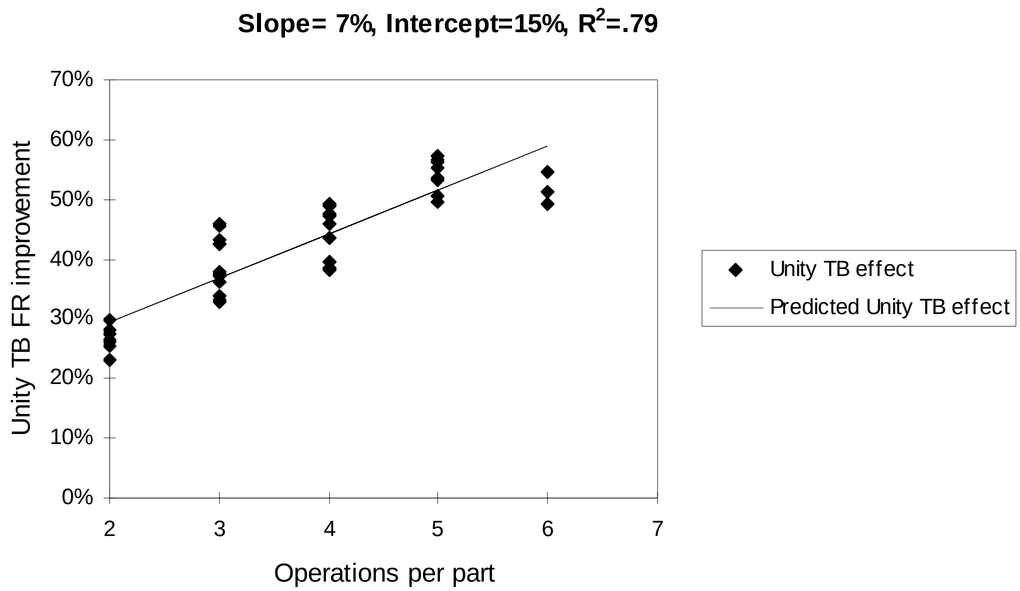

Figure 3-3. Flow time improvement using unity transfer batches as a function of operations per part (data set 6).

3.7.2 Job Shop Loading Sensitivity. Our computational runs have shown that pooling loss must be linked to the manufacturing load. As mentioned previously, bottlenecks may occur as the pooled resources of the job shop are segregated into cells. If such bottlenecks occur, their effect on flow time will be more pronounced as the overall utilization increases.

We use data set 2 (using standardized formation) to illustrate the case where the average machine utilization is reduced as a result of conversion, but the maximum machine utilization deteriorates in the CMS. For this data set, we varied the level of utilization from 55% to 85% and ran the simulation repeatedly. The results appear in Figure 3-4. Recall that the JS utilization sets the level of demand since the relative part demands are adjusted until the average machine utilization gets within 2% of the desired utilization value. Utilization levels above 85% could not be tested for using this data set since the maximum utilization in the CMS reaches 100%. We see that the flow time suffers in the CMS when the job shop is loaded at 85%, but for machines with lower utilization (in the 65% ±10% range), the effect on flow time is modest. This example shows a point we have observed in other data sets: the flow time in CMS is more sensitive to machine utilization than in JS. Therefore, cell layouts may not exhibit superior flow times if bottlenecks appear.

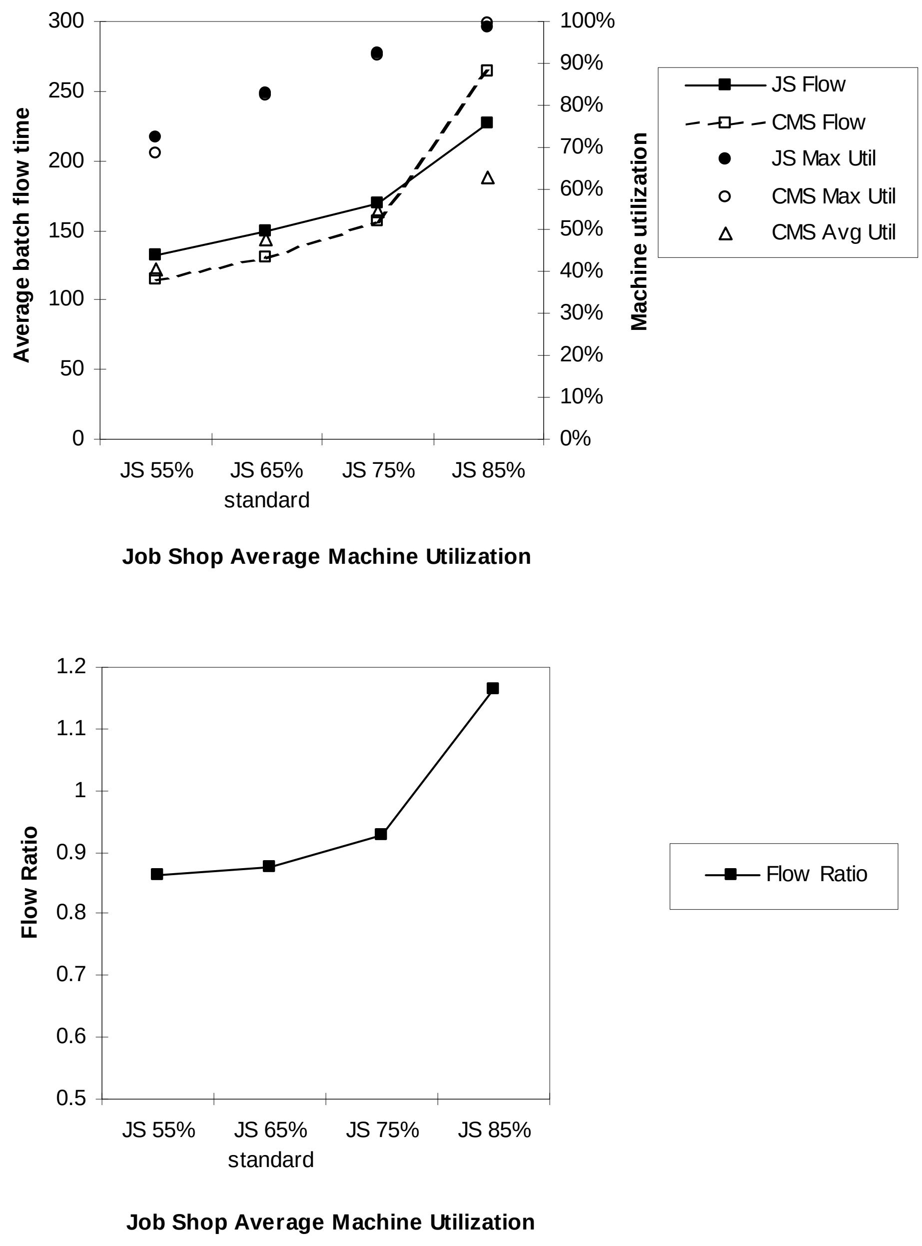

Figure 3-4. Job shop loading sensitivity (data set 2).

3.7.3 Setup Potential. We tested the sensitivity of flow time to the setup potential by varying both the setup ratio and the setup fraction. We ran all nine combinations of the two factors with three levels per factor. The highest potential occurs when the setup ratio $s/br = 2$ and setup fraction equals 0.1, while the lowest potential occurs at the pairing (0.5, 0.4). We chose data set 2 to perform the setup sensitivity runs. We kept the batch size (b) and part processing time ($s + br$) constant when varying the setup ratio (s/br) and ran each experiment at the standard 65% target average machine utilization.

We expected the (2, 0.1) setting to produce results better than the standard (1, 0.2) setting and expected the CMS flow ratios to increase as the potential for setup reduction is lowered. The results in Figure 3-5 are consistent with this expectation: the lowest flow ratio corresponds to the highest setup potential.

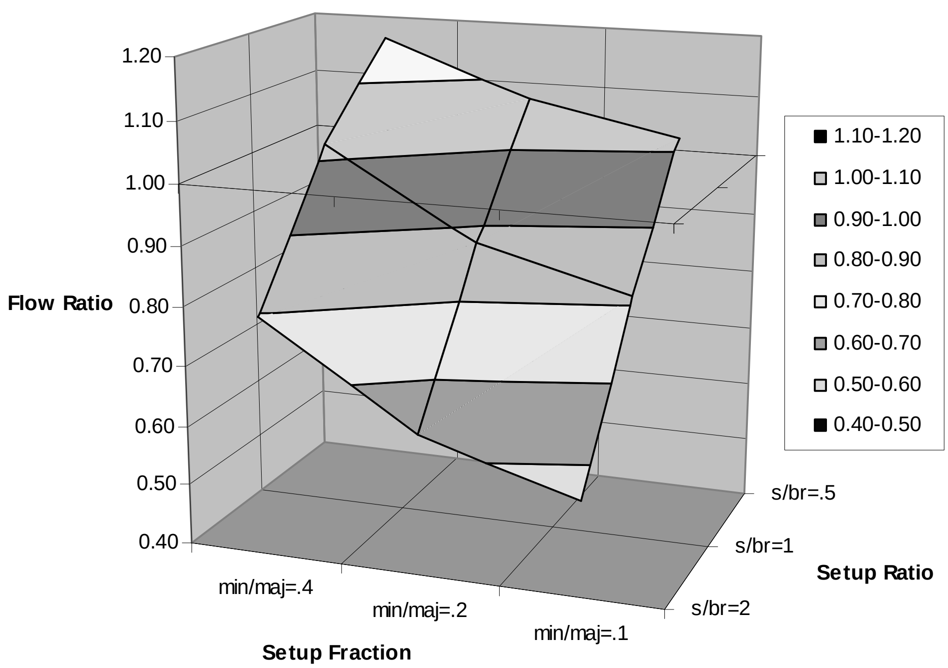

Figure 3-5. Response of the flow ratio to the two setup parameters.

3.7.4 Dispatching Rule. Although we chose the repetitive lot (RL) dispatching rule for our analyses, we recognize that not all shops may use a rule tailored to minimize the incidence of setup. We, therefore, compare the use of this rule to first-come-first-served (FCFS) dispatching to understand the dispatching rule's effect. We chose one data set from the six (data set 1) and evaluated its flow time at both the JS and CMS layouts at a common level of demand using the same simulation model. We set the factory load using the same method as before, but used the FCFS job shop as the basis: we measured the average machine utilization and then set the demand relative to the original demand mix such that the average was within two percent of 70%. We chose a slightly different value for the target to keep them clear from the results of the conversion study above. Since these dispatching rules directly affect the incidence of setup, we list more detail simulation measurement results in Table 3-8. As before, we report the average setup as a fraction of the average JS flow time, but here for a given dispatching rule. We include detailed measures of the incidence of setup paid: none, minor and major. We do this because it enables us to separate setup time incurred (which the reader will recall is a function of the ratio of minor to major setup) from the setup incidence. The flow times are listed along with the calculated flow ratios. Finally, the average machine utilization measures are listed (the range data for utilization is similar to that shown in Figure 3-1 above).

			Dispatching Rule			
			FCFS		RL	
			mean	stdev	mean	stdev
Flow Time		JS	1290	12.34	1169	6.03
		CMS	844	2.63	837	2.28
		Flow Ratio	0.65	0.005	0.72	0.003
Setup		JS	0.295	0.002	0.293	0.002
		CMS	0.084	0.001	0.090	0.001
		Setup reduction	72%	0.001	69%	0.001
Setup Incidence	None	JS	5%	0.001	6%	0.001
		CMS	10%	0.002	12%	0.002
	Minor	JS	40%	0.003	46%	0.002
		CMS	90%	0.002	88%	0.002
	Major	JS	55%	0.003	48%	0.002
		CMS	0%	-	0%	-
Average machine utilization		JS	68%	0.005	66%	0.004
		CMS	49%	0.003	50%	0.003

Table 3-8. RL dispatching avoids more major setups in the job shop than FCFS.

The RL flow time in the JS is 9% lower than when using FCFS (1169 versus 1290). If we look first at the setup, the impact of either rule seems to be similar. The fraction of flow time in both the JS and CMS as well as the setup reduction are all within five percent across dispatching rules. They are, however, fractions of their respective job shop flow times so the FCFS setup is 0.295*1290=381 and the RL setup is 0.293*1169=343. This difference is significant with >95% confidence since the mean difference between the FCFS and RL setup times (381-343=38) is within 1% of its estimate using the paired-t test. The setup incidence reveals that RL requires fewer major setups (48% as compared to 55%). The reader will recall from Table 3-1 that this data set contains 60 discrete part types that make up six part families. The average queue size (not shown in Table 3-8) for the FCFS job shop is

0.64 so it is not surprising that the RL dispatching rule rarely has an opportunity to bring forward a like part from the queue to process in sequence. Although RL isn't able to leverage part-to-part sequencing often, it is able to leverage the common family parts currently in queue generating more minor setups (46% versus 40% using FCFS). The lower utilization measure is a direct result of the reduced setup paid using RL. The range of machine utilization across the machine types is roughly unchanged.

Once the factory is converted to cells, there seems to be little flow time advantage to RL over FCFS. This may be because the major setup reduction is complete and no longer a factor. This particular data set has 10 parts per part family and the average queue size in the cells for FCFS (and RL) was 0.10. The FCFS rule in the cells paid a minor setup 90% of the time (which corresponds to the number of discrete part types per cell). Therefore, for RL to improve upon FCFS there must be more than one part in queue (and of the same type being processed) so the dispatching rule can pull it forward and avoid the minor setup.

3.8 Move times

While we do not focus on move time effects in this research, it is useful to briefly explore the magnitude of this effect. We note that when move times are known and not subject to congestion, these times can be added in ex post. We evaluated this effect for data set 2 with 31 total machines, 10 machine types, 50 part types, and five part families forming five cells (Suresh and Meredith, 1994). We set the move time equal to $\alpha(s+br)$, where α is a multiplier that we can vary, so that the move time is proportional to the standard processing time per batch (major setup plus batch run time). We used this time every time work was transported between a pair of departments in the job shop. Since we assume that move times in the CMS are negligible, the flow ratio should improve as α increases. The value $\alpha = 2$ corresponds to the high level of move time used in Suresh and Meredith (1994). We found that the flow ratio improves 12% each time α is increased by 1. The move is therefore an independent compensatory factor that can be used to overcome pooling loss. But the preceding example shows that the magnitude of move times has to be significant (compared to the batch run time) for it to have an impact.

3.9 Discussion on Dispersion of Simulation Results in the Literature

We now return to the issue that motivated this study: the large dispersion in the results of simulation studies that compare functional and cellular layouts as shown previously in Figure 2-1. In Figure 3-6 we add our results. The topmost bar of Figure 3-6 is reserved for the results of our test bed of six data sets. It is immediately clear that the range of results for our runs is narrower than the results of most of the other studies and lies consistently in the band that favors CMS. This remains true even when we compare our results to the first group of bars in Figure 3-6 that represent the sources of data for our test bed. This shows that standardization can significantly reduce the dispersion across six different data sets.

The second and third bars in Figure 3-6 show the reduction in flow time for CMS resulting from the use of reduced batch sizes or the implementation of transfer batches in cells. For our test bed, the numerical averages reported in Table 3-7 indicate that while retaining the original batch size in CMS produces flow ratios in the range 0.78 - 0.89, using a reduced batch size or transfer batches in the cells can further reduce the flow ratios to lie in the range 0.37-0.63.

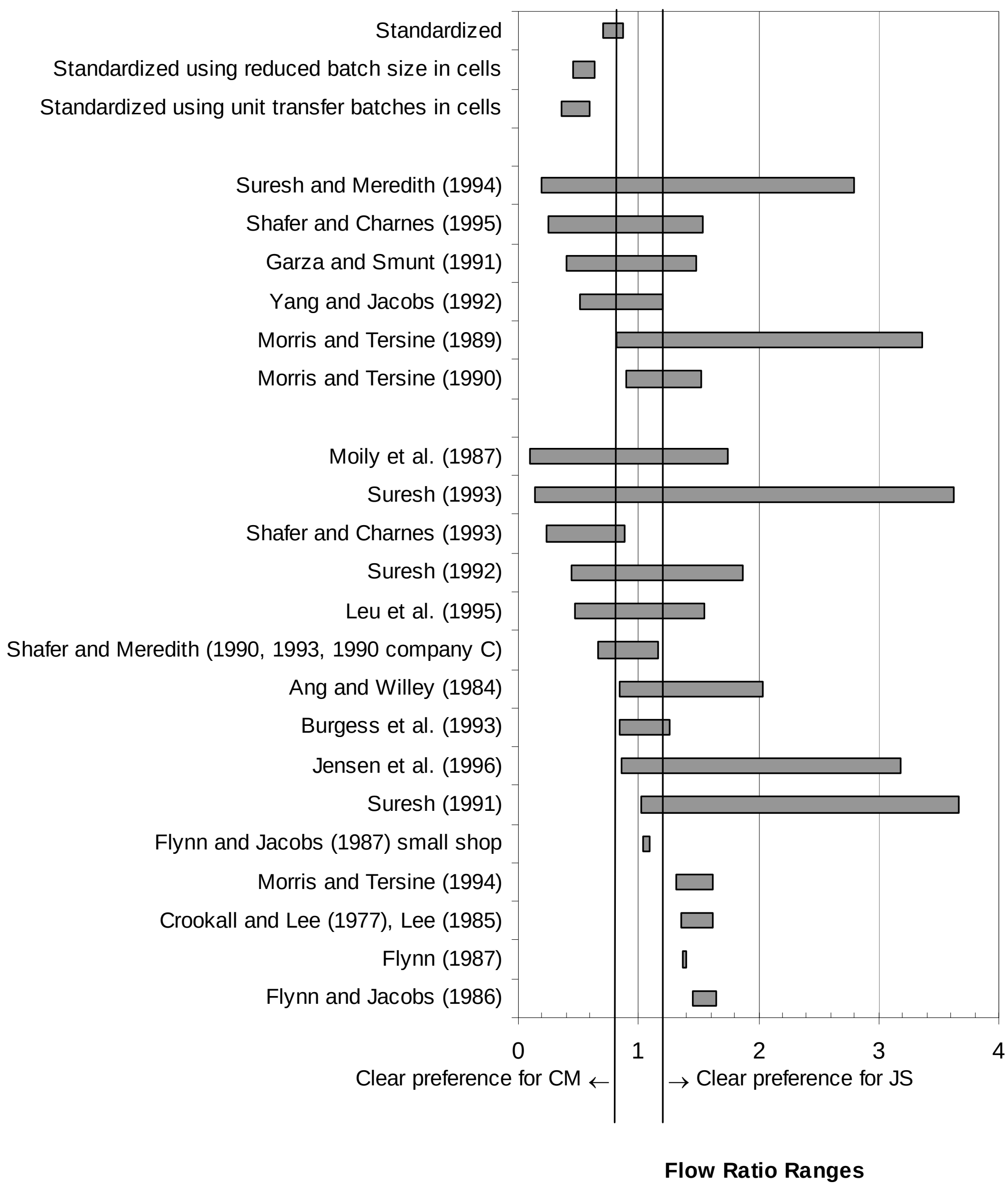

Figure 3-6. Results from standardized approach reduce variability and favor CM.

One may inquire as to the possible sources of the wide dispersion seen in Figure 3-6. Of the 17 data sets where the job shop flow times are superior, eight did not discount setups at all. On the other hand, ten data sets showed better flow times for CMS. Seven of these ten data sets used a high ratio of setup to run time (some going up to 6.0, compared to our baseline values of 1.0). The other three used transfer batches in the cells. For the specific studies included in our test bed, Table 3-9 compares the flow time results reported in the literature with our results and provides our choice of the most likely factors that can explain the difference for each study.

Data Set ID	Source	Source simulation results: JS to CM TB = *b* mean	Source simulation results: JS to CM TB = 1 mean	Standardized simulation results: JS to CM TB = *b* mean	Standardized simulation results: JS to CM TB = *b* stdev	Standardized simulation results: JS to CM TB = 1 mean	Standardized simulation results: JS to CM TB = 1 stdev	Source setting explaining the difference
1	Garza and Smunt 1991	1.42	n/a	0.72	0.003	0.37	0.001	low *s* /*br* range
4	Suresh and Meredith 1994	0.93	n/a	0.78	0.005	0.38	0.003	high JS utilization
5	Yang and Jacobs 1992	0.59	n/a	0.80	0.004	0.51	0.004	large material handling effect present in JS
6a	Morris and Tersine 1989	1.19	0.82	0.82	0.006	0.45	0.004	low *s* /*br* high minor/major setup
6b	Morris and Tersine 1990	1.05	n/a	0.82	0.006	0.45	0.004	high minor/major setup
6c	Shafer and Charnes 1995	n/a	0.90	0.82	0.006	0.45	0.004	low *s* /*br* high minor/major setup

Table 3-9. JS to CMS flow ratios in the modeling literature.

The results of our runs also allow us to compare the relative impact of utilization level, setup potential, and batch size reduction. We have shown this in Figure 3-7 for a single data set (#2). The topmost bar shows the range of flow ratios obtained by changing the utilization levels, the second bar shows the results for different combinations of the setup ratios and setup fraction, and the last bar shows the effect of using a smaller batch size or adopting transfer batches.

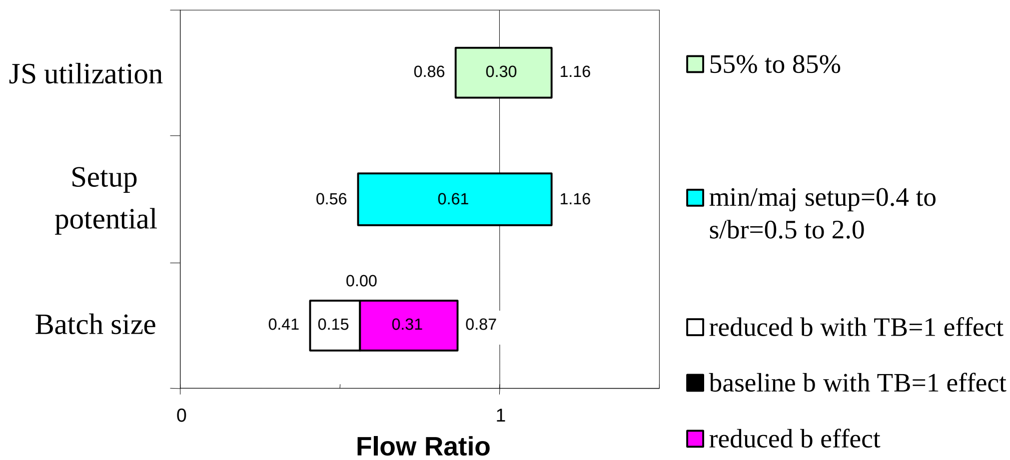

Figure 3-7. Results of sensitivity analysis for data set 2.

3.10 Summary

In this research, we argue that the wide divergence reported in the literature occurs because of differences in the choice of demand data, production environments, setup structures, utilization levels, cell formation, and significant disparities in the operation of the production system. The present study attempts to study the sources of variation more systematically by standardizing the operating rules of the factories and adopting a common set of key parameters ranges, while retaining the differences in demand and part type characteristics across data sets. By performing a set of baseline runs with standardized values and a host of sensitivity runs on the level of the standardized factors individually, we seek to gauge the effect of each factor more reliably.

Of pivotal importance to our computational study is the use of six different data sets selected from different sources in the open literature, so that the results would not be tied to a single profile of part types, mix, or demands. To our knowledge, this is the first study that compares CMS conversion benefits across disparate data sets. In addition, by using consistent operating principles in the simulation runs, we took utmost care to make the comparison between the job shop and CMS environments even handed.

Based on over 2000 simulation runs conducted in this study, we can summarize our main conclusions as follows.

- The conversion of job shops to cells consistently improves flow time by 10% to 20%, for the test bed used in this study. This result provides a conservative estimate of the advantages of CMS because it does not take advantage of such additional factors as reduced batch sizes, transfers batches, or move times. We conclude that setup reduction can overcome the effects of pooling loss as long as the magnitude of the setups is not too small and no significant bottlenecks develop in the cells upon conversion.
- The use of reduced batch sizes, or the implementation of transfer batches, can each provide cells with an additional improvement in flow time. Typically, each of these two factors has a significant effect on reducing the flow time for CMS, and the amount of reduction is usually at least as large as that obtained by conversion to CMS without any changes in the batch sizes.
- The sensitivity runs show that the overall factory utilization and the potential for setup reduction can both affect the conversion results obtained. Our tests indicate that conversion to CMS may not be advantageous if the utilization level is high or there is not sufficient potential to reduce setups.

- The design of cells also has a clear impact on the conversion improvements obtained. Typically, we observed better performance in cells when the original source design was used. However, conversion benefits continue to be present even after we use a uniform cell formation procedure due to Vakharia and Wemmerlöv (1990). This indicates that careful allocation of machines to cells to avoid heavy utilization helps to keep the pooling loss within tight control.
- Our experimental runs support the conclusions of previous authors that RL dispatching provides less overall setup and supports lower flow times than FCFS in a job shop with setup. The effect of RL seems to diminish in the same factory setting once it incorporates cells.

In summary, we believe that this part of the dissertation has shown that the comparison of job shops and cellular systems with respect to the flow time measure can produce reasonably consistent results when the same operating rules and key parameter ranges are used across different data sets. Moreover, our research shows that setup reduction can overcome pooling losses, even under the conservative assumptions where batch size remain unchanged and the material transport times in the job shop are assumed to be negligible. Overall, the conclusions of our research are consistent with the qualitative insights cited in the literature when comparing CMS and job shops. However, our research clarifies that the quantitative comparisons using the flow time metric must be interpreted in the context of the region of the parameter space spanned by the data sets, as well as the particular design used for the cells.

Chapter 4

PARTIAL CELLULAR MANUFACTURING SYSTEMS

Conversion from a job shop environment to cellular manufacturing does not need to proceed all the way: one can consider a partial implementation of cellular layout. One can investigate what the benefits of a partial cellular layout may be as compared to full conversion. For example, we may ask if a *few* cells can provide *most* of the flow time benefits associated with full conversion. To answer this question, we use the same data sets we analyzed fully in Chapter 3. We consider partial cellular layouts at all levels ranging between the two extremes of JS (no cells) and CMS (all cells). For each hybrid layout, we evaluate the flow times in both the cells and the remainder shop and relate this to congestion effects. We find that cell selection, sequence of cell application, level of cellular implementation and load balance are all important considerations in the implementation of partial layouts.

4.1 Simulation Analysis of PCMS

The evaluation of partial layouts follows the schema used in Chapter 3. For each data set considered, there is a complete cellular layout that is known in advance. This is the all-cell layout corresponding to full conversion. Suppose that this layout uses NC cells. We can consider each partial layout as a choice of a subset S of the set $T = \{1, \ldots, NC\}$. Given a subset S of selected cells, let $FR(S)$ be the flow ratio of the configuration represented by the cells in S and the remainder shop handling all parts not assigned to these selected cells. We will use simulation to evaluate $FR(S)$ for all subsets of a fixed cardinality n, where n is successively increased from 1 to NC. The exhaustive evaluation of all subsets of n cells allows us to rank sort all subsets of size n with respect to total factory flow time. For each n, we record the *best* pick as the subset S of size n that results in the lowest flow ratio and label it $BP(n)$ and denote its flow ratio $BFR(n)$. Similarly, the worst pick subset of cells at level n is associated with the highest overall flow ratio is denoted by $WP(n)$ with flow ratio $WFR(n)$.

Table 4.1 presents the results of this analysis for all six data sets discussed in Chapter 3. As in Chapter 3, the setup reduction reflects the total setup paid relative to the total setup paid in the JS layout. At each fixed n, we also compare the best and worst flow ratios obtained at that level with the best overall pick that gives the lowest flow ratio across all n. We denote this best overall flow ratio as $BFR^* = \min\{BFR(n)\}$ with the minimum taken over all n from 1 to NC. This minimum may be achieved for the all-cell option where $n = NC$ or a partial layout using a smaller number of cells. We identify the optimum level of cellular implementation for each data set as the smallest n for which there is no further marginal reduction in flow ratio. The marginal reduction in flow ratio at any level $n < NC$ is calculated as $\frac{BFR(n-1) - BFR(n)}{1 - BFR^*}$ or $\frac{BFR(n-1) - WFR(n)}{1 - BFR^*}$ and for $n = NC$ is $\frac{BFR(NC-2) - BFR(NC)}{1 - BFR^*}$.

In order to assess the impact of the cellular investment at a given implementation level n, we try to relate the factory flow ratio to the fraction of machines and part demands sent allocated to the cells. Specifically, these ratios are computed as follows: We indicate the number of machines sent to cells for the best and worst pick at level n as $BM(n)$ and $WM(n)$, respectively. Therefore, the fraction of machines sent to the cells is calculated as $BM(n)\big/\sum_j NM_j$ and $WM(n)\big/\sum_j NM_j$ (we remind the reader from our notation in Chapter 3 that the number of machines of type j in the factory is NM_j). Similarly, we indicate the total batch demand sent to cells, $\sum_{f(i)\in F(S)} \lambda_i$ where $F(S)$ is the family of parts assigned to the cells in S, for the best and worst pick at level n as $BD(n)$ and $WD(n)$, respectively. The fraction of batch demands sent to the cells are calculated as $BD(n)\big/\sum_i \lambda_i$ and $WD(n)\big/\sum_i \lambda_i$.

To illustrate the contents of Table 4-1, we now review the information presented for data set 3. We see from the maximum number of cells formed that there are four cells to choose from. At $n = 2$, where we allow two cells to be formed, $BP(2) = \{3,4\}$. The simulation results of that pick list that the overall factory will enjoy a 70% setup reduction as compared to the original JS. The measured flow ratio from the simulation is 0.890. This particular pick happens to be equivalent in flow time to the all-cell pick. In this case only 66% of machines and 47% of batch demands and have been sent to the (two) cells. If we read the $n = 4$ data we see that there is no further reduction in flow ratio if we split up the remaining resources and demands.

The last data set entry, $6^{\dagger}$, represents a perturbation to data set 6. We created a bottleneck by shifting the load on a particular machine type: we changed the routing of the parts requiring machine type 6 common to cells 4 and 5 such that the machine in cell 4 (when selected) was only 20% utilized. Therefore, whenever cell 4 was selected the residual was left with type 6 machine utilization in excess of 90%. Data set $6^{\dagger}$ is a case where the best partial cell option is better than the all-cell option (the difference in the all-cell and partial option 1,2,3 flow times is significant with >95% confidence using a paired-t test).

Data Set	Number of cells formed	Cell Ids: Best Worst	Setup Reduction mean (%)	Setup Reduction stdev	Flow Raio mean	Flow Raio stdev	Machines in Cell(s) (%)	Batch Demands in Cell(s) (%)
1	1	5	19	0.012	0.937	0.011	17	17
		4	18	0.012	0.948	0.010	17	17
	2	1,3	37	0.010	0.862	0.011	33	33
		2,3	36	0.010	0.890	0.011	33	33
	3	4,5,6	54	0.008	0.787	0.009	50	50
		2,4,6	45	0.009	0.836	0.010	50	50
	4	3,4,5,6	66	0.006	0.734	0.008	67	67
		1,3,5,6	58	0.007	0.777	0.009	67	67
	6	1,2,3,4,5,6	69	0.004	0.716	0.008	100	100
2	1	4	30	0.007	0.949	0.008	25	25
		1	31	0.007	0.986	0.009	25	25
	2	2,4	62	0.004	0.897	0.009	50	50
		1,3	51	0.005	0.956	0.009	50	50
	4	1,2,3,4	77	0.002	0.867	0.010	100	100
3	1	3	45	0.004	0.911	0.009	40	26
		2	32	0.005	0.984	0.012	23	31
	2	3,4	70	0.002	0.890	0.013	66	47
		1,2	45	0.004	0.956	0.012	34	53
	4	1,2,3,4	70	0.002	0.890	0.013	100	100
4	1	4	25	0.004	0.944	0.004	23	21
		2	21	0.005	0.959	0.005	19	20
	2	4,5	45	0.004	0.878	0.005	42	40
		2,3	41	0.004	0.904	0.005	39	41
	3	3,4,5	66	0.002	0.807	0.004	61	61
		1,2,3	59	0.004	0.841	0.005	58	60
	5	1,2,3,4,5	72	0.002	0.781	0.005	100	100
5	1	1	50	0.003	0.877	0.004	40	33
		3	38	0.004	0.915	0.004	30	33
	3	1,2,3	74	0.001	0.798	0.004	100	100
6	1	2	28	0.008	0.940	0.004	27	26
		4	19	0.008	0.989	0.008	17	18
	2	1,2	53	0.007	0.891	0.009	50	49
		3,4	37	0.007	0.967	0.009	37	36
	3	1,2,3	72	0.005	0.835	0.008	70	66
		3,4,5	53	0.006	0.920	0.009	50	51
	5	1,2,3,4,5	76	0.005	0.824	0.009	100	100
6†	1	1	26	0.003	0.932	0.006	23	23
		4	6	0.095	2.008	0.309	17	17
	2	1,2	53	0.003	0.861	0.008	50	49
		3,4	21	0.009	2.123	0.335	37	34
	3	1,2,3	75	0.001	0.749	0.005	70	66
		2,3,4	59	0.009	1.171	0.079	63	60
	5	1,2,3,4,5	78	0.001	0.766	0.005	100	100

Table 4-1. Simulation results for best and worst picks at each level of cellular implementation.

4.1.1 Cell Selection. To ensure that every potential layout is assessed, we ran the simulation model exhaustively for all subsets S of the set of cells for each of the six data sets plus the a perturbed data set 6. The resulting comparison reveals that the choice of the cells at each level makes a difference. For any n, we observe a difference in the flow ratios between the best and worst picks. Data set 6 shows this clearly: at $n = 1$ the best pick, cell 2, results in flow ratio of 0.940 whereas the worst pick, cell 4, results in a flow ratio of 0.989.

When we look across results from all the data sets we can compare the last two columns with the flow ratios. We see that $BP(n), n < NC$ always results in a greater flow time reduction than the batch demands or machines invested, but this is not the case with the worst picks. Again, using data set 6 as an example, $BP(1)$ results in 34% [(1-0.940)/(1-0.824)*100%] of the possible flow ratio reduction for that data set while requiring only 27% of the machines to be located in cells to work on 20% of the batch demands. We contrast this with $WP(1)$ resulting in six percent of flow ratio reduction [(1-0.989)/(1-0.824)*100%], but requiring 17% of the machines in the cells working on 25% of the batch demands. So, even though there may be several choices available that will improve the overall factory flow time, the best pick leverages the resources of batch arrivals and machines most effectively.

We also observe that $BP(n)$ has setup reduction that matches and often exceeds the setup reduction achieved by $WP(n)$. Although large differences in setup reduction can account for a portion of the difference between factory flow times, it is not the only source of such differences. A good example is available for data set 1 for $n = 2$. The setup reductions achieved by $BP(n)$ and $WP(n)$ are equal, yet there is a three percent difference in factory flow times ($BFR(2) - WFR(2)$). To explain this disparity we must also review the machine utilization as shown in Figure 4-1.

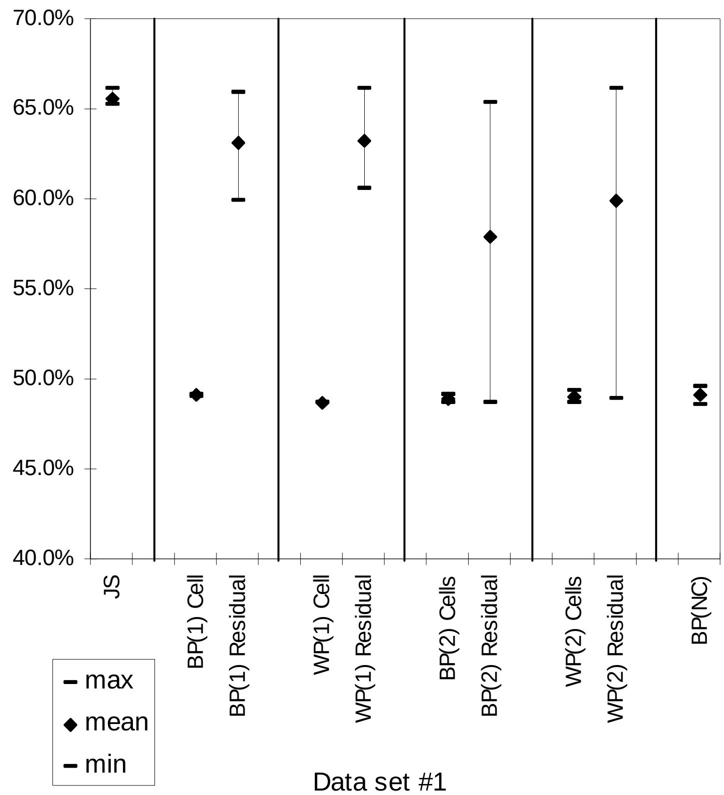

Figure 4-1. Machine utilization ranges during early stages of CMS implementation.

Both cell utilization levels are well below the original JS. The best choice $BP(2)$ shows a lower average and maximum utilization in the residual while the utilization is comparable in the cells. We get an indication from this example that in comparing subsets S of the same size, a pair of subsets may show equal performance on the cell side of the shop but the preferred choice may be the subset that achieves superior performance in the remainder shop. We also note that there can be flow time differences even when the cells seem to allocate the resources equally. Like the CMS analysis in Chapter 3, we find that ex post setup reduction information alone is not sufficient to discern the best cell pick(s).

4.1.2 Effect of Sequence. Full conversions from JS to CMS reported in the computational studies found in the literature do not address the order in which cells are implemented. However, the empirical literature clearly shows that firms tend to implement "one step at a time." Here we address the sequence question. Using the same data sets we ask the natural question, "is there always a nested picking order from a single cell to the all-cell conversion option?" To put it in practical terms, the manager should be alerted if a cell that appears to be the best choice at a given stage turns out to be an inferior pick once other cells come into being. In any event, the manager prefers nested sequences of subsets S with increasing cardinality since dismantling a cell formed earlier is unattractive.

In our limited number of data sets tested here we found the occurrence of mutually exclusive sets of cells picked at different levels of CMS implementation suggesting sequence of cells picked *can* matter.

We look at data set 1 for an example of this phenomenon: $BP(1) = \{5\}$, but $BP(2) = \{1,3\}$ and then $BP(3) = \{4,5,6\}$. While not shown in Table 4-1, it turns out that in this case there is little difference in the factory flow times of the $S = \{5,6\}$ and the best pick at $n = 2$, $S = \{1,3\}$. In fact, the former set was ranked second best in a close contest. Given the best choice for $n = 3$ level, it is clear that the manger would prefer the sequence of cells 5, followed by 6, followed by 4 to a blind implementation of the best subset at each level. Such considerations suggest look-ahead strategies and the use of a richer set of criteria in selecting the cells for partial implementation.

4.1.3 Stopping Rule. The results of this chapter confirm our statement in Chapter 1 that the best overall flow may be achieved by a hybrid layout, rather than either a pure JS or all-cell options. In such cases, one should look for rules or strategies to halt conversion at some intermediate state instead of proceeding to full conversion. This is apparent in the results of the simulation runs for data sets 3 and $6^{\dagger}$. In data set 3, $BFR(2) = BFR(NC) = BFR^*$. Any further implementation of cells after $n = 2$ will not result in further reduction in flow time. In data set $6^{\dagger}$ further cell picks (equivalent to all-cell conversion) will actually degrade the factory overall flow time, $BFR(3) = BFR^* < BFR(NC)$.

4.2 Summary

The analysis performed in this chapter provides some insights into implementing partial cell layouts (hybrids) using the same test bed as in Chapter 3. Below we summarize some of the lessons learned from the exhaustive computational evaluation of all partial layouts. We did not pursue this line of investigation any further because we could not identify general and robust rules that applied across all data sets. Our observations may be summarized as follows:

1. Even when the number of cells to be included in the partial layout is fixed, the choice of the correct subset of cells can have a significant impact on the flow time. In short, selection matters.
2. The sequence of best subsets to pick as n increases from 1 to NC is not necessarily nested, so sequence matters.
3. Factory flow time of a partial cellular implementation may be as good as or even better than the all-cell option as we have shown in our perturbed data set 6, so it is important to stop short of full conversion where appropriate.
4. The differences in factory flow times are due to the same factors recognized in the all-cell CMS analysis, setup reduction and machine utilization, but neither factor alone is sufficient to reliably determine the best subset of cells to select. The best picks are characterized by large setup reductions along with reduction of utilization in the residual job shop and the lack of bottlenecks in the cell(s), so setup reduction and load balance in both the cells and residual job shop matter.

Chapter 5

ANALYTIC MODELING OF A SIMPLE SYSTEM WITH SETUP

The analysis of a job shop under the assumptions of the factory environment in Chapter 3 presents major challenges in modeling. The simplest model appears to be a queueing network model with setups. We do not intend to address the approximations made by queueing models in this work, especially since adjustments for setups are generally not made in any exact fashion. Instead, in this chapter, we use analytic models to gain insights into the extent of setup economies that can be obtained by using dispatching disciplines designed to avoid unnecessary setups and compare these with first-come-first-serve (FCFS) protocols. We focus on the simplest queueing model we could find that handles the effect of setups on flow time exactly. This system involves two customer classes with general service time distributions and setups are incurred when switching from one class to the other. The dispatching rule we investigate is designed to minimize the incidence of setups in a queue with two customer classes. This will provide a theoretical underpinning for our empirical findings in Chapter 3, where we found that the dispatching rule selected does make a difference.

5.1 Zero Setup

We start by establishing a baseline in the absence of setups, evaluating flow times under FCFS versus a dispatching rule that minimizes the incidence of changeovers. Our comparison involves a system with two customer classes, where each customer requires a single operation at the service facility. Initially, we assume that the setup time equals zero, and study the queueing system under two different dispatching regimes: Alternating Priority (AP) and FCFS. We already know from Avi-Itzhak et al. (1965) that if the two classes have the same service distribution, then the mean flow times of both systems are the same (assuming zero setup). Here, we focus on the asymmetric case where the service distributions are different. Further, we choose cases where the first and second moments are easily related and therefore develop our result with the assumption of exponential service since $E\left(S_i^2\right) = 2E\left(S_i\right)^2$. We employ two general results for our comparison. To measure the AP (two-queue) flow time, we start with the general result from Eisenberg (1967). We measure the flow time of the FCFS (single queue) using the familiar Pollaczek-Khintchine (P-K) formula for the M/G/1. We follow the analytic comparison of AP versus FCFS in the zero setup case with numerical comparisons at two arrival rate settings.

Because setup times are not involved, there is no difference between service times paid in either regime, so we focus on the average wait time until service, versus the flow time, $\overline{F}$. We use the notation W_q for the wait in queue when there is zero setup, consistent with queueing notation. We use the notation $\overline{W}$ when the wait includes non-zero setup. The flow time always includes any setup time paid.

5.1.1 Analytic comparison of AP versus FCFS. From Eisenberg (1967) the general wait time for AP after removing setup for the class-1 queue is:

$$W_{q_1}^{AP} = \frac{\lambda_1 E(S_1^2)}{2(1-\rho_1)} + \frac{\rho_2^2 \lambda_1 E(S_1^2) + (1-\rho_1)^2 \lambda_2 E(S_2^2)}{2(1-\rho_1)(1-\rho)[(1-\rho_1)(1-\rho_2)+\rho_1\rho_2]}$$

and for the class-2 queue is:

$$W_{q_2}^{AP} = \frac{\lambda_2 E(S_2^2)}{2(1-\rho_2)} + \frac{\rho_1^2 \lambda_2 E(S_2^2) + (1-\rho_2)^2 \lambda_1 E(S_1^2)}{2(1-\rho_2)(1-\rho)[(1-\rho_1)(1-\rho_2)+\rho_1\rho_2]}$$

Together, the overall average wait time is:

$$\begin{aligned} W_q^{AP} &= \frac{1}{2\lambda}\left\{\frac{\lambda_1^2 E(S_1^2)}{(1-\rho_1)} + \frac{\lambda_2^2 E(S_2^2)}{(1-\rho_2)}\right\} \\ &\quad + \frac{1}{2\lambda D(1-\rho)}\left\{\frac{\rho_1^2 \lambda_2^2 E(S_2^2) + (1-\rho_2)^2 \lambda_1 \lambda_2 E(S_1^2)}{(1-\rho_2)}\right. \\ &\quad \left. + \frac{\rho_2^2 \lambda_1^2 E(S_1^2) + (1-\rho_1)^2 \lambda_1 \lambda_2 E(S_2^2)}{(1-\rho_1)}\right\} \end{aligned} \tag{1}$$

where $D = (1-\rho_1)(1-\rho_2) + \rho_1\rho_2$

but for FCFS, $$W_q^{FCFS} = \frac{1}{2(1-\rho)}\left[\lambda_1 E(S_1^2) + \lambda_2 E(S_2^2)\right]. \tag{2}$$

This follows from the Pollaczek-Khintchine formula for the single M/G/1 queue:

$$W_q = \frac{\lambda E(S^2)}{2(1-\rho)}. \tag{3}$$

For our case $$E\left(S^2\right)=\frac{\lambda_1}{\lambda}E\left(S_1^2\right)+\frac{\lambda_2}{\lambda}E\left(S_2^2\right)$$

so $$\lambda E\left(S^2\right)=\lambda_1 E\left(S_1^2\right)+\lambda_2 E\left(S_2^2\right).$$

Now convert to exponential case using $E\left(S_i^2\right)=2E\left(S_i\right)^2$ or $\lambda_i^2 E\left(S_i^2\right)=2\rho_i^2$

$$W_q^{AP}=\frac{\rho_1^2}{\lambda\left(1-\rho_1\right)}+\frac{\rho_2^2}{\lambda\left(1-\rho_2\right)}+\frac{1}{\lambda D(1-\rho)}\times \left\{\frac{\rho_1^2\rho_2^2+\lambda_1\left(1-\rho_1\right)^2\rho_2 E\left(S_2\right)}{\left(1-\rho_1\right)}+\frac{\rho_1^2\rho_2^2+\lambda_2\left(1-\rho_2\right)^2\rho_1 E\left(S_1\right)}{\left(1-\rho_2\right)}\right\}. \tag{4}$$

For the exponential case, (2) becomes the following

$$W_q^{FCFS}=\frac{\rho_1 E\left(S_1\right)+\rho_2 E\left(S_2\right)}{1-\rho}. \tag{5}$$

We can re-write the expression for W_q^{AP} in (4) slightly differently:

$$\begin{aligned}
W_q^{AP} &= \frac{\rho_1^2}{\lambda}\frac{1}{1-\rho_1}+\frac{\rho_1^2}{\lambda D(1-\rho)}\frac{\rho_2^2}{1-\rho_1}+\frac{\lambda_1\left(1-\rho_1\right)\rho_2 E\left(S_2\right)}{\lambda D(1-\rho)} \\
&\quad +\frac{\rho_2^2}{\lambda}\frac{1}{1-\rho_2}+\frac{\rho_2^2}{\lambda D(1-\rho)}\frac{\rho_1^2}{1-\rho_2}+\frac{\lambda_2\left(1-\rho_2\right)\rho_1 E\left(S_1\right)}{\lambda D(1-\rho)} \\
&= \frac{\rho_1^2\left[D(1-\rho)+\rho_2^2\right]}{\lambda D(1-\rho)\left(1-\rho_1\right)}+\frac{\rho_2^2\left[D(1-\rho)+\rho_1^2\right]}{\lambda D(1-\rho)\left(1-\rho_2\right)} \\
&\quad +\frac{\lambda_1\left(1-\rho_1\right)\rho_2 E\left(S_2\right)+\lambda_2\left(1-\rho_2\right)\rho_1 E\left(S_1\right)}{\lambda D(1-\rho)}.
\end{aligned}$$

Consider the bracketed expression within the first term:

$$\begin{aligned}
D(1-\rho)+\rho_2^2 &= \left(1-\rho+2\rho_1\rho_2\right)(1-\rho)+\rho_2^2 \\
&= (1-\rho)^2+2\rho_1\rho_2(1-\rho)+\rho_2^2
\end{aligned}$$

write $(1-\rho)=\left(1-\rho_1\right)-\rho_2$ then

$$
\begin{aligned}
D(1-\rho)+\rho_2^2 &= (1-\rho_1)^2+2\rho_2(1-\rho_1)+\rho_2^2+2\rho_1\rho_2(1-\rho_1)-2\rho_1\rho_2^2+\rho_2^2 \\
&= (1-\rho_1)^2-2\rho_2(1-\rho_1)^2+2\rho_2^2-2\rho_1\rho_2^2 \\
&= (1-\rho_1)^2-2\rho_2(1-\rho_1)^2+2\rho_2^2(1-\rho_1) \\
&= (1-\rho_1)^2(1-2\rho_2)+2\rho_2^2(1-\rho_1)
\end{aligned}
$$

so $$\frac{\rho_1^2\left[D(1-\rho)+\rho_2^2\right]}{\lambda D(1-\rho)(1-\rho_1)}=\frac{\rho_1^2}{\lambda D(1-\rho)}\left[(1-\rho_1)(1-2\rho_2)+2\rho_2^2\right]$$

similarly $$\frac{\rho_2^2\left[D(1-\rho)+\rho_1^2\right]}{\lambda D(1-\rho)(1-\rho_2)}=\frac{\rho_2^2}{\lambda D(1-\rho)}\left[(1-\rho_2)(1-2\rho_1)+2\rho_1^2\right].$$

Also note that

$$
\begin{aligned}
(1-\rho_1)(1-2\rho_2)+2\rho_2^2 &= (1-\rho_1)(1-2\rho_2+\rho_2^2)+2\rho_2^2-\rho_2^2(1-\rho_1) \\
&= (1-\rho_1)(1-\rho_2)^2+\rho_2^2+\rho_1\rho_2^2 \\
&= (1-\rho_1)(1-\rho_2)^2+\rho_2^2(1+\rho_1)
\end{aligned}
$$

so we can simplify the expression for W_q^{AP}

$$
\begin{aligned}
W_q^{AP} &= \frac{1}{\lambda D(1-\rho)}\Big\{\rho_1^2\left[(1-\rho_1)(1-2\rho_2)+2\rho_2^2\right]+\rho_2^2\left[(1-\rho_2)(1-2\rho_1)+2\rho_1^2\right] \\
&\quad +\left[\lambda_1(1-\rho_1)\rho_2E(S_2)+\lambda_2(1-\rho_2)\rho_1E(S_1)\right]\Big\} \\
&= \frac{1}{\lambda D(1-\rho)}\Big\{\left[4\rho_1^2\rho_2^2+\rho_1^2(1-\rho_1)(1-2\rho_2)+\rho_2^2(1-\rho_2)(1-2\rho_1)\right] \\
&\quad +\left[\lambda_1(1-\rho_1)\rho_2E(S_2)+\lambda_2(1-\rho_2)\rho_1E(S_1)\right]\Big\}
\end{aligned}
$$

also note that

$$
\begin{aligned}
(1-\rho_1)(1-2\rho_2)+2\rho_2^2 &= 1-\rho_1-2\rho_2+2\rho_1\rho_2+2\rho_2^2 \\
&= 1-\rho_1-\rho_2+2\rho_1\rho_2+\left(2\rho_2^2-\rho_2\right) \\
&= D-\rho_2(1-2\rho_2)
\end{aligned}
$$

and similarly $$(1-\rho_2)(1-2\rho_1)+2\rho_1^2=D-\rho_1(1-2\rho_1)$$

so we can re-write the first two terms of W_q^{AP}

$$
\begin{aligned}
W_q^{AP} &= \frac{1}{\lambda D(1-\rho)}\left[\rho_1^2\left(D-\rho_2(1-2\rho_2)\right)+\rho_2^2\left(D-\rho_1(1-2\rho_1)\right)\right] \\
&\quad + \frac{1}{\lambda D(1-\rho)}\left[\lambda_1(1-\rho_1)\rho_2 E(S_2)+\lambda_2(1-\rho_2)\rho_1 E(S_1)\right] \\
&= \frac{1}{\lambda(1-\rho)}\left(\rho_1^2+\rho_2^2\right)-\frac{\left[\rho_1^2\rho_2(1-2\rho_2)+\rho_1\rho_2^2(1-2\rho_1)\right]}{\lambda D(1-\rho)} \\
&\quad + \frac{1}{\lambda D(1-\rho)}\left[\lambda_1(1-\rho_1)\rho_2 E(S_2)+\lambda_2(1-\rho_2)\rho_1 E(S_1)\right] \\
W_q^{AP} &= \frac{\left(\rho_1^2+\rho_2^2\right)}{\lambda(1-\rho)}-\frac{\rho_1\rho_2\left(\rho_1+\rho_2-4\rho_1\rho_2\right)}{\lambda D(1-\rho)} \\
&\quad + \frac{\left[\lambda_1(1-\rho_1)\rho_2 E(S_2)+\lambda_2(1-\rho_2)\rho_1 E(S_1)\right]}{\lambda D(1-\rho)}.
\end{aligned}
$$

We now try to relate this to W_q^{FCFS} by replacing the first expression using the relation:

$$
\begin{aligned}
W_q^{FCFS} &= \frac{\rho_1 E(S_1)+\rho_2 E(S_2)}{(1-\rho)} \\
&= \frac{(\lambda_1+\lambda_2)\left[\rho_1 E(S_1)+\rho_2 E(S_2)\right]}{\lambda(1-\rho)} \\
&= \frac{\rho_1^2+\rho_2^2+\lambda_2\rho_1 E(S_1)+\lambda_1\rho_2 E(S_2)}{\lambda(1-\rho)}
\end{aligned}
$$

so

$$
\begin{aligned}
W_q^{AP} &= \left[W_q^{FCFS}-\frac{\lambda_2\rho_1 E(S_1)+\lambda_1\rho_2 E(S_2)}{\lambda(1-\rho)}\right]-\frac{\rho_1\rho_2\left(\rho_1+\rho_2-4\rho_1\rho_2\right)}{\lambda D(1-\rho)} \\
&\quad + \frac{\left[\lambda_1(1-\rho_1)\rho_2 E(S_2)+\lambda_2(1-\rho_2)\rho_1 E(S_1)\right]}{\lambda D(1-\rho)}
\end{aligned}
$$

therefore

$$W_q^{AP} - W_q^{FCFS} = \frac{1}{\lambda D(1-\rho)}\{\lambda_1\rho_2 E(S_2)(1-\rho_1-D) + \lambda_2\rho_1 E(S_1)(1-\rho_2-D)\}$$
$$-\frac{\rho_1\rho_2(\rho_1+\rho_2-4\rho_1\rho_2)}{\lambda D(1-\rho)}.$$

Term within braces is $\rho_1(1-\rho_2-D)\lambda_2 E(S_1) + \rho_2(1-\rho_1-D)\lambda_1 E(S_2)$.

Use $$1-\rho_1-D = \rho_2(1-2\rho_1)$$

and $$1-\rho_2-D = \rho_1(1-2\rho_2)$$

to write above as $\rho_1^2(1-2\rho_2)\lambda_2 E(S_1) + \rho_2^2(1-2\rho_1)\lambda_1 E(S_2)$

and substitute $$\lambda_i = \frac{\rho_i}{E(S_i)}$$

to get $$\rho_1^2(1-2\rho_2)\rho_2\frac{E(S_1)}{E(S_2)} + \rho_2^2(1-2\rho_1)\rho_1\frac{E(S_2)}{E(S_1)} =$$

$$\rho_1\rho_2\left[\rho_1(1-2\rho_2)\frac{E(S_1)}{E(S_2)} + \rho_2(1-2\rho_1)\frac{E(S_2)}{E(S_1)}\right].$$

The final result is:

$$W_q^{AP} - W_q^{FCFS} = \frac{\rho_1\rho_2}{\lambda D(1-\rho)} \times \left[\rho_1(1-2\rho_2)\frac{E(S_1)}{E(S_2)} + \rho_2(1-2\rho_1)\frac{E(S_2)}{E(S_1)} - (\rho_1+\rho_2-4\rho_1\rho_2)\right]. \tag{6}$$

We can now ask when the expression within brackets is negative.

If we let $Q = \frac{E(S_1)}{E(S_2)}$, then we have an expression $f(Q) = AQ + \frac{B}{Q} - (A+B)$ then we can re-write (6) as

$$W_q^{AP} - W_q^{FCFS} = \frac{C}{\lambda} f(Q) \tag{7}$$

where $A = \rho_1(1-2\rho_2)$, $B = \rho_2(1-2\rho_1)$ and $C = \rho_1\rho_2 / [D(1-\rho)]$.

It is well known that the minimum value of $AQ + \frac{B}{Q}$ equals $2\sqrt{AB}$ if $\rho_i < 1/2$. So $f(Q)$ has minimum value

$$2\sqrt{AB} - (A+B) = -\left[(A+B) - 2\sqrt{AB}\right] = -\left(\sqrt{A} - \sqrt{B}\right)^2.$$

Observation:

$$W_q^{AP} - W_q^{FCFS} \geq -\frac{\rho_1\rho_2}{\lambda D(1-\rho)}\left[\sqrt{\rho_1(1-2\rho_2)} - \sqrt{\rho_2(1-2\rho_1)}\right]^2 \quad \text{if both } \rho_i < 1/2.$$

So, as long as both $\rho_i < 1/2$, we have a bound on how much better W_q^{AP} can do as compared to W_q^{FCFS}. From this analysis, it is clear that $\text{Min } f(Q) < 0$ if $A \neq B$.

Also, if $Q = 1$ then clearly $f(Q) = 0$. Since $f(Q)$ is U-shaped, we know that there is another root with $Q < 1$ and $f(Q) = 0$ as illustrated in Figure 5-1.

If $B < A$, the roots are $Q = \frac{B}{A}$ and 1 with $\frac{B}{A} < \sqrt{\frac{B}{A}} < 1$ if $0 < B < A$.

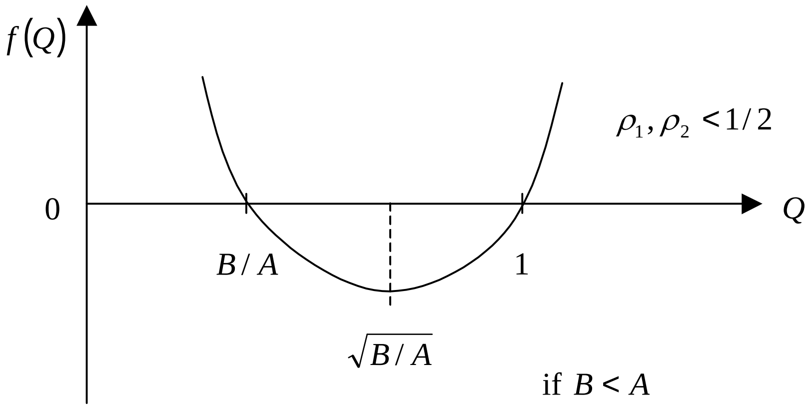

Figure 5-1. Roots and minimum for $f(Q)$ when $\rho_1, \rho_2 < 1/2$.

We now address the case where the condition $\rho_i < 1/2$ does *not* hold. The stability of the queueing system requires that $1 - \rho_1 - \rho_2 > 0$ or $\rho_1 + \rho_2 < 1$. Thus, $\rho_1 > 1/2$ forces $\rho_2 < 1/2$.

Since $f(1) = 0$ in all cases, from (7) we see that $W_q^{AP} = W_q^{FCFS}$ for $Q = 1$, so $Q = 1$ is a root for the function f. Since $\rho_2 < 1/2 < \rho_1$ implies that $A > 0$ and $B < 0$, $f'(Q) = A - \dfrac{B}{Q^2} > 0$ for all values of Q. So f is strictly increasing over $[0, \infty)$ and $Q = 1$ is the only root. As Figure 5-2 shows, this implies that

$$W_q^{AP} < W_q^{FCFS} \text{ if } Q < 1$$

and

$$W_q^{AP} > W_q^{FCFS} \text{ if } Q > 1.$$

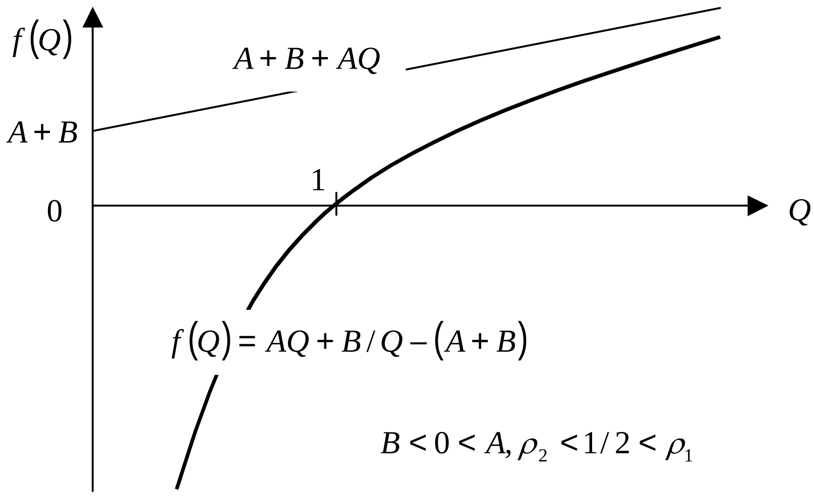

Figure 5-2. Single root of $f(Q)$ when $\rho_1 > 1/2$.

We summarize the preceding discussion in the form of a theorem.

Theorem 1 Consider the two-class single server system with zero setups, exponential service times, and Poisson arrivals. Let the average wait times for the AP and FCFS be denoted as W_q^{AP} and W_q^{FCFS} and set $\Delta W_q = W_q^{AP} - W_q^{FCFS}$.

Then
$$\Delta W_q = \frac{C}{\lambda} f(Q)$$

where
$$f(Q) = AQ + \frac{B}{Q} - (A+B)$$

$$A = \rho_1(1-2\rho_2),\ B = \rho_2(1-2\rho_1),\ C = \rho_1\rho_2/[D(1-\rho)]$$

$$D = (1-\rho_1)(1-\rho_2) + \rho_1\rho_2,\ Q = \frac{E(S_1)}{E(S_2)}, \text{ and assuming } \lambda_1 \geq \lambda_2.$$

If $\rho_1 > 1/2$, then $f(Q)$ is strictly increasing and has a single root at $Q = 1$.

So $$\Delta W_q < 0 \quad \text{if } Q < 1 \quad \text{and} \quad \Delta W_q \geq 0 \quad \text{if } Q \geq 1.$$

If $\rho_2 < \rho_1 < 1/2$, so that both ρ_i's are less than ½, then $f(Q)$ is U-shaped and has two roots at $Q = \frac{B}{A}$ and 1, so that

$$\Delta W_q < 0 \quad \text{if } \frac{B}{A} < Q < 1$$

$$\Delta W_q \geq 0 \quad \text{if } Q \leq \frac{B}{A} \quad \text{or } Q \geq 1.$$

Theorem 1 applies to exponential service. We now extend it for use with non-exponential service. Previously, we used the relationship between the moments, $E(S_i^2) = kE(S_i)^2$, with $k = 2$ for the exponential case. We know that $k = 1$ for constant service times. We note how k is related to the coefficient of variation:

$$k = \frac{E(S_i^2)}{E(S_i)^2} = \frac{E(S_i)^2 + \sigma^2}{E(S_i)^2} = 1 + C_{S_i}^2.$$

Then using k,

$$W_q^{AP} = \frac{k}{2}\left\{\frac{\rho_1^2}{\lambda(1-\rho_1)} + \frac{\rho_2^2}{\lambda(1-\rho_2)}\right\}$$
$$+ \frac{k}{2}\frac{1}{\lambda D(1-\rho)}\left\{\frac{\rho_1^2\rho_2^2 + \lambda_1(1-\rho_1)^2\rho_2 E(S_2)}{(1-\rho_1)} + \frac{\rho_1^2\rho_2^2 + \lambda_2(1-\rho_2)^2\rho_1 E(S_1)}{(1-\rho_2)}\right\}$$

and

$$W_q^{FCFS} = \frac{k}{2}\frac{\rho_1 E(S_1) + \rho_2 E(S_2)}{1-\rho}.$$

So we have introduced a new factor, $k/2$, and therefore know the maximum benefit received by AP in an exponential service environment will be twice that of a constant service environment.

The preceding theorem summarizes the two types of behavior exhibited by the difference ΔW_q. We now proceed to map the regions where either AP or FCFS is superior in the full parameter space of the problem.

Consider any system with parameters

$$\left(\lambda_1, \lambda_2, E(S_1), E(S_2), \rho_1, \rho_2\right)$$

when $\rho_i = \lambda_i E(S_i)$. We define a reference system with parameters

$$\left(\lambda_1, \lambda_2, Q, 1, \lambda_1 Q, \lambda_2\right)$$

where $$E(S_2) = 1, \quad Q = \frac{E(S_1)}{E(S_2)}$$

and with no loss of generality, assume that $\lambda_1 \geq \lambda_2$.

It is clear that we can convert any system to the reference system by a simple re-indexing (if necessary) and re-scaling. Stated otherwise, from the arbitrary

system $$\left(\tilde{\lambda}_1, \tilde{\lambda}_2, E(\tilde{S}_1), E(\tilde{S}_2), \rho_1, \rho_2\right)$$

we get $$\left(\lambda_1, \lambda_2, \frac{E(\tilde{S}_1)}{E(\tilde{S}_2)}, 1, \rho_1, \rho_2\right)$$

by defining $\lambda_i = \tilde{\lambda}_i E(S_i)$. Note that in such a re-scaling, the ρ_i's remain invariant so the expression for ΔW_q changes by the scaling factor alone, that is:

$$\Delta \tilde{W}_q = \frac{C}{\tilde{\lambda}} f(Q) = \frac{CE(\tilde{S}_1)}{\lambda} f(Q) = E(S_1)\Delta W_q .$$

This shows that it is sufficient to map the behavior of the reference system as long as ΔW_q is of interest.

Consider the system with $(\lambda_1, \lambda_2, Q, 1, \rho_1, \rho_2)$

where $\rho_1 = \lambda_1 Q$ and $\rho_2 = \lambda_2$.

The stability conditions are $\rho_1 = \lambda_1 Q < 1$, $\rho_2 = \lambda_2 < 1$

and $\rho_1 + \rho_2 = \lambda_1 Q + \lambda_2 < 1$. (8)

We also assume that $\lambda_1 \geq \lambda_2$. (9)

We consider four cases as listed below. The first three correspond to $\lambda_2 < 1/2$ and the last one to $\lambda_2 \geq 1/2$. We discuss each case briefly and then summarize the results in Table 5-1 and Table 5-2.

Case 1:	$\lambda_2 < 1/2$	$\rho_2 < 1/2 < \rho_1$
2:	$\lambda_2 < 1/2$	$\rho_2 \leq \rho_1 < 1/2$
3:	$\lambda_2 < 1/2$	$\rho_1 < \rho_2 < 1/2$
4:	$\lambda_2 \geq 1/2$	$\rho_1 < 1/2 \leq \rho_2$

Case 1: The stability conditions and the $\lambda_1 \geq \lambda_2$ requirement define the relevant

region as $$\max\left\{\lambda_2, \frac{1}{2Q}\right\} \leq \lambda_1 < \frac{1-\lambda_2}{Q} \quad \text{with } \lambda_2 < 1/2. \tag{10}$$

In this case $\rho_2 < 1/2 < \rho_1$ implies that $A > 0$ and $B < 0$, so $f(Q)$ as defined in Theorem 1 is increasing for $Q > 0$ and has a single root at $Q = 1$. So

$$\Delta W_q \leq 0 \qquad \text{if } Q \leq 1$$

and $$\Delta W_q > 0 \qquad \text{if } Q > 1.$$

Case 2: $\rho_2 \leq \rho_1 < 1/2$. The region is defined by

$$\max\left\{\lambda_2, \frac{\lambda_2}{Q}\right\} \leq \lambda_1 < \frac{1}{2Q} \quad \text{with } \lambda_2 < 1/2. \tag{11}$$

Since $0 < B \leq A$ in this region, $f(Q)$ has two roots, at $Q = B/A$ and $Q = 1$, so $\Delta W_q < 0$ if Q lies between these two roots. We need to express the condition

$\frac{B}{A} < Q < 1$ as a condition on λ_1.

$$\frac{B}{A} = \frac{\rho_2(1-2\rho_1)}{\rho_1(1-2\rho_2)} < Q$$

means $$\left(\frac{1}{\rho_1} - 2\right) < \left(\frac{1-2\lambda_2}{\lambda_2}\right)Q$$

or $$\frac{1}{\lambda_1 Q} < 2 + \left(\lambda_2^{-1} - 2\right)Q.$$

So the condition is $$\lambda_1 > \frac{1}{Q\left[2 + \left(\lambda_2^{-1} - 2\right)Q\right]} \quad \text{with } Q < 1 \tag{12}$$

given $\lambda_2 < 1/2$.

Note that the right-hand-side is decreasing in Q for $Q > 0$, and that its value for $Q = 1$ equals λ_2. Since $\lambda_1 \geq \lambda_2$ at all times, the range of validity of this condition is up to $Q = 1$.

Case 3: $\rho_1 < \rho_2 < 1/2$. The region requires

$$\lambda_2 \leq \lambda_1 < \frac{\lambda_2}{Q} \qquad \text{for } \lambda_2 < 1/2. \tag{13}$$

This immediately implies that $Q \leq 1$. While $f(Q)$ has two roots at $Q = 1$ and $Q = B/A > 1$, the latter root does not fall into this region, so we conclude that

$$\Delta W_q \geq 0 \qquad \text{for } 0 < Q \leq 1.$$

Case 4: $\rho_1 < 1/2 \leq \rho_2$. The region is defined by

$$\lambda_2 \leq \lambda_1 < \frac{1-\lambda_2}{Q} \qquad \text{for } \lambda_2 \geq 1/2. \tag{14}$$

The relation (14) forces $Q < \dfrac{1-\lambda_2}{\lambda_2}$ and since λ_2 satisfies $1/2 \leq \lambda_2 \leq 1$, Q must satisfy $0 < Q < 1$. Since $A \leq 0$ and $B > 0$ in this region, $f(Q)$ is strictly decreasing over $(0,1)$ and $f(1) = 0$. So, in this region, we always have:

$$\Delta W_q \geq 0 \qquad \text{for } 0 < Q \leq 1.$$

The four cases are summarized in Table 5-1 for the region $0 < Q \leq 1$ and in Table 5-2 for $1 < Q$.

Case	Region for $0 < Q \le 1$	ΔW_q
(1) $\rho_2 < 1/2 < \rho_1$	$\frac{1}{2Q} \le \lambda_1 < \frac{1-\lambda_2}{Q}$ $\lambda_2 < 1/2$	$\Delta W_q \le 0$
(2) $\rho_2 < \rho_1 \le 1/2$	$\frac{\lambda_2}{Q} \le \lambda_1 < \frac{1}{2Q}$ $\lambda_2 < 1/2$	$\Delta W_q < 0$ if $\lambda_1 > \frac{1}{Q\left[2 + \left(\lambda_2^{-1} - 2\right)Q\right]}$
(3) $\rho_1 < \rho_2 < 1/2$	$\lambda_2 \le \lambda_1 < \frac{\lambda_2}{Q}$ $\lambda_2 < 1/2$	$\Delta W_q \ge 0$
(4) $\rho_1 < 1/2 \le \rho_2$	$\lambda_2 \le \lambda_1 < \frac{1-\lambda_2}{Q}$ $\lambda_2 \ge 1/2$	$\Delta W_q \ge 0$

Table 5-1. Four cases defining the parameter space for $0 < Q \le 1$.

Case	Region for $Q > 1$	ΔW_q
(1) $\rho_2 < 1/2 < \rho_1$	$\max\left(\lambda_2, \frac{1}{2Q}\right) \le \lambda_1 < \frac{1-\lambda_2}{Q}$ $\lambda_2 < 1/2$	$\Delta W_q \ge 0$
(2) $\rho_2 < \rho_1 \le 1/2$	$\lambda_2 \le \lambda_1 < \frac{1}{2Q}$ $\lambda_2 < 1/2$	$\Delta W_q > 0$
(3) $\rho_1 < \rho_2 < 1/2$	N/A	
(4) $\rho_1 < 1/2 \le \rho_2$	N/A	

Table 5-2. Four cases defining the parameter space for $Q > 1$.

Focusing on the sign of ΔW_q, we can state the results in the following form.

<u>Theorem 2</u> For any system with parameters $(\lambda_1, \lambda_2, E(S_1), E(S_2), \rho_1, \rho_2)$ with the conventions $\lambda_1 \geq \lambda_2$ and $E(S_2) = 1$, the Alternating Priority policy is superior to FCFS if and only if

$$\rho_2 < 1/2 < \rho_1 \quad \text{and} \quad 0 < Q \leq 1$$

or $$\rho_2 < \rho_1 \leq 1/2, \quad 0 < Q \leq 1 \quad \text{and} \quad \lambda_1 > \frac{1}{Q\left[2 + \left(\lambda_2^{-1} - 2\right)Q\right]}$$

where $$Q = E(S_1)/E(S_2).$$

We now illustrate the relevant regions for representative values of the parameter λ_2. We start with the choice $\lambda_2 = 1/4$. The stability condition is $\lambda_1 < \frac{1-\lambda_2}{Q} = \frac{3}{4Q}$, so λ_1 must lie below the graph for $y = \frac{3}{4Q}$ in the λ_1 versus Q-space. The condition $\lambda_1 \geq \lambda_2 = 1/4$ must also be satisfied at all times. The region of superiority of AP is given by $\Delta W_q < 0$ and corresponds to

$$\frac{1}{2Q} < \lambda_1 < \frac{3}{4Q} \qquad \text{for } 0 < Q \leq 1.$$

For $Q > 1$, the region $\frac{1}{2Q} < \lambda_1 < \frac{3}{4Q}$ is where $\Delta W_q > 0$ until Q reaches 3 where the constraint $\lambda_1 \geq 1/4$ becomes binding.

For Case 2, the relevant region is defined by

$$\frac{1}{4Q} < \lambda_1 < \frac{1}{2Q} \qquad \text{for } 0 < Q \leq 1$$

and $\frac{1}{4} < \lambda_1 < \frac{1}{2Q}$ with $1 < Q \le 2$.

The condition for $\Delta W_q < 0$ is $\lambda_1 > \frac{1}{2Q(Q+1)}$ for $0 < Q < 1$.

The relevant regions are illustrated in Figure 5-3. Moving on to Figure 5-4, the regions are shown for $\lambda_2 = 0.10$. We see that the regions corresponding to Cases 1 and 2 for $Q < 1$ have both widened. Conversely, in Figure 5-5, when λ_2 increases to 0.4, we see that these regions have narrowed compared to the $\lambda_2 = 1/4$ case. This behavior remains in effect as long as $\lambda_2 < 1/2$.

Now consider the scenario when $\lambda_2 > 1/2$. When λ_2 exceeds $1/2$, only Case 4 applies and the region is defined by

$$\lambda_2 \le \lambda_1 < \frac{1-\lambda_2}{Q} \qquad \text{with } Q < \frac{1-\lambda_2}{\lambda_2}.$$

For $\lambda_2 = 0.6$, for example, we have

$$0.6 \le \lambda_1 < \frac{0.4}{Q} = \frac{2}{5Q} \qquad \text{with } Q < \frac{0.4}{0.6} = \frac{2}{3}$$

so the only relevant region lies between the horizontal line at 0.6 and the curve $\frac{2}{5Q}$ as shown in Figure 5-6. Within this region $\Delta W_q \ge 0$ and outside this region, the system is unstable.

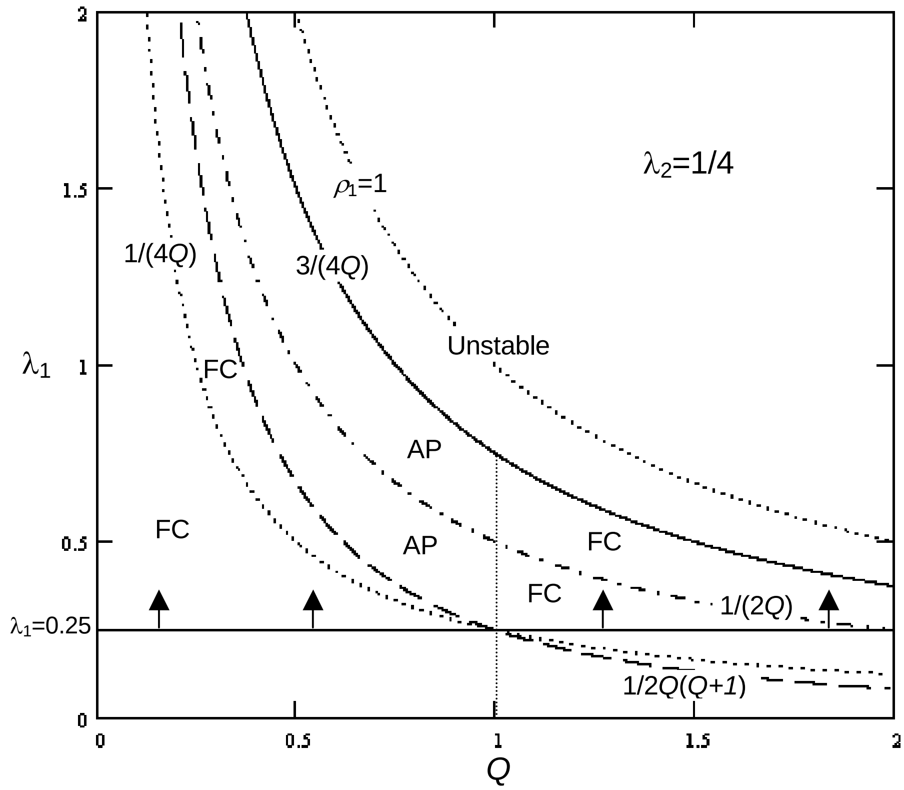

Figure 5-3. Graph of λ_1 versus Q when $\lambda_2 = 0.25$. AP and FC indicates superiority in that region.

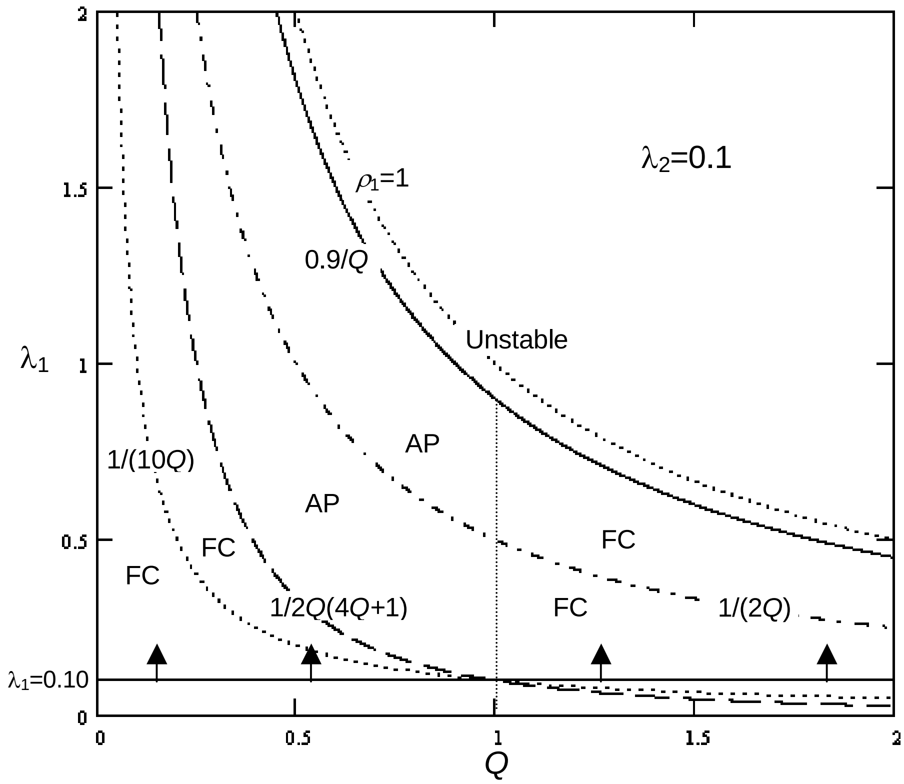

Figure 5-4. Graph of λ_1 versus Q when $\lambda_2 = 0.10$.

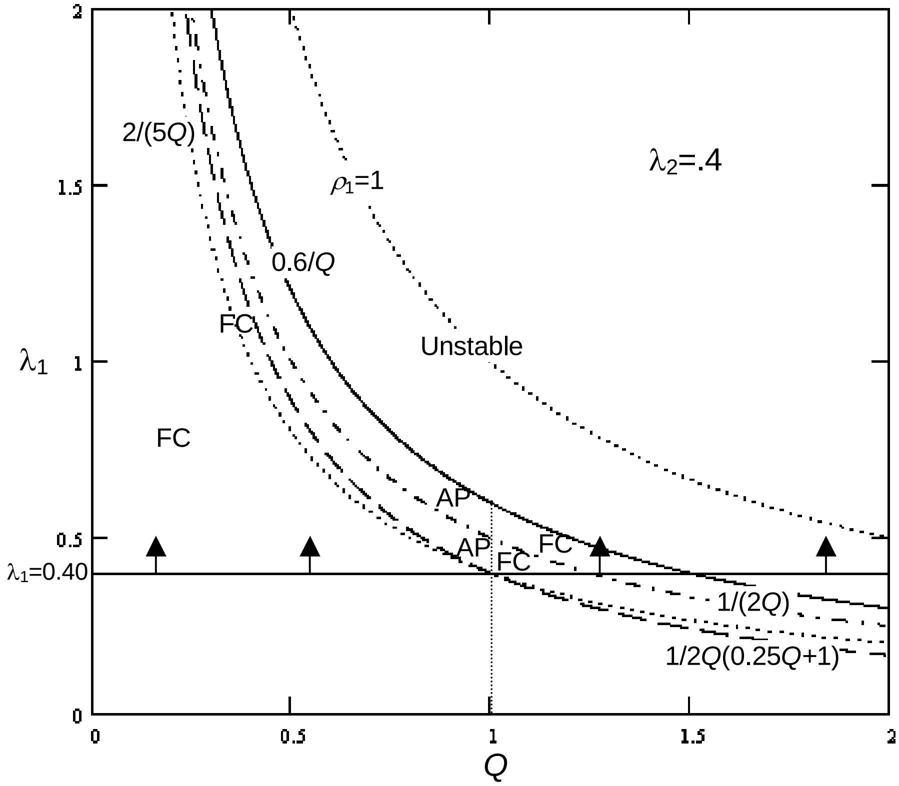

Figure 5-5. Graph of λ_1 versus Q when $\lambda_2 = 0.40$.

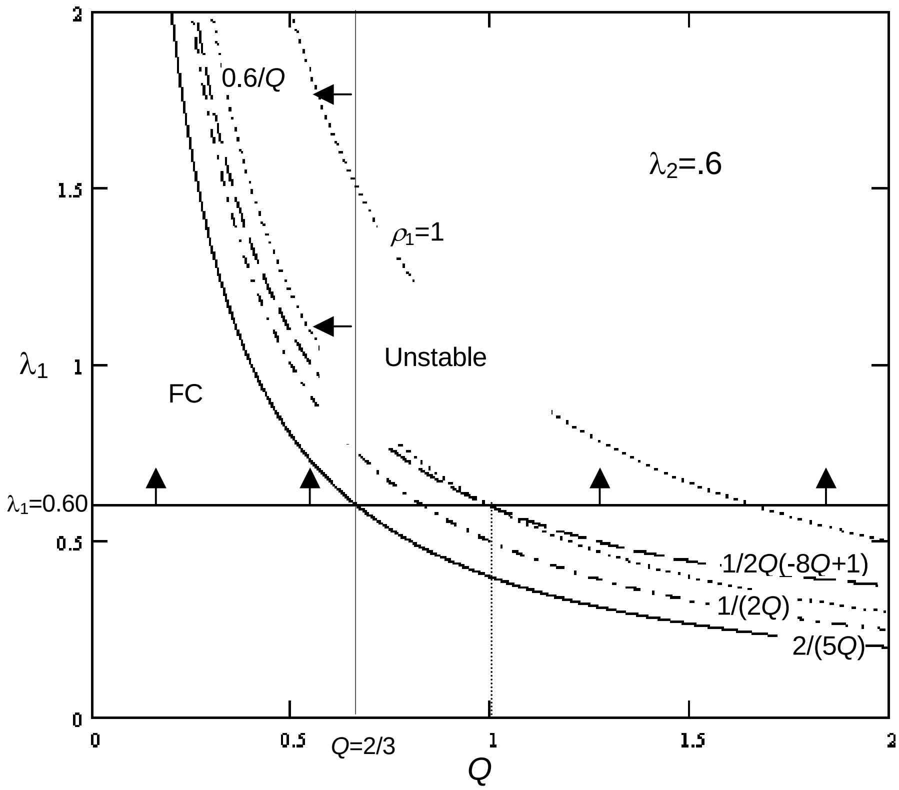

Figure 5-6. Graph of λ_1 versus Q when $\lambda_2 = 0.60$.

5.1.2 Baseline numerical comparisons. We choose two of the preceding λ_2 settings for our zero-setup baseline, $\lambda_2 = 0.25$ and $\lambda_2 = 0.60$. Figure 5-7 contains a matrix of discrete values at equal 0.05 intervals of λ_1 and Q where the numerical value at each location is $\left(W_q^{AP} - W_q^{FCFS}\right)*100 = \Delta W_q *100$ as defined in Section 5.1.1. Figure 5-7 therefore resembles Figure 5-3. We label and italicize the cells that unstable due to ρ_1 saturation, "R1," the cells that are unstable due to the sum of the ρ_i's as "RS," and cells that violate $\lambda_1 \geq \lambda_2$, "LV". We assist the reader by adding a light shade to the $\Delta W_q < 0$ region and darker shading to the $\Delta W_q > 0$ region. We leave the region of $\Delta W_q = 0$ un-shaded (for example at $Q = 1$). The reader will note that although not coincident with our specific measurement points and therefore not shown without shade, the transition from $\Delta W_q < 0$ to $\Delta W_q > 0$ includes the $\Delta W_q = 0$ curve. This is not true when transitioning to a zone of instability or $\lambda_1 \geq \lambda_2$ violation.

We remind the reader that we have assumed $E(S_2) = 1$, so that $E(S_1) = Q$ and therefore the expected service time equals $E(S) = \dfrac{\lambda_1 Q + \lambda_2}{\lambda_1 + \lambda_2}$. For Figure 5-7, $\lambda_2 = 0.25$, so $E(S) = \dfrac{\lambda_1 Q + 0.25}{\lambda_1 + 0.25}$. The actual wait difference is useful because the four largest differences that favor AP in Figure 5-7 are less than 1.5 and all four occur when $\rho > 0.95$ (not shown). AP, therefore, has little positive impact in the absence of setup when $\lambda_2 < 1/2$. If $Q > 1$ then AP can be significantly worse than FCFS, but only when λ_1 approaches $3/(4Q)$.

For Figure 5-8, $\lambda_2 = 0.60$, so $\rho_2 = 0.60$ and $E(S) = \dfrac{\lambda_1 Q + 0.60}{\lambda_1 + 0.60}$. We simplified Figure 5-8 by trimming off a majority of the unavailable space: where $\lambda_1 < \lambda_2$ and for this case $(\lambda_2 > 1/2)$ where $Q > 1$. The load offered by each class in the absence of setup is ρ_i. AP is biased towards the class that provides the majority of the load (we will call this the *dominant* class. Since AP will not changeover until the current queue is exhausted there is a greater likelihood that a dominant class arrival will occur continuing the work session than when working on the lesser class. Continuing work on the dominant class is done at the expense of the other class. The net result for the $\lambda_2 > 1/2$ case is higher wait times when using AP where the feasible area for this case starts with $\rho > 0.60$. We will see in Figures 5-9 and 5-10 that AP does require fewer changeovers as compared to FCFS, but the tradeoff is not always beneficial to the overall system flow time, especially when there is no setup time at stake.

λ_1 \ Q	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1
1	8.38	7.71	6.87	5.82	4.5	2.85	0.77	-1.9	-5.3	-10	-17	-27	-47	-104	RS	RS	RS	RS	RS	R1
0.95	8.29	7.65	6.86	5.87	4.66	3.18	1.35	-0.9	-3.8	-7.5	-12	-19	-31	-53	-125	RS	RS	RS	RS	RS
0.9	8.2	7.59	6.83	5.9	4.8	3.45	1.84	-0.1	-2.5	-5.4	-9.2	-14	-21	-32	-54	-133	RS	RS	RS	RS
0.85	8.1	7.51	6.79	5.93	4.9	3.7	2.27	0.58	-1.4	-3.8	-6.7	-10	-15	-21	-31	-49	-113	RS	RS	RS
0.8	8	7.42	6.74	5.93	4.99	3.9	2.64	1.17	-0.5	-2.5	-4.7	-7.4	-11	-14	-19	-26	-38	-71	RS	RS
0.75	7.87	7.32	6.68	5.92	5.06	4.06	2.94	1.67	0.23	-1.4	-3.2	-5.2	-7.5	-10	-13	-16	-20	-24	-28	RS
0.7	7.74	7.22	6.6	5.9	5.1	4.2	3.19	2.08	0.85	-0.5	-2	-3.6	-5.2	-7	-8.8	-10	-12	-12	-10	0
0.65	7.59	7.08	6.5	5.85	5.11	4.3	3.4	2.43	1.36	0.22	-1	-2.2	-3.5	-4.8	-6	-7	-7.6	-7.3	-5.4	0
0.6	7.42	6.94	6.4	5.78	5.1	4.37	3.56	2.71	1.78	0.82	-0.2	-1.2	-2.2	-3.2	-4	-4.7	-4.9	-4.6	-3.2	0
0.55	7.23	6.77	6.25	5.69	5.07	4.39	3.68	2.92	2.12	1.29	0.45	-0.4	-1.2	-2	-2.6	-3.1	-3.2	-3	-2	0
0.5	7.01	6.57	6.1	5.57	5	4.39	3.75	3.08	2.38	1.67	0.95	0.24	-0.4	-1	-1.6	-1.9	-2.1	-1.9	-1.3	0
0.45	6.77	6.35	5.91	5.41	4.89	4.34	3.77	3.18	2.56	1.95	1.34	0.74	0.18	-0.3	-0.8	-1.1	-1.2	-1.1	-0.8	0
0.4	6.48	6.1	5.67	5.23	4.75	4.26	3.75	3.22	2.69	2.15	1.62	1.12	0.63	0.2	-0.2	-0.5	-0.6	-0.6	-0.4	0
0.35	6.14	5.79	5.4	4.99	4.56	4.13	3.67	3.21	2.74	2.27	1.82	1.37	0.96	0.57	0.24	-0	-0.2	-0.3	-0.2	0
0.3	5.74	5.41	5.06	4.7	4.32	3.93	3.53	3.13	2.72	2.31	1.92	1.53	1.18	0.84	0.54	0.29	0.1	-0	-0.1	0
0.25	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.2	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.15	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.1	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.05	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV

λ_1 \ Q	1.05	1.1	1.15	1.2	1.25	1.3	1.35	1.4	1.45	1.5	1.55	1.6	1.65	1.7	1.75	1.8	1.85	1.9	1.95	2
1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.95	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.9	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.85	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.8	RS	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.75	RS	RS	RS	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.7	80.8	RS	RS	RS	RS	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.65	14.7	65.8	1594	RS	RS	RS	RS	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.6	6.69	20.7	53.6	164	RS	RS	RS	RS	RS	RS	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1
0.55	3.69	10.2	21.8	43.3	88.6	215	1318	RS	RS	RS	RS	RS	RS	RS	RS	RS	R1	R1	R1	R1
0.5	2.19	5.75	11.4	20.2	34.5	58.5	102	197	494	RS	RS	RS	RS	RS	RS	RS	RS	RS	RS	R1
0.45	1.32	3.38	6.47	11	17.5	27	41	62.3	96.1	154	268	565	2699	RS	RS	RS	RS	RS	RS	RS
0.4	0.78	1.99	3.74	6.21	9.62	14.3	20.5	29.1	40.8	57	80	114	165	251	411	797	2769	RS	RS	RS
0.35	0.43	1.11	2.1	3.47	5.32	7.76	10.9	15	20.3	27	35.7	46.8	61.4	80.5	106	141	191	264	380	583
0.3	0.19	0.54	1.06	1.79	2.77	4.06	5.71	7.77	10.4	13.5	17.4	22.2	28.1	35.2	44	54.8	68.1	84.5	105	131
0.25	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.2	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.15	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.1	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.05	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV

Figure 5-7. Wait time differences (AP-FCFS)*100 when setup is zero and $\lambda_2 = 0.25$.

λ_1 \ Q	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1
1	134	129	126	125	130	147	212	RS	RS	RS	RS	RS	RS	RS	RS	RS	RS	RS	RS	R1
0.95	132	126	123	121	124	135	170	404	RS	RS	RS	RS	RS	RS	RS	RS	RS	RS	RS	RS
0.9	129	123	120	117	118	125	145	225	RS	RS	RS	RS	RS	RS	RS	RS	RS	RS	RS	RS
0.85	126	120	116	114	113	116	128	165	407	RS	RS	RS	RS	RS	RS	RS	RS	RS	RS	RS
0.8	123	117	113	110	108	109	116	134	200	RS	RS	RS	RS	RS	RS	RS	RS	RS	RS	RS
0.75	119	114	109	106	103	103	106	115	141	263	RS	RS	RS	RS	RS	RS	RS	RS	RS	RS
0.7	115	110	106	102	99	97	98	101	113	148	366	RS	RS	RS	RS	RS	RS	RS	RS	RS
0.65	111	106	102	97	94	92	90	91	96	108	147	467	RS	RS	RS	RS	RS	RS	RS	RS
0.6	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV

Figure 5-8. Wait time differences (AP-FCFS)*100 when setup is zero and $\lambda_2 = 0.60$.

5.2 Non-Zero Setup

In this section, we introduce a nonzero setup into the comparison of the two dispatching rules AP and FCFS. We are no longer able to use P-K formula for the FCFS wait because it assumes independence in the processing times and we know that the setup times are correlated to the service times by the customer class. The solution given by Gaver (1963) allows for the processing time correlation by class. To provide a baseline for comparison, we use the results of the last section to report measured differences in wait time as well as differences in the incidence of part changeovers (number of switches). The introduction of setup starts at a low level. The magnitude of the setup is then increased until it equals the batch service time, a level that is consistent with our simulation studies in Chapters 3 and 4.

5.2.2 FCFS versus AP in the Non-Zero Setup Environment. We continue with the comparison started in section 5.1 comparing AP to FCFS now with non-zero setup.

The inputs to both FCFS and AP flow time calculations are the same:

- Two streams of Poisson arrivals with mean arrival rates $\lambda_i \; i = 1,2$,

  $\lambda_1 + \lambda_2 = \lambda$, $a_i = \dfrac{\lambda_i}{\lambda}$

- Distribution function of the service time of a type-i customer: $F_{S_i}(t)$,

  first moment: $E(S_i)$, second moment: $E(S_i^2)$. Laplace-Stieltjes

  transform of distributions: $\gamma_i(z) = \int_0^\infty e^{-zt} \mathrm{d}F_{S_i}(t)$ (15)

  Note: If the service time is exponential then $\gamma_i(z) = \dfrac{1}{1 + zE(S_i)}$ (16)

- Distribution function of the setup time of a type-i customer: $F_{U_i}(t)$,

  first moment: $E(U_i)$, second moment: $E(U_i^2)$. Laplace-Stieltjes

  transform of distributions: $\kappa_i(z) = \int_0^\infty e^{-zt} \mathrm{d}F_{U_i}(t)$ (17)

  Note: If the setup time is exponential then $\kappa_i(z) = \dfrac{1}{1 + zE(U_i)}$ (18)

The FCFS wait time (wait in queue prior to setup or service) of Gaver (1963) is based on a Markov process with a simple integro-differential forward Kolmogorov equation. The waiting time of a random arrival at t, $W(t)$, depends on the class of the last service which will determine whether or not a setup is required. If the arrival is of the same class then there is no setup required, otherwise a setup must occur prior to service. The joint probabilities result:

$$F_1(x,t) = P\{W(t) \le x, \text{ last demand prior to t in class } 1\}$$

and

$$F_2(x,t) = P\{W(t) \leq x \text{, last demand prior to t in class } 2\}$$

Under suitable conditions, the functions $F_i(x,t)$ will have a limit as $t \to \infty$. If we denote the limiting functions by $F_i(x)$, the Laplace-Stieltjes transforms by $f_i(s)$, then we have

$$f_1(s) = \frac{s[F_1(0)\{s - \lambda + \lambda_2\gamma_2(s)\} - F_2(0)\lambda_1\kappa_1(s)\gamma_1(s)]}{D(s)} \tag{19a}$$

and

$$f_2(s) = \frac{s[F_2(0)\{s - \lambda + \lambda_1\gamma_1(s)\} - F_1(0)\lambda_2\kappa_2(s)\gamma_2(s)]}{D(s)} \tag{19b}$$

where

$$D(s) = [s - \lambda + \lambda_1\gamma_1(s)][s - \lambda + \lambda_2\gamma_2(s)] - \lambda_1\lambda_2\kappa_1(s)\gamma_1(s)\kappa_2(s)\gamma_2(s) \tag{20}$$

and
$$\lim_{s \to 0}\left(f_1(s) + f_2(s)\right) = 1. \tag{21}$$

By taking the limit of (19) we note the probabilities, $F_1(x)$ and $F_2(x)$, are related by:

$$F_1(0) + F_2(0) = 1 - \overline{U} \tag{22}$$

where $\overline{U} = \rho + \lambda\left[\frac{\lambda_1\lambda_2}{\lambda^2}\{E(U_1) + E(U_2)\}\right]$ and $\rho = \lambda_1 E(S_1) + \lambda_2 E(S_2)$.

This is exactly the same utilization measure obtained using conditional probabilities as outlined in Section 2.2.1.

The expected wait is prior to setup or service is

$$E_1(W) + E_2(W) \tag{23}$$

where $\quad E_1(W) = \lim_{s \to 0} (-1) \dfrac{df_1(s)}{ds}$ and $E_2(W) = \lim_{s \to 0} (-1) \dfrac{df_2(s)}{ds}$

and

$$\begin{aligned} E_1(W) = {} & \frac{F_1(0)\{1 - \lambda_2 E(S_2)\} + F_2(0)\lambda_1 E(U_1 + S_1)}{\lambda(1-\overline{U})} \\ & + \frac{\lambda_1}{2\lambda^2(1-\overline{U})} \times \{\lambda_1 \lambda_2 \ [-2E(S_1)E(S_2) + 4E(U_1)E(S_1) + 4E(U_2)E(S_2) \\ & + 2E(U_1)E(S_2) + 2E(U_2)E(S_1) + E(U_1^2) + E(U_2^2)] \\ & + \lambda\lambda_1 E(S_1^2) + \lambda\lambda_2 E(S_2^2) + 2\lambda_1 E(S_1) + 2\lambda_2 E(S_2) - 2\} \end{aligned} \tag{24}$$

$$\begin{aligned} E_2(W) = {} & \frac{F_2(0)\{1 - \lambda_1 E(S_1)\} + F_1(0)\lambda_2 E(U_2 + S_2)}{\lambda(1-\overline{U})} \\ & + \frac{\lambda_2}{2\lambda^2(1-\overline{U})} \times \{\lambda_1 \lambda_2 \ [-2E(S_1)E(S_2) + 4E(U_1)E(S_1) + 4E(U_2)E(S_2) \\ & + 2E(U_1)E(S_2) + 2E(U_2)E(S_1) + E(U_1^2) + E(U_2^2)] \\ & + \lambda\lambda_1 E(S_1^2) + \lambda\lambda_2 E(S_2^2) + 2\lambda_1 E(S_1) + 2\lambda_2 E(S_2) - 2\} \end{aligned} \tag{25}$$

using $\quad F_1(0) = \dfrac{(1-\overline{U})\{\lambda_1 \kappa_1(s) \gamma_1(s)\}}{\{s - \lambda + \lambda_2 \gamma_2(s)\} + \lambda_1 \kappa_1(s) \gamma_1(s)} \qquad (26)$

and from (20) $\quad F_2(0) = 1 - \overline{U} - F_1(0) \qquad (27)$

Numeric methods are required to solve for the positive real root of $D(s)$ which is required to eliminate the singularity of (19).

The AP system state definition Eisenberg (1967) uses is based on service completions. For AP, Eisenberg provides an expression for the probability that a service completion by an arbitrary customer is followed by a changeover. To compute this probability, he uses numerical methods even in the case of zero setup times. The changeover probability is a result of the AP flow time calculations by Eisenberg (1967) which we review after the FCFS flow time calculations by Gaver (1963).

FCFS and AP Flow Time Calculations. Due to the complexity of the computations we provide the necessary background for the reader to replicate results. For both FCFS and AP calculations we provide step-by-step details of the computations leading to the mean flow time. We also include a description of the imbedded state probabilities for the AP model. The changeover probability is pointed out after each wait equation is stated.

FCFS Flow Time Calculations.

1 Determine the positive root of $D(s)$. Using Newton-Raphson method:

1.1 Set $s = 0.9$ as the first guess of the root.

1.2 If $|D(s)| < \varepsilon$, stop and retain positive root, s. Otherwise compute a new estimate for the root using $s = s - D(s)/D'(s)$.

If exponential setup and service distributions,

$$D(s) = \frac{\lambda_1\lambda_2}{[1+sE(S_1)][1+sE(S_2)]} - \frac{\lambda_1\lambda_2}{[1+sE(S_1)][1+sE(S_2)][1+sE(U_1)][1+sE(U_2)]} - \frac{\lambda\lambda_1}{[1+sE(S_1)]} - \frac{\lambda\lambda_2}{[1+sE(S_2)]} + \frac{s\lambda_1}{[1+sE(S_1)]} + \frac{s\lambda_2}{[1+sE(S_2)]} + \lambda^2 - 2\lambda s + s^2 \tag{28}$$

and

$$D'(s) = \lambda_1 \lambda_2 \left[-\frac{E(S_2)}{(1+sE(S_1))} - \frac{E(S_1)}{(1+sE(S_2))} + \frac{E(S_1)}{(1+sE(U_1))(1+sE(U_2))(1+sE(S_2))} \right.$$
$$+ \frac{E(S_2)}{(1+sE(U_1))(1+sE(U_2))(1+sE(S_1))} + \frac{E(U_1)}{(1+sE(S_1))(1+sE(U_2))(1+sE(S_2))}$$
$$\left. + \frac{E(U_2)}{(1+sE(S_1))(1+sE(U_1))(1+sE(S_2))} \right] + \lambda\lambda_1 E(S_1) + \lambda\lambda_2 E(S_2)$$
$$-2\lambda - \lambda_1 sE(S_1) + \frac{\lambda_1}{1+sE(S_1)} + \lambda_2 sE(S_2) + \frac{\lambda_2}{1+sE(S_2)} + 2s. \tag{29}$$

2 Calculate utilization including expected setup:

$$\overline{U} = \rho + \lambda \left[\frac{\lambda_1 \lambda_2}{\lambda^2} \{ E(U_1) + E(U_2) \} \right] \text{ where } \rho = \lambda_1 E(S_1) + \lambda_2 E(S_2) \tag{30}$$

3 $$F_1(0) = \frac{(1-\overline{U})\{\lambda_1 \kappa_1(s) \gamma_1(s)\}}{\{s - \lambda + \lambda_2 \gamma_2(s)\} + \lambda_1 \kappa_1(s) \gamma_1(s)} \text{ using } s \text{ from step 1.2} \tag{31}$$

If exponential setup and service distributions,

$$F_1(0) = \frac{(1-\overline{U})\left\{\lambda_1 \frac{1}{1+sE(U_1)} \cdot \frac{1}{1+sE(S_1)}\right\}}{\left\{s - \lambda + \lambda_2 \frac{1}{1+sE(S_2)}\right\} + \lambda_1 \frac{1}{1+sE(U_1)} \cdot \frac{1}{1+sE(S_1)}} \tag{32}$$

4 $$F_2(0) = 1 - \overline{U} - F_1(0) \tag{33}$$

5 $$E_1(W) = \lim_{s \to 0} (-1) \frac{df_1(s)}{ds}$$

$$= \frac{F_1(0)\{1-\lambda_2 E(S_2)\} + F_2(0)\lambda_1 E(U_1 + S_1)}{\lambda(1-\overline{U})}$$

$$+ \frac{\lambda_1}{2\lambda^2(1-\overline{U})} \times \{\lambda_1\lambda_2 [2E(U_1)E(U_2) + 2E(U_1)E(S_2) + 2E(S_1)E(U_2) \qquad (34)$$

$$+ 2E(U_1)E(S_1) + 2E(U_2)E(S_2) + E(U_1^2) + (U_2^2)]$$

$$+ \lambda\lambda_1 E(S_1^2) + \lambda\lambda_2 E(S_2^2) - 2 + 2\lambda_1 E(S_1) + 2\lambda_2 E(S_2)\}$$

6 $E_2(W) = \lim_{s\to 0}(-1)\frac{df_2(s)}{ds}$

$$= \frac{F_2(0)\{1-\lambda_1 E(S_1)\} + F_1(0)\lambda_2 E(U_2 + S_2)}{\lambda(1-\overline{U})}$$

$$+ \frac{\lambda_2}{2\lambda^2(1-\overline{U})} \times \{\lambda_1\lambda_2 [2E(U_1)E(U_2) + 2E(U_1)E(S_2) + 2E(S_1)E(U_2) \qquad (35)$$

$$+ 2E(U_1)E(S_1) + 2E(U_2)E(S_2) + E(U_1^2) + (U_2^2)]$$

$$+ \lambda\lambda_1 E(S_1^2) + \lambda\lambda_2 E(S_2^2) - 2 + 2\lambda_1 E(S_1) + 2\lambda_2 E(S_2)\}$$

7 Wait in queue prior to processing (does not include setup) is

$$E_1(W) + E_2(W) \qquad (36)$$

8 Wait in queue prior to service (comparable to AP wait) is

$$\overline{W}_{FCFS} = E_1(W) + E_2(W) + \frac{\lambda_1\lambda_2}{\lambda^2}[E(U_1) + E(U_2)] \qquad (37)$$

9 Flow time is

$$\overline{F} = E_1(W) + E_2(W) + \frac{\lambda_1}{\lambda}\left[E(S_1) + \frac{\lambda_2}{\lambda}E(U_2)\right] + \frac{\lambda_2}{\lambda}\left[E(S_2) + \frac{\lambda_1}{\lambda}E(U_1)\right] \qquad (38)$$

The probability that an arbitrary customer is followed by a changeover is

$$\frac{2\lambda_1\lambda_2}{\lambda^2}. \qquad (39)$$

AP Imbedded Markov State Probabilities. Recalling from Chapter 2, Eisenberg considers the imbedded Markov process of queue lengths at the instant of service completion, and includes the class of service just completed. Thus, state ${}^{i}_{mn}$ denotes "server is at line i and m customers are waiting at line 1 and n customers are waiting at line 2." The imbedded process is described as follows.

- State is $(i;m,n)$ where i is customer type of service just completed, m and n are numbers of customers present in queues 1 and 2, respectively.
- Equilibrium probability that an arbitrary service completion leaves the system in state $(i;m,n)$ is π^{i}_{mn}.

Now we define the transition probabilities of the imbedded Markov chain $P\left[(i;m,n)\to\left(i^{'};m^{'},n^{'}\right)\right]$. Using equilibrium equations:

$$\pi^{i'}_{m'n'}=\sum_{i=1}^{2}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\pi^{i}_{mn}P\left[(i;m,n)\to\left(i^{'};m^{'},n^{'}\right)\right] \tag{40}$$

and normalization condition, $$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\pi^{i}_{mn}=a_i=\lambda_i/\lambda\,, \tag{41}$$

the fraction of all possible states left by customer type-i completions (noting $\sum_{i=1}^{2}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\pi^{i}_{mn}=1$). The generating functions of the imbedded state probabilities

are $$\pi^{i}(y,v)\equiv\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\pi^{i}_{mn}y^{m}v^{n}\,. \tag{42}$$

The transition probabilities for the process are defined as p_{ij} = prob(i type-1 customers and j type-2 customers arrive during the service time of a type-1 customer)

$$= \int_0^\infty \left[(\lambda_1 t)^i / i!\right]\left[(\lambda_2 t)^i / j!\right] e^{-(\lambda_1+\lambda_2)t} \mathrm{d}F_{S_1}(t) \tag{43}$$

- q_{ij} = prob(i type-1 customers and j type-2 customers arrive during the service time of a type-2 customer

$$= \int_0^\infty \left[(\lambda_1 t)^i / i!\right]\left[(\lambda_2 t)^i / j!\right] e^{-(\lambda_1+\lambda_2)t} \mathrm{d}F_{S_2}(t) \tag{44}$$

- r_{ij} = prob(i type-1 customers and j type-2 customers arrive during the changeover from 2 to 1)

$$= \int_0^\infty \left[(\lambda_1 t)^i / i!\right]\left[(\lambda_2 t)^i / j!\right] e^{-(\lambda_1+\lambda_2)t} \mathrm{d}F_{U_1}(t) \tag{45}$$

- $R(y,v) \equiv$ generating function of transition probabilities (of type-1 and type-2 arrivals) during type-1 setup so

$$R(y,v) = \kappa_1(\lambda_1 - \lambda_1 y + \lambda_2 - \lambda_2 v) \tag{46}$$

- h_{ij} = prob(i 1-customers and j 2-customers during changeover from 1 to 2

$$= \int_0^\infty \left[(\lambda_1 t)^i / i!\right]\left[(\lambda_2 t)^i / j!\right] e^{-(\lambda_1+\lambda_2)t} \mathrm{d}F_{U_2}(t) \tag{47}$$

- $H(y,v) \equiv$ generating function of transition probabilities (of type-1 and type-2 arrivals) during type-2 setup so

$$H(y,v) = \kappa_2(\lambda_1 - \lambda_1 y + \lambda_2 - \lambda_2 v) \tag{48}$$

$\beta_i(z)$ is the Laplace-Stieltjes transform of the customer type-i busy period distribution function in isolation where

$$\beta_i(z) = \gamma_i(z + \lambda_i - \lambda_i \beta_i(z)). \tag{49}$$

Note: If the service time is exponential then

$$\beta_i(z) = \frac{1}{2\rho_i}\left[1 + \rho_i + \frac{(z)}{\mu_i} - \sqrt{\left(1 + \rho_i + \frac{(z)}{\mu_i}\right)^2 - 4\rho_i}\right] \tag{50}$$

Let $g \equiv$ ratio of number of times the system is emptied by completing service on type-2 customer to type-1 (a constant). We must solve for g because it relates the limits of the generating functions used in the mean wait equation. These generating functions are boundary conditions for the states of the system and are defined as:

$$\begin{aligned}\eta^1(v) = {} & R[\beta_1(\lambda_2 - \lambda_2 v), v]g\eta^2[\beta_1(\lambda_2 - \lambda_2 v)] + a_1\beta_1(\lambda_2 - \lambda_2 v) \\ & + gR[\beta_1(\lambda_2 - \lambda_2 v), v][a_1\beta_1(\lambda_2 - \lambda_2 v) - 1]\end{aligned} \tag{51}$$

$$\begin{aligned}g\eta^2(y) = {} & H[y, \beta_2(\lambda_1 - \lambda_1 y)]\eta^1[\beta_2(\lambda_2 - \lambda_1 y)] + ga_2\beta_2(\lambda_1 - \lambda_1 y) \\ & + H[y, \beta_2(\lambda_1 - \lambda_1 y)][a_2\beta_2(\lambda_1 - \lambda_1 y) - 1]\end{aligned} \tag{52}$$

In solving for g, we also solve for the limiting value of the generating function $\eta^1(1)$. The limits of the generating functions are related using

$$\eta^1(1) - a_1 = g[\eta^2(1) - a_2] \ . \tag{53}$$

Only one value of g leads to a consistent solution of the functional equations. We build the functional equations with many different sizes of their arguments by first initiating them with either $v_0 = 0$ or $y_0 = 0$. We use the fact that $\lim_{i \to \infty} v_i = 1$ which implies $\lim_{i \to \infty} \eta^1(v_i) = \eta^1(1)$ and therefore they converge regardless of the starting point. This is *only* true when g is chosen correctly. The solution is calculated as follows.

Select two arbitrary values of $g : g_k \ k = 1,2$. Since g is a ratio of incidences, restrict $g \geq 0$. For each value of g compute two limiting $\eta^1(v)$ values, $\eta^1(1)$, by calculating it with two different initial conditions: $y_0 = 0$ and $v_0 = 0$ per the procedure below and define the result as follows: $\Gamma_k(v_0 = 0) = \{\eta^1(1) | v_0 = 0, g = g_k\}$ and $\Gamma_k(y_0 = 0)$ similarly.

Using $g = g_k$ set $k = 1$

1 Set $j = 0, y(0) = 0$, let $\eta^1 1(j) = \eta^1(v_j)$ and $\eta^2 1(j) = \eta^2(y_j)$, $\eta^2 1(0) = 1$

1.1 $v(j) = \beta_2[\lambda_1 - \lambda_1 y(j)]$

1.2 $y(j+1) = \beta_1[\lambda_2 - \lambda_2 v(j)]$

Starting iterations are therefore:

$y(0) = 0$, $v(0) = \beta_2[\lambda_1]$, $y(1) = \beta_1[\lambda_2 - \lambda_2 v(0)] = \beta_1[\lambda_2 - \lambda_2 \beta_2[\lambda_1]]$, and $v(1) = \beta_2[\lambda_1 - \lambda_1 y(1)] = \beta_2[\lambda_1 - \lambda_1 \beta_1[\lambda_2 - \lambda_2 \beta_2[\lambda_1]]]$.

1.3 $\eta^1 1(j) = 1 - a_2 v(j) + \dfrac{g[\eta^2 1(j) - a_2 v(j)]}{H(y(j), v(j))}$ where

$$H[y(j), v(j)] = \kappa_2[\lambda_1 - \lambda_1 y(j) + \lambda_2 - \lambda_2 v(j)]$$

since $\eta^1[\beta_2(\lambda_2 - \lambda_1 y)] = 1 - a_2\beta_2(\lambda_1 - \lambda_1 y) + \dfrac{g[\eta^2(y_i) - a_2\beta_2(\lambda_1 - \lambda_1 y)]}{H[y_i, \beta_2(\lambda_1 - \lambda_1 y)]}$ and

$\eta^2(0) = g$

1.4 $\eta^2 1(j+1) = 1 - a_1 y(j+1) + \dfrac{\eta^1 1(j) - a_1 y(j+1)}{gR[y(j+1), v(j)]}$ where

$R[y(j+1), v(j)] = \kappa_1[\lambda_1 - \lambda_1 y(j+1) + \lambda_2 - \lambda_2 v(j)]$

since $\eta^2[\beta_1(\lambda_2 - \lambda_2 v)] = 1 - a_1\beta_1(\lambda_2 - \lambda_2 v) + \dfrac{\eta^1(v) - a_1\beta_1(\lambda_2 - \lambda_2 v)}{gR[\beta_1(\lambda_2 - \lambda_2 v), v]}$

1.5 Assign $\Gamma_k(v_0 = 0) = \{\eta^1 1(j) | g = g_k\}$

1.6 Repeat steps (1.1 – 1.5) until sign of convergence: $|v(j) - v(j-1)| < \varepsilon$

1.7 Retain $\Gamma_k(v_0 = 0) = \eta^1 1(j) | v_0 = 0, g = g_k$ since $\eta^1 1(j)$ at the last value of j represents $\eta^1(1)$

2 Reset $j = 0, v(0) = 0$, let $\eta^1 1(j) = \eta^1(v_j)$ and $\eta^2 1(j) = \eta^2(y_j)$ and $\eta^1 1(0) = 1$

2.1 $y(j) = \beta_1[\lambda_2 - \lambda_2 v(j)]$

2.2 $v(j+1) = \beta_2[\lambda_1 - \lambda_1 y(j)]$

Starting iterations are therefore:

$v(0) = 0$, $y(0) = \beta_1[\lambda_2 - \lambda_2 v(0)] = \beta_1[\lambda_2]$, $v(1) = \beta_2[\lambda_1 - \lambda_1 y(0)] = \beta_2[\lambda_1 - \lambda_1\beta_1[\lambda_2]]$,

and $y(1) = \beta_1[\lambda_2 - \lambda_2 v(1)] = \beta_1[\lambda_2 - \lambda_2\beta_2[\lambda_1 - \lambda_1\beta_1[\lambda_2]]]$.

2.3 $\eta^2 1(j) = 1 - a_1 y(j) + \dfrac{\eta^1 1(j) - a_1 y(j)}{gR[y(j), v(j)]}$ where

$R[y(j), v(j)] = \kappa_1[\lambda_1 - \lambda_1 y(j) + \lambda_2 - \lambda_2 v(j)]$

since $\eta^2[\beta_1(\lambda_2-\lambda_2 v)]=1-a_1\beta_1(\lambda_2-\lambda_2 v)+\dfrac{\eta^1(v)-a_1\beta_1(\lambda_2-\lambda_2 v)}{gR[\beta_1(\lambda_2-\lambda_2 v),v]}$ and

$\eta^1(0)=1$

2.4 $\eta^1 1(j+1)=1-a_2 v(j+1)+\dfrac{g[\eta^2 1(j)-a_2 v(j+1)]}{H(y(j),v(j+1))}$ where

$H[y(j),v(j+1)]=\kappa_2[\lambda_1-\lambda_1 y(j)+\lambda_2-\lambda_2 v(j+1)]$

since $\eta^1[\beta_2(\lambda_2-\lambda_1 y)]=1-a_2\beta_2(\lambda_1-\lambda_1 y)+\dfrac{g[\eta^2(y_i)-a_2\beta_2(\lambda_1-\lambda_1 y)]}{H[y_i,\beta_2(\lambda_1-\lambda_1 y)]}$

2.5 Assign $\Gamma_k(y_0=0)=\{\eta^1 1(j)|g=g_k\}$

2.6 Repeat steps (2.1 – 2.5) until sign of convergence: $|y(j)-y(j-1)|<\varepsilon$

2.7 Retain $\Gamma_k(y_0=0)=\{\eta^1 1(j)|y_0=0,g=g_k\}$ since $\eta^1 1(j)$ at the last value of j represents $\eta^1(1)$

3 Set $k=2$, repeat steps 1 and 2.

4 The convergence is linearly dependent on g so we evaluate the differences in $\eta^1(1)$ starting with $v_0=0$ and $y_0=0$ at the two arbitrary values of g and then get $g=g^*$ by

$$g^*=\frac{g_1[\Gamma_2(y_0=0)-\Gamma_2(v_0=0)]-g_2[\Gamma_1(y_0=0)-\Gamma_1(v_0=0)]}{[\Gamma_2(y_0=0)-\Gamma_2(v_0=0)]-[\Gamma_1(y_0=0)-\Gamma_1(v_0=0)]} \tag{54}$$

5 Set $g=g^*$, repeat steps in section 1 of this procedure above to determine $\Gamma_k(v_0=0)=\{\eta^1 1(j)|v_0=0,g=g^*\}$ which represents $\eta^1(1)$ and using (32) we get $\eta^2(1)$.

At this point we can calculate the idle state probabilities using

$$\pi_{00}^{1} = \frac{1-\rho_1-\rho_2}{1+g+\lambda\left[E(U_1)+E(U_2)\right]\left[\eta^1(1)-a_1\right]}\,. \tag{55}$$

The total idle fraction is then $$\pi_{00}^{1}(1+g)\,. \tag{56}$$

The wait prior to service for a class-1 customer with AP dispatching and non-zero setup is finally:

$$\begin{aligned}\overline{W_1} = {} & \frac{1}{(1-\rho_1-\rho_2)(1-\rho_1-\rho_2+2\rho_1\rho_2)(C_1+C_2)}\times \\ & \Big\{E(U_1)\left[(1-\rho_1)(1-\rho_2)^2+\rho_1\rho_2^2\right]\left[\rho_2 C_1+(1-\rho_1)C_2\right] \\ & +E(U_2)\rho_2(1-\rho_1)\left[(1-\rho_2)C_1+\rho_1 C_2\right] \\ & +\left(\frac{(C_1+C_2)}{2}\right)\left[\left[(1-\rho_1)(1-\rho_2)^2+\rho_1\rho_2^2+\rho_2{}^2\right]\lambda_1 E\left(S_1^2\right)+(1-\rho_1)\lambda_2 E\left(S_2^2\right)\right] \\ & +\left(\frac{C}{2}\right)(1-\rho_1-\rho_2)\left[\left[(1-\rho_1)(1-\rho_2)^2+\rho_1\rho_2^2+\rho_2{}^2\right]E\left(U_1^2\right)+(1-\rho_1)E\left(U_2^2\right)\right]\Big\}\end{aligned} \tag{57}$$

using $$C \equiv \lambda\left[\eta^1(1)-a_1\right],\ C_1 \equiv 1+CE(U_1),\ C_2 \equiv g+CE(U_2)$$

and $\overline{W_2}$ is the same equation with the subscripts switched.

The overall expected wait time is the convex combination of the expected wait times of the two classes: $\overline{W_{AP}} = a_1\overline{W_1}+a_2\overline{W_2}$. The probability that an arbitrary customer is followed by a changeover is $2\pi_{00}^{1}\left(\eta^1(1)-a_1\right)$.

Changeover Comparisons. Each cell in Figure 5-9 and 5-10 contains the FCFS probability of setup above the AP probability of setup. We see in both figures that AP always requires fewer changeovers than FCFS in the zero setup case. The FCFS probability is invariant to Q since from equation (39) the probability of a random arrival requiring a changeover is $\frac{2\lambda_1\lambda_2}{\lambda^2}$. The AP probabilities monotonically decrease with increasing Q at any λ_1 (increasing utilization) and approach zero at saturation. Queue sizes grow with load; therefore, AP has a greater probability of a non-empty queue of the class currently being serviced from which to draw at higher utilization levels. This fact will lead to an increase in system capacity when compared with FCFS when setup is non-zero.

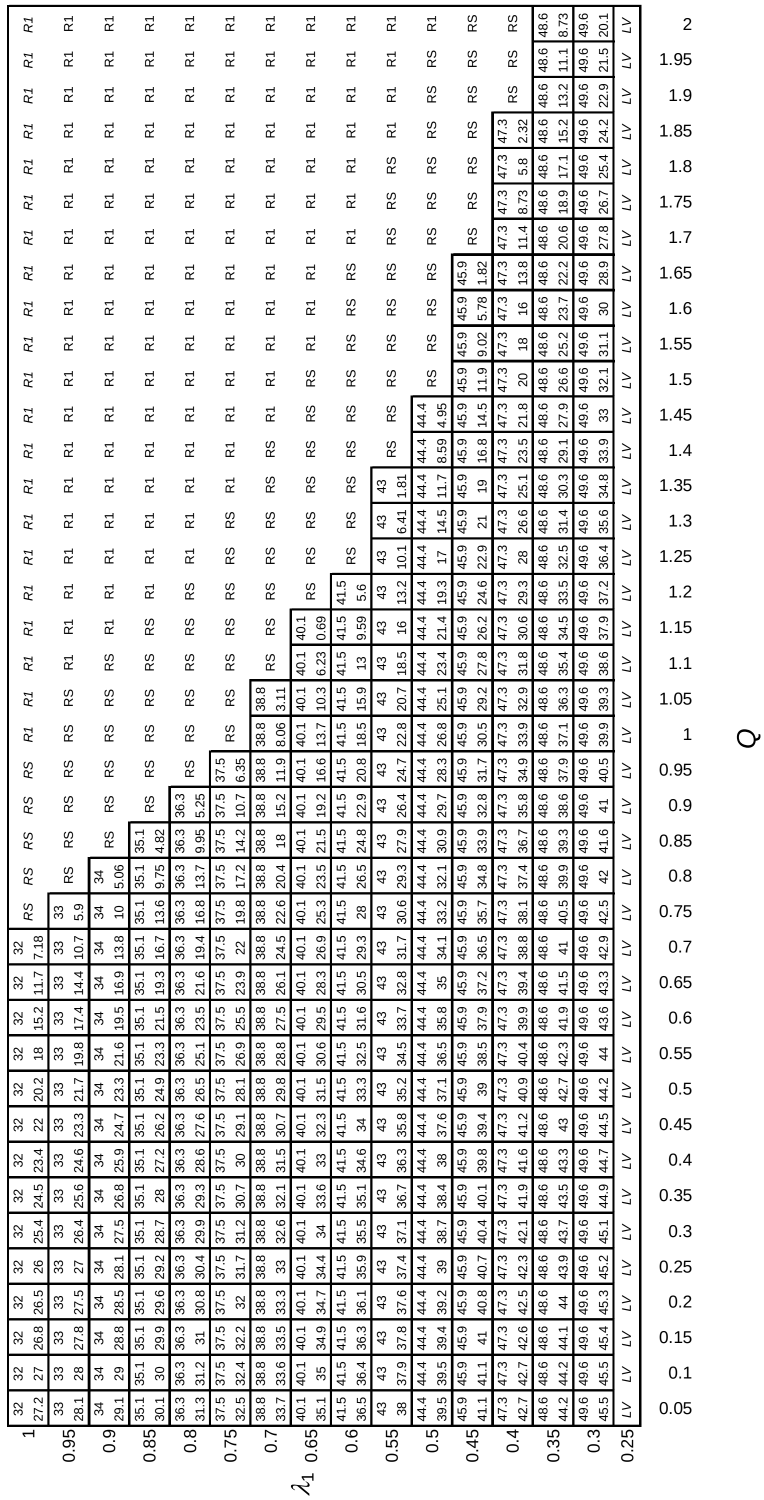

Figure 5-9. Probability of setup (FCFS% above AP%) when setup is zero and $\lambda_2 = 0.25$.

λ_1 \ Q	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1
1	47 25	47 24	47 22	47 20	47 17	47 13	47 8	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*R1*
0.95	47 25	47 25	47 23	47 21	47 18	47 15	47 11	47 4	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*
0.9	48 26	48 25	48 24	48 22	48 20	48 16	48 13	48 7	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*
0.85	49 27	49 26	49 25	49 23	49 21	49 18	49 14	49 10	49 4	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*
0.8	49 27	49 26	49 25	49 24	49 22	49 19	49 16	49 12	49 8	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*
0.75	49 28	49 27	49 26	49 25	49 23	49 21	49 18	49 15	49 11	49 5	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*
0.7	50 28	50 28	50 27	50 25	50 24	50 22	50 20	50 17	50 13	50 9	50 4	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*
0.65	50 29	50 28	50 27	50 26	50 25	50 23	50 21	50 19	50 16	50 12	50 8	50 3	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*	*RS*
0.6	*LV*	*LV*	*LV*	*LV*	*LV*	*LV*	*LV*	*LV*	*LV*	*LV*	*LV*	*LV*	*LV*	*LV*	*LV*	*LV*	*LV*	*LV*	*LV*	*LV*

Figure 5-10. Probability of setup (FCFS% above AP%) when setup is zero and $\lambda_2 = 0.60$.

We add setup time in a way consistent with our analyses of Chapters 3 and 4, using the setup fraction, as a ratio of the expected batch service time (with a batch size of one). We evaluate a range of setup magnitudes starting with a very low setup fraction of 0.001. Our highest level is 1.0, the level we use in our operational standardization in Chapter 3 and Chapter 4. We compare the baseline AP-FCFS wait differences of Figure 5-7 to non-zero setup using numerical methods. We identify regions of interest that we explain as follows. The FCFS system stability is limited as stated in Chapter 2 by

$$0 \le \overline{U} = \lambda_1 E(S_1) + \lambda_2 E(S_2) + \frac{\lambda_1 \lambda_2}{\lambda}\left[E(U_1) + E(U_2)\right] < 1,$$

but the AP system is only limited by $0 \le \lambda E(S) = \lambda_1 E(S_1) + \lambda_2 E(S_2) < 1$. We identify this disparity in system capacity for the non-zero setup cases in the figures by AP.

The first comparison is made for the symmetric cases where $\lambda_1 = \lambda_2$, $E(S_1) = E(S_2)$, and $E(U_1) = E(U_2)$. Figure 5-11 with $\lambda_i = 0.2$, is characteristic of the symmetric comparisons. We immediately see that in the presence of setup AP always requires less wait than FCFS and without setup ($U_i = 0.0$), there is no difference between AP and FCFS wait. We also note that the AP wait is monotonically better than FCFS with both increasing setup and service. As setup is introduced AP will minimize the changeovers and in the symmetric case provide lower wait times. Given any fixed λ_i, as setup and service times increase so does the utilization and, thus, expected lengths of the queues. AP by avoiding changeovers is able to provide a stable system in areas where FCFS is saturated.

$E(S_i)$ \ $E(U_i)$	0.0	0.001	0.01	0.1	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
2.0	0	-0.8	-8.1	-91.71	-938.3	-1.E+3	-2.E+3	-4.E+3	-1.E+4	AP	AP	AP	AP	AP	AP	AP	AP	AP	AP	AP
1.9	0	-0.5	-5.08	-56.58	-511	-741.1	-1.E+3	-2.E+3	-2.E+3	-4.E+3	-1.E+4	AP	AP	AP	AP	AP	AP	AP	AP	AP
1.8	0	-0.33	-3.34	-36.92	-310.3	-435.3	-603.2	-835.8	-1.E+3	-2.E+3	-3.E+3	-4.E+3	-1.E+4	AP	AP	AP	AP	AP	AP	AP
1.7	0	-0.22	-2.28	-25.04	-201.2	-277	-374.6	-502.3	-673.3	-909.7	-1.E+3	-2.E+3	-3.E+3	-4.E+3	-1.E+4	AP	AP	AP	AP	AP
1.6	0	-0.16	-1.59	-17.45	-136.3	-185.6	-247.3	-325.5	-425.9	-557	-732.3	-974	-1.E+3	-2.E+3	-3.E+3	-5.E+3	-1.E+4	AP	AP	AP
1.5	0	-0.11	-1.14	-12.38	-95.21	-128.7	-170.1	-221.4	-285.6	-366.7	-470.7	-606.2	-786.9	-1.E+3	-1.E+3	-2.E+3	-3.E+3	-5.E+3	-1.E+4	AP
1.4	0	-0.08	-0.82	-8.91	-67.9	-91.46	-120.3	-155.6	-199	-252.9	-320.1	-404.8	-513.1	-653.9	-841.3	-1.E+3	-1.E+3	-2.E+3	-3.E+3	-5.E+3
1.3	0	-0.06	-0.59	-6.45	-49.14	-66.1	-86.73	-111.9	-142.5	-180	-226	-282.8	-353.6	-442.5	-555.9	-703.1	-898.6	-1.E+3	-2.E+3	-2.E+3
1.2	0	-0.04	-0.43	-4.69	-35.9	-48.33	-63.42	-81.74	-104	-131	-163.8	-203.8	-252.7	-313	-387.9	-481.8	-601.1	-755.6	-960.6	-1.E+3
1.1	0	-0.03	-0.32	-3.4	-26.37	-35.59	-46.81	-60.42	-76.91	-96.84	-121	-150.1	-185.4	-228.3	-280.7	-345	-424.5	-523.9	-650.1	-813.1
1.0	0	-0.02	-0.23	-2.46	-19.41	-26.31	-34.75	-45	-57.44	-72.46	-90.58	-112.4	-138.7	-170.3	-208.5	-254.7	-310.9	-379.6	-464.4	-570.1
0.9	0	-0.02	-0.16	-1.76	-14.26	-19.46	-25.86	-33.68	-43.18	-54.68	-68.55	-85.23	-105.2	-129.2	-157.9	-192.4	-233.7	-283.6	-344	-417.7
0.8	0	-0.01	-0.11	-1.24	-10.43	-14.36	-19.25	-25.25	-32.59	-41.51	-52.29	-65.25	-80.78	-99.34	-121.5	-147.8	-179.3	-216.8	-261.7	-315.6
0.7	0	-0.01	-0.08	-0.86	-7.57	-10.55	-14.29	-18.94	-24.67	-31.66	-40.13	-50.34	-62.59	-77.21	-94.61	-115.3	-139.8	-168.8	-203.2	-244.2
0.6	0	-0.01	-0.05	-0.57	-5.42	-7.69	-10.57	-14.19	-18.69	-24.22	-30.96	-39.11	-48.89	-60.59	-74.48	-90.96	-110.4	-133.4	-160.5	-192.4
0.5	0	-0.01	-0.03	-0.37	-3.84	-5.56	-7.78	-10.62	-14.17	-18.59	-24	-30.57	-38.49	-47.96	-59.22	-72.55	-88.3	-106.8	-128.6	-154.1
0.4	0	0	-0.02	-0.21	-2.67	-3.98	-5.69	-7.93	-10.77	-14.32	-18.71	-24.07	-30.54	-38.3	-47.55	-58.49	-71.4	-86.55	-104.3	-125.1
0.3	0	0	-0.01	-0.12	-1.82	-2.82	-4.16	-5.93	-8.22	-11.11	-14.69	-19.11	-24.45	-30.88	-38.56	-47.66	-58.38	-70.96	-85.67	-102.8
0.2	0	0	-0.01	-0.05	-1.23	-1.99	-3.06	-4.48	-6.33	-8.7	-11.66	-15.33	-19.8	-25.19	-31.63	-39.28	-48.3	-58.87	-71.21	-85.57
0.1	0	0	0	-0.02	-0.84	-1.43	-2.28	-3.43	-4.96	-6.92	-9.41	-12.49	-16.26	-20.82	-26.29	-32.8	-40.45	-49.44	-59.92	-72.1

Figure 5-11. Wait time differences (AP-FCFS)*100 for symmetric cases when $\lambda_i = 0.2$.

Figures 5-12 through 5-15 show a progression of the effects of setup when $\lambda_2 = 0.25$. With minimum setup added $(E(U) = 0.001 * E(S))$ we see the equality at $Q = 1$ has been replaced entirely by AP. The dominance of AP, where only AP yields the lesser wait as compared to FCFS, is quickly realized. We note that only 5% setup is needed for AP to dominate the $Q > 1$ region as shown in Figure 5-13. As we expect from the stability limits, the AP win area increases with setup magnitude, especially approaching the region of instability.

λ_1 \ Q	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1
1	8.37	7.7	6.85	5.79	4.47	2.81	0.71	-2	-5.5	-10	-17	-28	-49	-114	RS	RS	RS	RS	RS	R1
0.95	8.28	7.64	6.84	5.85	4.63	3.13	1.29	-1	-3.9	-7.7	-13	-20	-32	-56	-145	RS	RS	RS	RS	RS
0.9	8.2	7.58	6.81	5.89	4.77	3.42	1.79	-0.2	-2.6	-5.6	-9.4	-15	-22	-34	-58	-166	RS	RS	RS	RS
0.85	8.1	7.5	6.78	5.91	4.88	3.66	2.22	0.52	-1.5	-3.9	-6.9	-11	-15	-22	-33	-54	-155	RS	RS	RS
0.8	7.98	7.41	6.73	5.92	4.97	3.87	2.6	1.11	-0.6	-2.6	-4.9	-7.6	-11	-15	-20	-28	-44	-107	RS	RS
0.75	7.87	7.31	6.66	5.91	5.04	4.04	2.91	1.61	0.16	-1.5	-3.3	-5.4	-7.8	-10	-14	-17	-22	-29	-53	RS
0.7	7.73	7.2	6.59	5.89	5.08	4.18	3.16	2.04	0.8	-0.6	-2.1	-3.7	-5.4	-7.3	-9.2	-11	-13	-14	-15	-14
0.65	7.58	7.08	6.49	5.84	5.1	4.28	3.38	2.39	1.32	0.16	-1.1	-2.4	-3.7	-5	-6.3	-7.4	-8.2	-8.3	-7.1	-3.3
0.6	7.42	6.93	6.38	5.77	5.09	4.35	3.54	2.66	1.74	0.77	-0.2	-1.3	-2.3	-3.3	-4.2	-4.9	-5.3	-5.2	-4.1	-1.3
0.55	7.23	6.76	6.24	5.67	5.05	4.38	3.66	2.89	2.09	1.25	0.4	-0.5	-1.3	-2.1	-2.7	-3.3	-3.5	-3.3	-2.5	-0.7
0.5	7.01	6.57	6.09	5.56	4.99	4.38	3.73	3.05	2.34	1.63	0.9	0.19	-0.5	-1.1	-1.7	-2.1	-2.2	-2.1	-1.6	-0.4
0.45	6.76	6.35	5.9	5.4	4.89	4.33	3.75	3.15	2.54	1.92	1.3	0.7	0.12	-0.4	-0.8	-1.2	-1.3	-1.3	-1	-0.2
0.4	6.48	6.08	5.67	5.22	4.75	4.24	3.73	3.21	2.66	2.13	1.59	1.07	0.58	0.14	-0.2	-0.5	-0.7	-0.8	-0.6	-0.2
0.35	6.14	5.77	5.39	4.99	4.56	4.11	3.65	3.19	2.72	2.25	1.79	1.35	0.92	0.54	0.2	-0.1	-0.3	-0.4	-0.3	-0.1
0.3	5.75	5.41	5.06	4.69	4.32	3.92	3.51	3.12	2.7	2.3	1.9	1.52	1.14	0.81	0.51	0.24	0.05	-0.1	-0.1	-0.1
0.25	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.2	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.15	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.1	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.05	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV

λ_1 \ Q	1.05	1.1	1.15	1.2	1.25	1.3	1.35	1.4	1.45	1.5	1.55	1.6	1.65	1.7	1.75	1.8	1.85	1.9	1.95	2
1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.95	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.9	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.85	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.8	RS	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.75	RS	RS	RS	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.7	-95	RS	RS	RS	RS	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.65	6.82	33.4	-7.E+3	RS	RS	RS	RS	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.6	4.4	16.2	42.51	115	RS	RS	RS	RS	RS	RS	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1
0.55	2.68	8.63	19.13	38.4	77.6	176	368	RS	RS	RS	RS	RS	RS	RS	RS	RS	R1	R1	R1	R1
0.5	1.65	5	10.3	18.7	32	54.3	94.3	177	408	RS	RS	RS	RS	RS	RS	RS	RS	RS	RS	R1
0.45	1	2.97	5.93	10.3	16.5	25.6	39	59.3	91.2	145	248	499	1515	RS	RS	RS	RS	RS	RS	RS
0.4	0.58	1.74	3.44	5.83	9.13	13.6	19.7	28	39.4	55	77.2	109	158	239	386	722	2026	RS	RS	RS
0.35	0.3	0.96	1.91	3.25	5.05	7.44	10.5	14.5	19.7	26.3	34.7	45.6	59.8	78.5	103	137	185	255	365	554
0.3	0.11	0.43	0.94	1.65	2.62	3.88	5.5	7.52	10.1	13.2	17	21.7	27.5	34.5	43.1	53.7	66.8	82.9	103	129
0.25	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.2	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.15	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.1	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.05	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV

Figure 5-12. Wait time differences (AP-FCFS)*100 when E(U)=0.001*E(S) and $\lambda_2 = 0.25$.

We note an interesting pattern in the figures illustrating the wait differences. Figure 5-12 at the $\lambda_1 = 0.65$ setting shows the AP-FCFS wait to be *negative* just prior to the region of saturation. This pattern is also seen in Figure 5-16 for three settings: $\lambda_1 = 0.40$, $\lambda_1 = 0.45$, and $\lambda_1 = 0.50$, but not in Figure-17. To explain this pattern we show the actual wait times for AP and then FCFS for $\lambda_1 = 0.50$ at four levels of setup magnitude in Figures 5-13 and 5-14. We then follow with a plot of the difference in flow time for the $\lambda_1 = 0.50$.

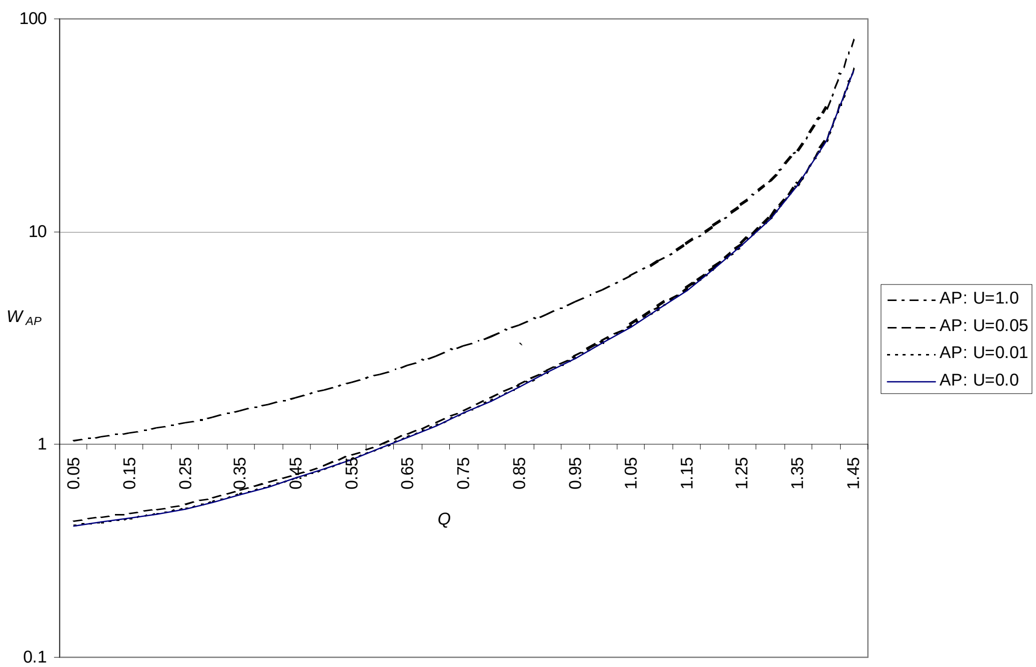

Figure 5-13. Wait time when $\lambda_1 = 0.50$, $\lambda_2 = 0.25$ using AP.

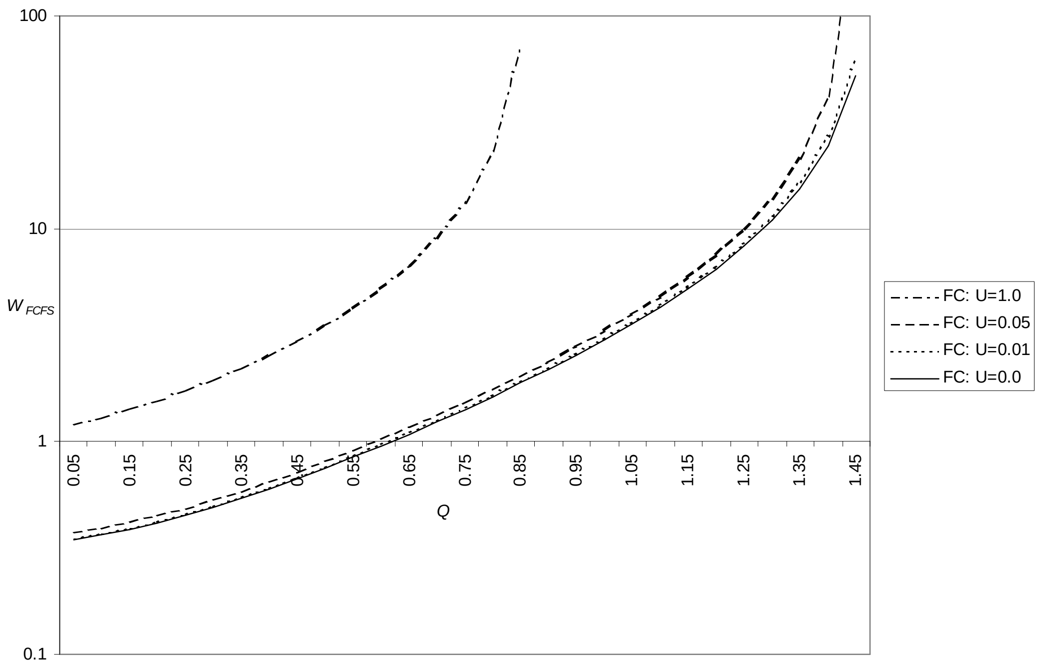

Figure 5-14. Wait time when $\lambda_1 = 0.50$, $\lambda_2 = 0.25$ using FCFS.

If we superimpose the wait time curves we find that the FCFS curves are steeper at the same $Q = E(S_1)$. This is because the wait time, driven by congestion, is a function of both service and setup times and AP pays less setup than FCFS. This steeper slope near saturation causes the wait curves to intersect. We show the case of $E(U) = 0.01 * E(S)$ and identify three points of intersection of the two curves. This does not happen with greater setup magnitude because the FCFS wait curve is shifted up, intersecting the AP wait curve in only one place. Figure 5-15 shows the three points of intersection for the $\lambda_1 = 0.50$, $\lambda_2 = 0.25$ and $E(U) = 0.01 * E(S)$ case.

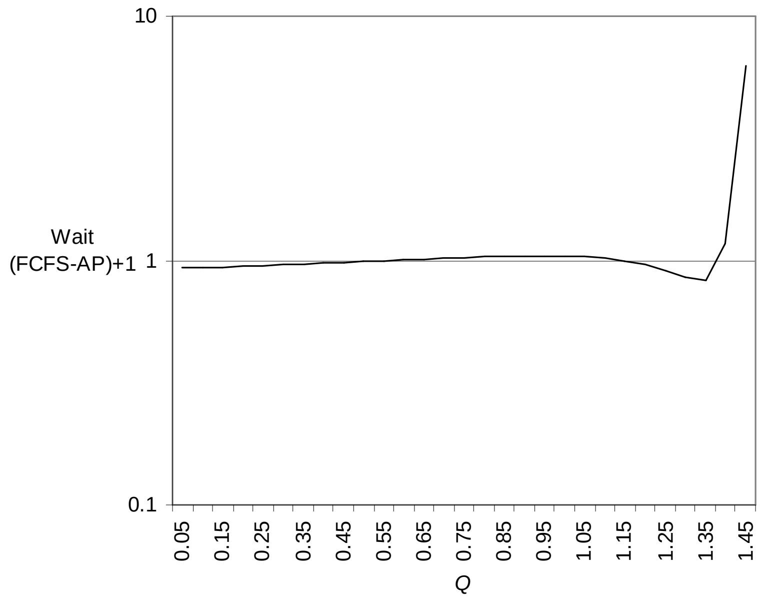

Figure 5-15. Three zeroes of intersection between wait curves of AP and FCFS when $\lambda_1 = 0.50$, $\lambda_2 = 0.25$ and $E(U) = 0.01 * E(S)$.

Q \ λ_1	1	0.95	0.9	0.85	0.8	0.75	0.7	0.65	0.6	0.55	0.5	0.45	0.4	0.35	0.3	0.25	0.2	0.15	0.1	0.05
2	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	RS	RS	266	104	LV	LV	LV	LV	LV
1.95	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	RS	RS	RS	220	84.1	LV	LV	LV	LV	LV
1.9	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	RS	RS	RS	171	67.9	LV	LV	LV	LV	LV
1.85	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	RS	RS	-1.E+4	131.2	54.75	LV	LV	LV	LV	LV
1.8	R1	R1	R1	R1	R1	R1	R1	R1	R1	RS	RS	RS	-69	101	44	LV	LV	LV	LV	LV
1.75	R1	R1	R1	R1	R1	R1	R1	R1	R1	RS	RS	RS	138	77.3	35.2	LV	LV	LV	LV	LV
1.7	R1	R1	R1	R1	R1	R1	R1	R1	R1	RS	RS	RS	124	59.3	28	LV	LV	LV	LV	LV
1.65	R1	R1	R1	R1	R1	R1	R1	R1	RS	RS	RS	-2.E+4	94.62	45.43	22.08	LV	LV	LV	LV	LV
1.6	R1	R1	R1	R1	R1	R1	R1	R1	RS	RS	RS	-193	69.9	34.6	17.2	LV	LV	LV	LV	LV
1.55	R1	R1	R1	R1	R1	R1	R1	R1	RS	RS	RS	53.3	51	26.1	13.2	LV	LV	LV	LV	LV
1.5	R1	R1	R1	R1	R1	R1	R1	RS	RS	RS	RS	58.9	36.8	19.4	9.99	LV	LV	LV	LV	LV
1.45	R1	R1	R1	R1	R1	R1	R1	RS	RS	RS	-525	44.5	26.2	14.2	7.35	LV	LV	LV	LV	LV
1.4	R1	R1	R1	R1	R1	R1	RS	RS	RS	RS	-18	31	18.3	10.1	5.24	LV	LV	LV	LV	LV
1.35	R1	R1	R1	R1	R1	R1	RS	RS	RS	-2.E+4	17	20.6	12.38	6.93	3.54	LV	LV	LV	LV	LV
1.3	R1	R1	R1	R1	R1	RS	RS	RS	RS	-215	14.9	13	7.97	4.48	2.22	LV	LV	LV	LV	LV
1.25	R1	R1	R1	R1	R1	RS	RS	RS	RS	-29	9	7.51	4.71	2.62	1.19	LV	LV	LV	LV	LV
1.2	R1	R1	R1	R1	RS	RS	RS	RS	-390	-7.6	4.04	3.65	2.33	1.23	0.43	LV	LV	LV	LV	LV
1.15	R1	R1	R1	RS	RS	RS	RS	AP	-64	-5.3	0.48	0.98	0.64	0.23	-0.1	LV	LV	LV	LV	LV
1.1	R1	R1	RS	RS	RS	RS	RS	-295	-26	-5.9	-1.9	-0.8	-0.5	-0.4	-0.5	LV	LV	LV	LV	LV
1.05	R1	RS	RS	RS	RS	RS	-2.E+3	-68.7	-17	-6.68	-3.31	-1.92	-1.24	-0.88	-0.68	LV	LV	LV	LV	LV
1	R1	RS	RS	RS	RS	RS	-152	-34	-14	-7	-4.1	-2.5	-1.7	-1.1	-0.7	LV	LV	LV	LV	LV
0.95	RS	RS	RS	RS	RS	-296	-56	-23	-12	-6.9	-4.3	-2.8	-1.8	-1.2	-0.7	LV	LV	LV	LV	LV
0.9	RS	RS	RS	RS	-483	-81	-33	-17	-10	-6.5	-4.2	-2.8	-1.7	-1.1	-0.6	LV	LV	LV	LV	LV
0.85	RS	RS	RS	-585	-101	-42	-23	-14	-8.9	-5.9	-3.9	-2.5	-1.6	-0.9	-0.4	LV	LV	LV	LV	LV
0.8	RS	RS	-502	-109	-49	-27	-17	-11	-7.5	-5	-3.3	-2.1	-1.2	-0.6	-0.1	LV	LV	LV	LV	LV
0.75	RS	-340	-103	-51	-30	-19	-13	-8.9	-6.1	-4.1	-2.7	-1.6	-0.8	-0.2	0.18	LV	LV	LV	LV	LV
0.7	-213	-88	-49	-31	-20	-14	-9.8	-6.9	-4.7	-3.1	-1.9	-1	-0.4	0.16	0.52	LV	LV	LV	LV	LV
0.65	-70	-43	-29	-20	-14	-10	-7.2	-5	-3.4	-2.1	-1.1	-0.4	0.18	0.61	0.9	LV	LV	LV	LV	LV
0.6	-36	-26	-18	-13	-9.7	-7	-5	-3.3	-2.1	-1.1	-0.3	0.27	0.74	1.07	1.3	LV	LV	LV	LV	LV
0.55	-21	-16	-12	-8.7	-6.3	-4.5	-3	-1.8	-0.9	-0.1	0.48	0.95	1.31	1.56	1.71	LV	LV	LV	LV	LV
0.5	-13	-9.5	-7.1	-5.1	-3.6	-2.3	-1.3	-0.4	0.28	0.84	1.29	1.63	1.88	2.05	2.14	LV	LV	LV	LV	LV
0.45	-6.9	-5.1	-3.6	-2.3	-1.3	-0.4	0.28	0.87	1.37	1.76	2.07	2.3	2.46	2.55	2.56	LV	LV	LV	LV	LV
0.4	-2.9	-1.8	-0.8	-0.1	0.6	1.17	1.64	2.04	2.37	2.63	2.82	2.96	3.03	3.05	2.99	LV	LV	LV	LV	LV
0.35	0.11	0.76	1.31	1.8	2.22	2.57	2.86	3.11	3.3	3.44	3.54	3.59	3.59	3.54	3.41	LV	LV	LV	LV	LV
0.3	2.4	2.76	3.08	3.35	3.59	3.79	3.95	4.07	4.16	4.21	4.23	4.19	4.13	4.01	3.83	LV	LV	LV	LV	LV
0.25	4.17	4.36	4.51	4.65	4.76	4.84	4.9	4.93	4.94	4.92	4.86	4.78	4.65	4.47	4.23	LV	LV	LV	LV	LV
0.2	5.58	5.65	5.7	5.74	5.75	5.76	5.75	5.71	5.65	5.56	5.46	5.32	5.14	4.92	4.63	LV	LV	LV	LV	LV
0.15	6.71	6.7	6.68	6.65	6.61	6.55	6.48	6.39	6.28	6.16	6	5.82	5.59	5.33	5	LV	LV	LV	LV	LV
0.1	7.6	7.54	7.47	7.41	7.33	7.23	7.12	6.99	6.85	6.69	6.51	6.29	6.03	5.73	5.36	LV	LV	LV	LV	LV
0.05	8.29	8.21	8.12	8.03	7.92	7.8	7.67	7.52	7.36	7.17	6.96	6.71	6.43	6.1	5.71	LV	LV	LV	LV	LV

Figure 5-16. Wait time differences (AP-FCFS)*100 when E(U)=0.01*E(S) and $\lambda_2 = 0.25$.

λ_1 \ Q	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1
1	7.92	7.08	6	4.59	2.79	0.41	-2.8	-7.2	-14	-24	-41	-78	-185	-887	RS	RS	RS	RS	RS	R1
0.95	7.86	7.05	6.02	4.72	3.07	0.96	-1.8	-5.5	-11	-18	-31	-53	-103	-267	-2.E+3	RS	RS	RS	RS	RS
0.9	7.78	7	6.04	4.83	3.33	1.44	-0.9	-4.1	-8.3	-14	-23	-37	-64	-128	-361	-4.E+3	RS	RS	RS	RS
0.85	7.7	6.95	6.04	4.93	3.56	1.88	-0.2	-2.9	-6.3	-11	-17	-27	-43	-74	-149	-434	-6.E+3	RS	RS	RS
0.8	7.6	6.89	6.03	4.99	3.76	2.24	0.41	-1.9	-4.7	-8.3	-13	-20	-30	-47	-79.9	-158	-446	-5.E+3	RS	RS
0.75	7.49	6.82	6.01	5.06	3.92	2.57	0.97	-1	-3.3	-6.2	-9.9	-15	-22	-32	-48.6	-79.9	-152	-389	-2.E+3	RS
0.7	7.38	6.73	5.97	5.09	4.05	2.84	1.44	-0.2	-2.2	-4.5	-7.4	-11	-16	-22	-31.8	-46.8	-73.9	-132	-295	-1.E+3
0.65	7.24	6.63	5.92	5.09	4.16	3.08	1.84	0.41	-1.2	-3.1	-5.4	-8.2	-12	-16	-21.6	-29.8	-42.3	-63.5	-105	-203
0.6	7.09	6.51	5.85	5.09	4.23	3.27	2.18	0.95	-0.4	-2	-3.8	-5.9	-8.4	-11	-15.1	-19.9	-26.4	-36	-51	-77.5
0.55	6.92	6.37	5.74	5.05	4.27	3.41	2.45	1.39	0.22	-1.1	-2.6	-4.2	-6	-8.2	-10.7	-13.7	-17.4	-22.2	-28.9	-38.6
0.5	6.72	6.21	5.62	4.98	4.28	3.51	2.67	1.75	0.75	-0.3	-1.5	-2.8	-4.2	-5.8	-7.52	-9.48	-11.8	-14.5	-17.8	-22
0.45	6.49	6.01	5.47	4.89	4.25	3.57	2.83	2.03	1.18	0.27	-0.7	-1.7	-2.8	-4	-5.24	-6.59	-8.05	-9.68	-11.5	-13.6
0.4	6.23	5.77	5.28	4.76	4.19	3.58	2.92	2.25	1.52	0.76	-0	-0.9	-1.7	-2.6	-3.56	-4.52	-5.53	-6.56	-7.65	-8.8
0.35	5.92	5.5	5.05	4.57	4.06	3.53	2.97	2.37	1.76	1.13	0.47	-0.2	-0.9	-1.6	-2.3	-3.03	-3.74	-4.45	-5.16	-5.86
0.3	5.54	5.16	4.76	4.33	3.89	3.42	2.93	2.44	1.91	1.38	0.84	0.29	-0.3	-0.8	-1.38	-1.92	-2.46	-2.98	-3.47	-3.94
0.25	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.2	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.15	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.1	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.05	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV

λ_1 \ Q	1.05	1.1	1.15	1.2	1.25	1.3	1.35	1.4	1.45	1.5	1.55	1.6	1.65	1.7	1.75	1.8	1.85	1.9	1.95	2
1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.95	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.9	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.85	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.8	RS	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.75	RS	RS	RS	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.7	AP	RS	RS	RS	RS	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.65	-538	-4.E+3	AP	RS	RS	RS	RS	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.6	-131	-267	-767	-7.E+3	RS	RS	RS	RS	RS	RS	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1
0.55	-54.2	-81.6	-137	-271	-722	-4.E+3	AP	RS	RS	RS	RS	RS	RS	RS	RS	RS	R1	R1	R1	R1
0.5	-27.8	-36.2	-49.4	-71.7	-114	-206	-455	-1475	-2.E+4	RS	RS	RS	RS	RS	RS	RS	RS	RS	RS	R1
0.45	-16.1	-19.3	-23.4	-29.2	-38	-51.9	-76	-122	-220	-473	-1386	-1.E+4	AP	RS	RS	RS	RS	RS	RS	RS
0.4	-10	-11.4	-12.9	-14.7	-17	-20	-24.3	-30.7	-40.8	-57.9	-88.8	-149	-279	-613	-1793	-1.E+4	AP	RS	RS	RS
0.35	-6.54	-7.21	-7.86	-8.51	-9.1	-9.84	-10.6	-11.5	-12.7	-14.5	-17.1	-21.4	-28.3	-40	-61.2	-99.8	-175	-337	-737	-2.E+3
0.3	-4.38	-4.78	-5.12	-5.41	-5.6	-5.76	-5.82	-5.8	-5.68	-5.47	-5.19	-4.85	-4.51	-4.3	-4.22	-4.66	-5.94	-8.8	-14	-24.5
0.25	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.2	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.15	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.1	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.05	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV

Figure 5-17. Wait time differences (AP-FCFS)*100 when E(U)=0.05*E(S) and $\lambda_2 = 0.25$.

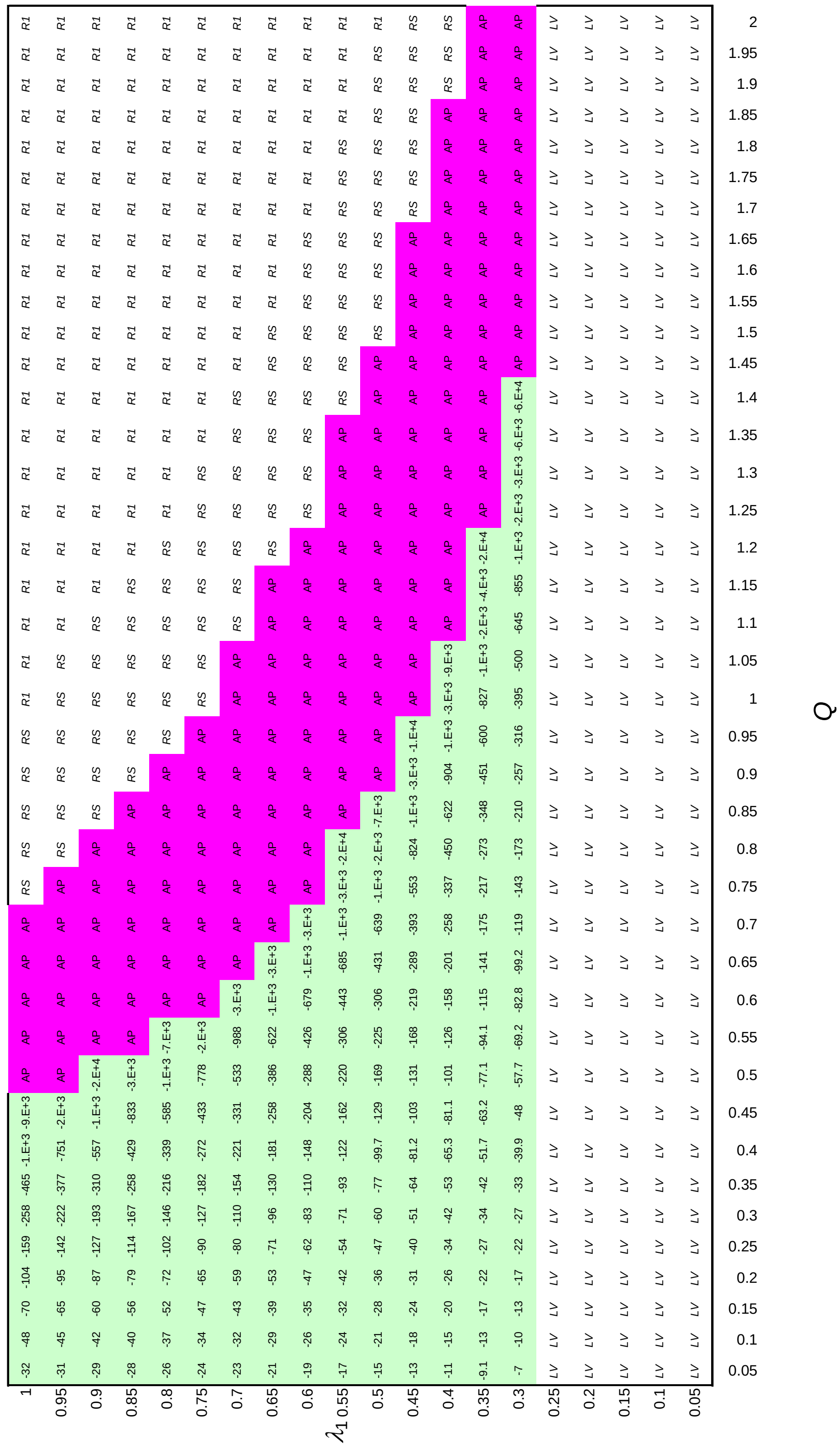

Figure 5-18. Wait time differences (AP-FCFS)*100 when E(U)=1.0*E(S) and $\lambda_2 = 0.25$.

Under certain circumstances FCFS will provide less wait than AP even when setup time is non-zero. There are two regions, one characterized by $E(S_2) > E(S_1)$ with $\lambda_1 > \lambda_2$ and the other $E(S_1)/E(S_2) = Q < 0.5$ with $\lambda_1 > \lambda_2$. Both of these regions decrease in size with increasing setup as shown in Figures 5-12 and 5-16 through 5-18 and 5-19 through 5-21 such that when $E(U) = E(S)$, AP dominates the entire feasible space.

λ_1 \ Q	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1
1	1.03	0.89	0.67	0.37	-0.1	-0.6	-1.3	-2.2	-3.3	-4.9	-6.9	-9.7	-14	-20	-30	-52	-138	RS	RS	R1
0.95	1.02	0.89	0.68	0.4	0.03	-0.5	-1.1	-1.8	-2.8	-4.1	-5.6	-7.7	-10	-14	-20	-30	-51	-141	RS	RS
0.9	1.02	0.89	0.7	0.44	0.1	-0.3	-0.9	-1.5	-2.4	-3.4	-4.6	-6.2	-8.2	-11	-14	-19	-27	-44	-116	RS
0.85	1.02	0.9	0.71	0.47	0.17	-0.2	-0.7	-1.3	-2	-2.8	-3.8	-5	-6.4	-8.2	-10	-13	-16	-21	-31	-71
0.8	1.01	0.9	0.73	0.51	0.22	-0.1	-0.5	-1	-1.6	-2.3	-3.1	-4	-5.1	-6.3	-7.7	-9.2	-11	-13	-14	-16
0.75	1.01	0.89	0.73	0.53	0.28	-0	-0.4	-0.8	-1.3	-1.9	-2.5	-3.3	-4.1	-4.9	-5.8	-6.7	-7.5	-8	-8	-6.7
0.7	1	0.9	0.74	0.56	0.33	0.06	-0.3	-0.6	-1.1	-1.5	-2.1	-2.6	-3.2	-3.8	-4.4	-5	-5.3	-5.4	-5	-3.5
0.65	0.99	0.89	0.75	0.59	0.37	0.13	-0.2	-0.5	-0.8	-1.2	-1.6	-2.1	-2.5	-3	-3.4	-3.7	-3.9	-3.8	-3.3	-2.1
0.6	0.99	0.89	0.76	0.6	0.41	0.2	-0.1	-0.3	-0.6	-1	-1.3	-1.7	-2	-2.3	-2.6	-2.8	-2.9	-2.7	-2.3	-1.3
0.55	0.97	0.88	0.76	0.62	0.45	0.26	0.05	-0.2	-0.5	-0.7	-1	-1.3	-1.6	-1.8	-2	-2.1	-2.1	-2	-1.6	-0.9
0.5	0.96	0.87	0.77	0.64	0.48	0.31	0.12	-0.1	-0.3	-0.5	-0.8	-1	-1.2	-1.4	-1.5	-1.6	-1.6	-1.5	-1.2	-0.6
0.45	0.95	0.87	0.77	0.64	0.51	0.36	0.2	0.02	-0.2	-0.4	-0.5	-0.7	-0.9	-1	-1.1	-1.2	-1.2	-1.1	-0.8	-0.4
0.4	0.93	0.85	0.76	0.65	0.53	0.4	0.26	0.11	-0	-0.2	-0.4	-0.5	-0.6	-0.8	-0.8	-0.9	-0.9	-0.8	-0.6	-0.3
0.35	0.9	0.83	0.75	0.66	0.54	0.44	0.31	0.18	0.05	-0.1	-0.2	-0.3	-0.4	-0.5	-0.6	-0.6	-0.6	-0.6	-0.4	-0.2
0.3	0.87	0.8	0.73	0.65	0.56	0.46	0.36	0.24	0.13	0.03	-0.1	-0.2	-0.3	-0.3	-0.4	-0.4	-0.4	-0.4	-0.3	-0.2
0.25	0.83	0.77	0.71	0.64	0.55	0.47	0.38	0.3	0.2	0.12	0.03	-0.1	-0.1	-0.2	-0.2	-0.3	-0.3	-0.3	-0.2	-0.1
0.2	0.78	0.72	0.67	0.61	0.53	0.46	0.39	0.33	0.25	0.18	0.11	0.04	-0	-0.1	-0.1	-0.1	-0.1	-0.2	-0.1	-0.1
0.15	0.7	0.65	0.61	0.56	0.51	0.45	0.39	0.34	0.28	0.22	0.17	0.12	0.07	0.03	0	-0	-0.1	-0.1	-0.1	-0.1
0.1	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.05	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV

λ_1 \ Q	1.05	1.1	1.15	1.2	1.25	1.3	1.35	1.4	1.45	1.5	1.55	1.6	1.65	1.7	1.75	1.8	1.85	1.9	1.95	2
1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.95	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.9	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.85	-4.E+3	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.8	-26.2	-358	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.75	-2.99	4.78	-9.6	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.7	-0.13	6.94	21.5	50.8	7.37	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.65	0.37	4.88	13.1	28.3	57.7	118	128	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.6	0.4	3.33	8.12	16	29.2	52	93.7	178	342	RS	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1
0.55	0.34	2.28	5.25	9.77	16.6	27.1	43.4	69.8	115	197	365	626	RS	RS	RS	RS	R1	R1	R1	R1
0.5	0.26	1.58	3.49	6.25	10.2	15.7	23.6	35	51.7	76.6	115	178	289	505	967	RS	RS	RS	RS	R1
0.45	0.19	1.09	2.35	4.11	6.49	9.69	14	19.8	27.6	38.3	52.8	73.1	102	144	208	310	485	826	1554	RS
0.4	0.13	0.74	1.59	2.73	4.21	6.15	8.66	11.9	16	21.3	28.2	36.9	48.3	63.2	82.8	109	145	194	265	371
0.35	0.08	0.5	1.06	1.79	2.74	3.94	5.46	7.33	9.66	12.6	16.1	20.5	25.8	32.4	40.6	50.7	63.3	79.1	98.9	124
0.3	0.04	0.31	0.68	1.15	1.75	2.49	3.41	4.52	5.87	7.5	9.45	11.8	14.5	17.8	21.7	26.3	31.8	38.3	46	55.2
0.25	0.01	0.18	0.41	0.71	1.07	1.52	2.06	2.71	3.49	4.41	5.49	6.75	8.22	9.93	11.9	14.2	16.8	19.8	23.2	27.2
0.2	-0.01	0.09	0.21	0.39	0.59	0.85	1.15	1.52	1.95	2.46	3.03	3.7	4.47	5.35	6.35	7.47	8.74	10.2	11.8	13.6
0.15	-0.02	0.02	0.08	0.16	0.27	0.39	0.55	0.73	0.94	1.19	1.47	1.81	2.17	2.59	3.06	3.59	4.17	4.83	5.55	6.34
0.1	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.05	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV

Figure 5-19. Wait time differences (AP-FCFS)*100 when E(U)=0.01*E(S) and $\lambda_2 = 0.10$.

λ_1 \ Q	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1
1	0.92	0.73	0.45	0.05	-0.5	-1.2	-2.2	-3.5	-5.4	-7.9	-12	-17	-26	-43	-80	-190	-937	RS	RS	R1
0.95	0.93	0.74	0.47	0.1	-0.4	-1	-1.9	-3	-4.6	-6.6	-9.5	-14	-20	-30	-50	-93	-229	-1.E+3	RS	RS
0.9	0.92	0.75	0.51	0.17	-0.3	-0.9	-1.6	-2.6	-3.9	-5.6	-7.8	-11	-15	-22	-33	-54	-101	-248	-1.E+3	RS
0.85	0.92	0.76	0.53	0.21	-0.2	-0.7	-1.4	-2.2	-3.3	-4.7	-6.5	-8.8	-12	-17	-24	-35	-56	-101	-237	-1.E+3
0.8	0.93	0.77	0.54	0.26	-0.1	-0.6	-1.2	-1.9	-2.8	-3.9	-5.4	-7.2	-9.5	-13	-17	-24	-34	-53.1	-92.4	-200
0.75	0.92	0.77	0.57	0.3	-0	-0.5	-1	-1.6	-2.4	-3.3	-4.4	-5.8	-7.6	-9.9	-13	-17	-23	-31.8	-47.3	-77.7
0.7	0.92	0.78	0.59	0.35	0.04	-0.3	-0.8	-1.3	-2	-2.7	-3.6	-4.7	-6.1	-7.7	-9.7	-12	-16	-20.7	-27.8	-39.4
0.65	0.91	0.78	0.6	0.38	0.1	-0.2	-0.6	-1.1	-1.6	-2.3	-3	-3.9	-4.9	-6.1	-7.5	-9.2	-11	-14.1	-17.8	-22.9
0.6	0.91	0.78	0.62	0.41	0.17	-0.1	-0.5	-0.9	-1.3	-1.8	-2.4	-3.1	-3.9	-4.8	-5.8	-6.9	-8.3	-9.93	-11.9	-14.5
0.55	0.9	0.78	0.64	0.45	0.23	-0	-0.3	-0.7	-1.1	-1.5	-2	-2.5	-3.1	-3.8	-4.5	-5.3	-6.2	-7.17	-8.3	-9.62
0.5	0.89	0.78	0.64	0.48	0.28	0.05	-0.2	-0.5	-0.8	-1.2	-1.6	-2	-2.5	-3	-3.5	-4	-4.6	-5.26	-5.93	-6.62
0.45	0.89	0.78	0.66	0.5	0.33	0.12	-0.1	-0.4	-0.6	-0.9	-1.2	-1.6	-1.9	-2.3	-2.7	-3.1	-3.5	-3.88	-4.29	-4.67
0.4	0.87	0.77	0.66	0.52	0.36	0.19	-0	-0.2	-0.4	-0.7	-0.9	-1.2	-1.5	-1.8	-2.1	-2.3	-2.6	-2.88	-3.12	-3.34
0.35	0.85	0.76	0.65	0.53	0.39	0.25	0.08	-0.1	-0.3	-0.5	-0.7	-0.9	-1.1	-1.3	-1.6	-1.8	-2	-2.13	-2.28	-2.41
0.3	0.82	0.74	0.64	0.54	0.42	0.3	0.15	0	-0.1	-0.3	-0.5	-0.6	-0.8	-1	-1.1	-1.3	-1.4	-1.55	-1.65	-1.74
0.25	0.79	0.71	0.63	0.54	0.44	0.33	0.21	0.1	-0	-0.2	-0.3	-0.4	-0.6	-0.7	-0.8	-0.9	-1	-1.11	-1.18	-1.24
0.2	0.74	0.67	0.6	0.53	0.44	0.35	0.26	0.17	0.07	-0	-0.1	-0.2	-0.3	-0.4	-0.5	-0.6	-0.7	-0.76	-0.82	-0.86
0.15	0.67	0.62	0.56	0.49	0.42	0.36	0.28	0.21	0.14	0.06	-0	-0.1	-0.2	-0.2	-0.3	-0.4	-0.4	-0.49	-0.54	-0.57
0.1	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.05	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV

λ_1 \ Q	1.05	1.1	1.15	1.2	1.25	1.3	1.35	1.4	1.45	1.5	1.55	1.6	1.65	1.7	1.75	1.8	1.85	1.9	1.95	2
1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.95	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.9	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.85	AP	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.8	-706	-8.E+4	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.75	-152	-412	-3.E+3	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.7	-60	-106	-229	-779	-2.E+4	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.65	-31	-43.9	-68.8	-125	-287	-1.E+3	-5.E+4	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.6	-18	-22.6	-29.9	-42	-66	-119	-266	-878	-1.E+4	RS	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1
0.55	-11	-13.1	-15.6	-19	-24.5	-33.6	-50.6	-87	-177	-465	-2.E+3	AP	RS	RS	RS	RS	R1	R1	R1	R1
0.5	-7.4	-8.18	-9.07	-10	-11.4	-13.2	-15.9	-20	-28.7	-46	-83.4	-179	-479	-2.E+3	-7.E+4	RS	RS	RS	RS	R1
0.45	-5	-5.36	-5.66	-5.9	-6.06	-6.18	-6.25	-6.3	-6.55	-7.2	-8.87	-13	-22	-44.5	-98.7	-245	-731	-3.E+3	AP	RS
0.4	-3.5	-3.65	-3.7	-3.7	-3.55	-3.28	-2.86	-2.3	-1.42	-0.3	1	2.61	4.4	6.2	7.56	7.48	3.72	-8.82	-43	-132
0.35	-2.5	-2.53	-2.52	-2.4	-2.23	-1.93	-1.5	-0.9	-0.14	0.87	2.13	3.72	5.66	8.03	10.88	14.3	18.2	22.79	27.9	33.2
0.3	-1.8	-1.79	-1.75	-1.7	-1.49	-1.25	-0.92	-0.5	0.09	0.8	1.67	2.76	4.07	5.64	7.52	9.76	12.4	15.52	19.2	23.5
0.25	-1.3	-1.27	-1.24	-1.2	-1.06	-0.89	-0.66	-0.4	-0.01	0.46	1.01	1.68	2.48	3.42	4.52	5.81	7.3	9.04	11	13.3
0.2	-0.9	-0.9	-0.88	-0.8	-0.78	-0.68	-0.54	-0.4	-0.15	0.11	0.43	0.81	1.26	1.79	2.4	3.09	3.89	4.8	5.82	6.99
0.15	-0.6	-0.62	-0.61	-0.6	-0.58	-0.54	-0.47	-0.4	-0.28	-0.1	0.01	0.2	0.43	0.69	0.99	1.33	1.72	2.15	2.64	3.18
0.1	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.05	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV

Figure 5-20. Wait time differences (AP-FCFS)*100 when E(U)=0.1*E(S) and $\lambda_2 = 0.10$.

λ_1 \ Q	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
1	-3.8	-5.4	-7.5	-11	-15	-21	-29	-42	-62	-95	-153	-267	-543	-2.E+3	AP	AP	AP	RS	RS
0.95	-3.7	-5.1	-7.2	-9.9	-14	-19	-27	-37	-54	-79	-122	-198	-355	-765	-3.E+3	AP	AP	AP	RS
0.9	-3.5	-4.9	-6.8	-9.3	-13	-17	-24	-33	-47	-67	-99	-152	-249	-453	-1.E+3	-5.E+3	AP	AP	AP
0.85	-3.4	-4.7	-6.5	-8.8	-12	-16	-22	-30	-41	-57	-81	-119	-183	-300	-550	-1.E+3	-8.E+3	AP	AP
0.8	-3.2	-4.5	-6.1	-8.2	-11	-15	-20	-26	-35	-48	-67	-95	-139	-211	-344	-626	-1.E+3	-1.E+4	AP
0.75	-3.1	-4.2	-5.7	-7.7	-10	-13	-18	-23	-31	-41	-56	-77	-108	-155	-233	-372	-660	-1.E+3	-7.E+3
0.7	-2.9	-4	-5.4	-7.1	-9.4	-12	-16	-21	-27	-36	-47	-63	-85	-117	-166	-245	-380	-645	-1.E+3
0.65	-2.7	-3.8	-5	-6.6	-8.6	-11	-14	-18	-24	-31	-40	-52	-68	-90.5	-123	-170	-244	-366	-588
0.6	-2.6	-3.5	-4.7	-6.1	-7.9	-10	-13	-16	-21	-26	-33	-43	-55	-70.9	-92.9	-123	-167	-232	-334
0.55	-2.4	-3.2	-4.3	-5.6	-7.2	-9.1	-11	-14	-18	-23	-28	-35	-44	-56.1	-71.4	-91.7	-119	-157	-211
0.5	-2.2	-3	-3.9	-5.1	-6.5	-8.1	-10	-13	-16	-19	-24	-29	-36	-44.7	-55.5	-69.3	-87.2	-111	-142
0.45	-2	-2.7	-3.5	-4.5	-5.8	-7.2	-8.9	-11	-13	-16	-20	-24	-29	-35.7	-43.4	-53	-64.9	-79.8	-98.9
0.4	-1.7	-2.4	-3.1	-4	-5.1	-6.3	-7.8	-9.5	-11	-14	-17	-20	-24	-28.5	-34	-40.7	-48.8	-58.6	-70.5
0.35	-1.5	-2	-2.7	-3.5	-4.4	-5.5	-6.7	-8.1	-9.7	-12	-14	-16	-19	-22.6	-26.6	-31.2	-36.8	-43.3	-51
0.3	-1.2	-1.7	-2.3	-3	-3.7	-4.6	-5.6	-6.7	-8	-9.5	-11	-13	-15	-17.7	-20.5	-23.8	-27.6	-31.9	-36.9
0.25	-0.9	-1.4	-1.8	-2.4	-3	-3.7	-4.5	-5.4	-6.4	-7.6	-8.8	-10	-12	-13.6	-15.6	-17.8	-20.4	-23.2	-26.4
0.2	-0.6	-1	-1.4	-1.8	-2.3	-2.9	-3.5	-4.2	-4.9	-5.8	-6.7	-7.7	-8.8	-10.1	-11.4	-12.9	-14.6	-16.4	-18.4
0.15	-0.3	-0.6	-0.9	-1.2	-1.6	-2	-2.5	-3	-3.5	-4.1	-4.7	-5.4	-6.1	-6.93	-7.81	-8.76	-9.79	-10.9	-12.1
0.1	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.05	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV

λ_1 \ Q	1	1.05	1.1	1.15	1.2	1.25	1.3	1.35	1.4	1.45	1.5	1.55	1.6	1.65	1.7	1.75	1.8	1.85	1.9	1.95	2
1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.95	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.9	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.85	AP	AP	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.8	AP	AP	AP	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.75	AP	AP	AP	AP	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.7	-4.E+3	AP	AP	AP	AP	AP	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.65	-1.E+3	-3.E+3	-4.E+4	AP	AP	AP	AP	AP	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1	R1	R1	R1
0.6	-506	-831	-2.E+3	-5.E+3	AP	AP	AP	AP	AP	AP	RS	RS	RS	RS	R1	R1	R1	R1	R1	R1	R1
0.55	-291	-416	-626	-1.E+3	-2.E+3	-6.E+3	AP	AP	AP	AP	AP	AP	AP	RS	RS	RS	RS	R1	R1	R1	R1
0.5	-184	-244	-330	-462	-675	-1.E+3	-2.E+3	-5.E+3	AP	AP	AP	AP	AP	AP	AP	AP	RS	RS	RS	RS	R1
0.45	-123	-155	-198	-256	-337	-454	-634	-930	-1.E+3	-3.E+3	-7.E+3	AP	AP	AP	AP	AP	AP	AP	AP	AP	RS
0.4	-85.3	-104	-127	-156	-194	-243	-309	-400	-528	-719	-1.E+3	-2.E+3	-3.E+3	-6.E+3	-5.E+4	AP	AP	AP	AP	AP	AP
0.35	-60.1	-71.1	-84.3	-100	-120	-144	-174	-211	-259	-320	-401	-511	-663	-884	-1.E+3	-2.E+3	-3.E+3	-6.E+3	-2.E+4	AP	AP
0.3	-42.7	-49.4	-57.2	-66.3	-77	-89.6	-105	-122	-144	-169	-200	-238	-285	-344	-418	-515	-641	-813	-1.E+3	-1.E+3	-2.E+3
0.25	-30.1	-34.2	-38.8	-44.2	-50.2	-57.1	-64.9	-73.9	-84.3	-96.2	-110	-126	-145	-166	-192	-222	-258	-300	-352	-415	-492
0.2	-20.7	-23.2	-25.9	-29	-32.4	-36.2	-40.4	-45	-50.3	-56.1	-62.6	-69.8	-78	-87.1	-97.4	-109	-122	-137	-154	-173	-194
0.15	-13.5	-14.9	-16.5	-18.1	-20	-22	-24.2	-26.6	-29.1	-32	-35	-38.4	-42	-46	-50.3	-55.1	-60.2	-65.9	-72.1	-78.8	-86.2
0.1	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV
0.05	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV	LV

Figure 5-21. Wait time differences (AP-FCFS)*100 when E(U)=1.0*E(S) and $\lambda_2 = 0.10$.

With the addition of setup, we make a number of observations:

1. There is an area where AP provides enough savings in the setups realized to increase the capacity of the system, relative to what the FCFS can achieve. In fact the region is stable when using AP, but unstable for FCFS. Recalling from Chapter 2, the FCFS system utilization is $\lambda E(P)$ where the processing time, $P = U + S$, includes the setup time and therefore is greater than $\rho = \lambda E(S)$ when the setup, U, is non-zero. The AP rule self-regulates the incidence of setup: in high traffic the queue is longer so there is less likelihood of switchover at the end of a service and in the limit when $\rho = 1$ there is zero probability of switchover at the end of a service. Thus, the AP system saturation is a function only of $\rho = \lambda E(S)$, regardless of the setup magnitude.
2. AP always requires fewer changeovers than FCFS. The FCFS probability is invariant to Q since the probability of a random arrival requiring a changeover is a function of only λ_1 and λ_2. The AP probabilities monotonically decrease with increasing Q at any λ_1 (increasing utilization) and approach zero at saturation. This fact will lead to an increase in system capacity for AP when compared with FCFS when setup is non-zero.
3. For the symmetric case where $\lambda_1 = \lambda_2$, $E(S_1) = E(S_2)$, and $E(U_1) = E(U_2)$ AP always requires less wait than FCFS ($U_i > 0.0$). We also note that at any $\lambda_1 = \lambda_2$ the AP wait is monotonically better than FCFS with both increasing setup and service.

4. There is always an area where AP provides less wait than FCFS. The region where AP wait is less than FCFS is much larger when setup is present. When the setup equals the service magnitude, AP dominates the entire feasible space. This may also suggest that $Q = 1$ has much less significance with non-zero setup.

Chapter 6

SUMMARY AND DIRECTIONS OF FUTURE RESEARCH

In this dissertation, we addressed two questions concerning the role of setup economies in discrete parts manufacturing. First, using simulation as the tool of choice, we design and conduct a computational study to evaluate the impact of setup reduction on the factory flow time in the setting of factory conversion from a job shop to full or partial cellular layout. A key component of the design is the construction of a framework for experimentation and a standardized test bed of scenarios with sufficient uniformity as to make meaningful comparisons possible. In the second segment of the dissertation, we focus on a queueing system that is simple enough so that the exact analysis of the extent of setup incidence and economies can be computed exactly. We use the results of analytic models of this system to gain insights into the role of the dispatching rule in the queueing system.

We now re-state the research questions in Chapter 1 and summarize the findings of Chapters 3-5 in the form of responses to these questions.

Factory Conversions to Cellular Manufacturing Systems

- Can consistent results be obtained as to when the conversion of the job shop can be expected to prove advantageous?

- What are the measured setup economies? When are setup economies large enough to overcome pooling losses?

The conversion of job shops to cells consistently improves flow time by 10% to 20%, for the test bed used in our research. This result provides a conservative estimate of the advantages of CMS because it does not take advantage of such additional factors as reduced batch sizes, transfers batches, or move times. We find that conversion to cells consistently reduces setup on the order of 65% to 75% for the test bed we used. We conclude that setup reduction can overcome the effects of pooling loss as long as the magnitude of the setups is not too small and no significant bottlenecks develop in the cells upon conversion.

- How do other cell factors including reduced batch sizes and the use of transfer batches affect flow times achieved in cells?

The use of reduced batch sizes, or the implementation of transfer batches, can each provide cells with an additional improvement in flow time. Typically, each of these two factors has a significant effect on reducing the flow time for CMS, and the amount of reduction is usually at least as large as that obtained by conversion to CMS without any changes in the batch sizes.

Our sensitivity runs show that the overall factory utilization and the potential for setup reduction can both affect the conversion results obtained. Our tests indicate that conversion to CMS may not be advantageous if the utilization level is high or there is not sufficient potential to reduce setups.

The design of cells also has a clear impact on the conversion improvements obtained. Typically, we observed better performance in cells when the original source design was used. However, conversion benefits continue to be present even after we use a uniform cell formation procedure due to Vakharia and Wemmerlöv (1990). This indicates that careful allocation of machines to cells to avoid heavy utilization helps to keep the pooling loss within tight control.

Regarding dispatching rules, our experimental runs support the conclusions of previous authors that Repetitive Lot dispatching provides less overall setup and supports lower flow times than FCFS in a job shop with setup. The effect of RL seems to diminish in the same factory setting once it incorporates cells.

- Is there value in considering a partial implementation of CMS?

Although we could not identify general and robust rules that applied across all data sets, we observed that the factory flow time of a partial cellular implementation may be as good as or even better than the all-cell option, so it is important to stop short of full conversion where appropriate. In addition, other considerations include the following. Even when the number of cells to be included in the partial layout is fixed, the choice of the correct subset of cells can have a significant impact on the flow time. The sequence of best subsets to pick as n increases from 1 to NC is not necessarily nested, so sequence matters. The differences in factory flow times are due to the same factors recognized in the all-cell CMS analysis, setup reduction and machine utilization, but neither factor alone is sufficient to reliably determine the best subset of cells to select. The best picks are characterized by large setup reductions along with reduction of utilization in the residual job shop and the lack of bottlenecks in the cell(s), so setup reduction and load balance in both the cells and residual job shop matter.

To our knowledge, this is the first simulation study that compares cell shop conversion benefits across disparate data sets. We believe that this dissertation has shown that the comparison of job shops and cellular systems with respect to the flow time measure can produce reasonably consistent results when the same operating rules and key parameter ranges are used across different data sets. Moreover, our research shows that setup reduction can overcome pooling losses, even under the conservative assumptions where batch size remain unchanged and the material transport times in the job shop are assumed to be negligible. Overall, the conclusions of our research are consistent with the qualitative insights cited in the literature when comparing cell shops and job shops. However, our research clarifies that the quantitative comparisons using the flow time metric must be interpreted in the context of the region of the parameter space spanned by the data sets, as well as the particular design used for the cells.

By investigating the efficacy of implementing partial cell layouts (hybrids) using the same test bed, we are able to define considerations for the cell implementation process. We find the selection of the subset of cells picked at any level of cellular implementation has an impact on factory flow time and that a partial cellular implementation may be as good as or even better than the all-cell option.

Analytic Modeling of a Simple System with Setup

- What is the role of dispatching rules in the reduction of setups?

We find that the Alternating Priority (AP) dispatching rule that minimizes setup incidence, and therefore, changeover incidence, can outperform the FCFS rule over significant regions of the two-class parameter space even when the setup time is taken to be zero (the metric for this comparison is average wait time in queue). We characterize the region of superiority of AP over FCFS analytically and provide bounds on the relative performance of the two rules.

When setup enters the comparison between these rules, we determine the extent of the difference in setup paid as well as the difference in setup incidence between AP and FCFS. We are able to identify regions where AP is always the better choice as well as regions where AP increases the service capacity due to reduction in the setups incurred. For the symmetric case of non-zero setup where $\lambda_1 = \lambda_2$, $E(S_1) = E(S_2)$, and $E(U_1) = E(U_2)$ AP always requires less wait than FCFS. We also note that at any $\lambda_1 = \lambda_2$ the AP wait is monotonically better than FCFS with both increasing setup and service. For the non-symmetric case we also note that by the time the setup is equal to the service in magnitude, AP dominates the entire feasible region. This may also suggest that $Q = 1$ has much less significance with non-zero setup.

Directions for Future Research

The following topics are offered as potentially fruitful areas of research that would extend the findings of this dissertation.

1. Analytic comparison of rules in the presence of non-zero setups: In the case of non-zero setups, further research should pursue the derivation of analytic results that constitute a counterpart to the analysis of Section 5.1. We think there is opportunity to examine regions of dominance for the AP rule using formulas for non-zero setup. This would also help explain the behavior of FCFS and how it can dominate AP even in the presence of setup.
2. Extension from two classes of customers to multiple classes. This research would extend the results of Sections 5.1 and 5.2 to the multi-class case. Analytically, this requires extending the results of Eisenberg (1967) to the multi-class case. While the mathematics of following Eisenberg's specific approach becomes extremely cumbersome, simpler schemes of analysis or approximate results may still reveal useful insights. Naturally, simulation remains open as a tool for performance evaluation for all such extensions.
3. Alternative rules for multiple customer classes: A quick search of the literature reveals that the analysis of queues with multiple classes in the presence of setups has let to a stream of research involving cyclic polling rules (where customers are serviced in a pre-determined order). Such

rules may be viewed as alternatives to extensions of the AP rule to the multi-class case (greater than two classes) such as the Repetitive Lot rule (Jacobs and Bragg, 1988) or its variants discussed in this dissertation. Further study is needed to evaluate such extensions. In particular, cyclic policies can be compared to dynamic policies that incorporate dynamic information into the switching decision. Of special interest is how setup impacts the comparative advantages of these policies.

4. Discount factors to reflect setup economies. Some studies use flat-rate discounts coupled with FCFS in analytic models to represent the effects of setup economies in job shops and cell shops. Further research is required to explore where this approximation can introduce severe distortion, especially as magnified by bottlenecks or increased congestion in the system.

APPENDIX A: SENSITIVITY TO THE SHAPE OF PROCESSING TIME DISTRIBUTIONS

The runs presented in the body of this research use a 2-Erlang distribution for both setup time and run time. The CV for this distribution is 0.707. To test the sensitivity of the flow time results to the shape of these distributions, we varied the CV while staying in the k-Erlang family and retaining the same mean. Of course, CV=1.0 corresponds to an exponential distribution (k=1) and CV=0.25 (k=16) captures the shape the normal curve. We also tested the effect of skewness by comparing the 2-Erlang with distributions from the beta(α_1, α_2) family, each skewed in a different direction.

Below in Table A-1 we tabulated the results of these runs for two data sets. Each cell with a dual entry shows the flow time for the job shop on top and CMS directly below it. Although the shape of the distribution affects both the job shop and CMS flow times, these values move together so that the flow ratio remains insensitive to the changes.

Data Set #2		CV Run		
	JS flow CMS flow	0.250	0.707	1.000
CV Setup	0.250	148 127	148 127	148 130
CV Setup	0.707	149 128	149 130	149 130
CV Setup	1.000	150 129	150 130	151 131

Data Set #6		CV Run		
	JS flow CMS flow	0.250	0.707	1.000
CV Setup	0.250	7582 6163	7586 6208	7595 6253
CV Setup	0.707	7613 6225	7612 6261	7634 6303
CV Setup	1.000	7657 6267	7644 6309	7659 6343

	Data Set #2	Data Set #6
2-Erlang CV=0.707	149 130	7612 6261
Beta(1.4,5.5) CV=0.705	148 127	7581 6148
Beta(5.5,1.4) CV=0.180	149 129	7607 6263

Table A-1. Sensitivity of Job Shop and CMS flow times to changes in distributions of setup and runtime.

APPENDIX B: OUTPUT MEASURES FOR SIMULATION RUNS

Our additional input parameters for Chapters 3 and 4 are as follows.

$T =$ duration of simulation window for releasing batches

$P =$ number of batch orders released during simulation release window, T

The output statistics gathered by the simulation are as follows.

$TQ =$ time at which last of P released batches is completed (simulation horizon)

$FT(p) =$ flow time of the p^{th} batch released within release window, T

$(p = 1, \ldots, P)$ [flow time measured from order release to shipping]

$ST(p) =$ total setup incurred for the production of the p^{th} batch $(p = 1, \ldots, P)$

$RT(p) =$ total run time incurred for the production of the p^{th} batch $(p = 1, \ldots, P)$

$SQ(j) =$ total setup time accrued on machine type j during TQ

$RQ(j) =$ total run time accrued on machine type j during TQ

The output measures are then calculated as follows.

The average batch flow time is

$$\sum_{p=1}^{P} FT(p) \Big/ P \tag{B-1}$$

Average time a batch spent in setup

$$\sum_{p=1}^{P} ST(p) \Big/ P \tag{B-2}$$

Average time a batch spent being run

$$\sum_{p=1}^{P} RT(p) \Big/ P \tag{B-3}$$

Average machine utilization for type j

$$\left(SQ(j) + RQ(j)\right)/TQ \cdot NM(j) \tag{B-4}$$

Overall average machine utilization for the factory (JS or CMS)

$$\sum_{j=1}^{J} \left(SQ(j) + RQ(j)\right) \Big/ TQ \sum_{j=1}^{J} NM(j) \tag{B-5}$$

Maximum machine utilization for the JS configuration

$$\max_{j} \left[\left(SQ(j) + RQ(j)\right)/TQ \cdot NM(j)\right] \tag{B-6}$$

The minimum calculations are analogous. For the CMS, the maximum and minimum utilization values consider machine types over all cells, so that equation (B-6) is computed once for each cell.

GLOSSARY

Alternating Priority	a dispatching rule from Maxwell (1961) designed to minimize setup incidence in a single-server queue with two customer classes: all jobs in queue of a given class are served before switching to the other class. The server thus alternates between strings of jobs of either class 1 or class 2 and the idle state, but never switches from class i to class j $(j \neq i)$ if there are jobs of class i still in queue
Cell	a collection of different machines positioned in proximity to work on a family of parts with similar shapes and processing requirements
Cellular Manufacturing	manufacturing part families using cells
Flow Ratio	ratio of the average batch flow time after cellular conversion to the average batch flow time of the job shop with the same factory operational parameters of load, machines and batch size

Job Shop	a manufacturing facility comprised of general-purpose machines organized into a collection of machine centers (departments) grouped on the basis of the operation performed
Major-Minor Model for Setup	a setup structure whereby the setup is a major setup, a minor setup, or no setup at all. A major setup is incurred if two parts belonging to distinct families are processed consecutively on the same machine. Switching between two different part types in the same family incurs a minor setup. No setup is required if a machine processes two batches of the same part type consecutively
Part Family	parts with similar features and common sequences of operations requiring similar tools or fixtures
Pooling Loss	the diseconomies of segregating a given machine type by assigning them to independent cells
Remainder Shop	that part of the factory that is not converted to cells and continues to operate as a job shop
Repetitive Lot Dispatching	a dispatching rule from Jacobs and Bragg (1988) designed to minimize setup: (1) a single (pooled)

	queue is formed for all batches arriving to be processed at a machine center, (2) Any arriving batch encountering an available machine upon entry is immediately routed to the available machine where it would encounter the least setup time. If no machines are available, the batch joins (or forms) a queue to wait for a machine, (3)When a machine becomes available, the next job assigned to it is selected based on the minimum setup among all jobs in queue. If multiple jobs tie at this minimum setup value, the FCFS discipline is used to break the tie.
Setup Fraction	the ratio of minor to major setup
Setup Ratio	the ratio of major setup to batch run time
Transfer Batch	lot quantities moved between workstations or production areas – typically equal to or smaller than the production lot size

BIBLIOGRAPHY

Avi-Itzhak, B, Maxwell, W. L. and Miller L. W., “Queueing with alternating priorities,” *Operations Research*, v13, 306-318 (1965).

Burgess, A. G., “Evaluation of group technology/cellular manufacturing: a computer simulation.” Diss. Boston Univ. (1989).

Burgess, A. G., Morgan, I. and Vollmann, T. E., “Cellular manufacturing: its impact on the total factory.” *International Journal of Production Research*, v31, n9, 2059-2077 (1993).

Buzacott, J. A., Shanthikumar, J. G, Stochastic Models of Manufacturing Systems. Prentice Hall, New Jersey (1993).

Chase, R. B., Jacobs, F. R. and Aquilano, N. J., Operations Management for Competitive Advantage. Tenth Edition, Irwin, Chicago (2004).

Cohen, M. and Apte, U. M., Manufacturing Automation. Irwin, Chicago (1997).

Cooper, R., Niu, S., Srinivasan, M., “Setups in polling models: does it make sense to set up if no work is waiting?” *Journal of Applied Probability*, v36, 585-592 (1999).

Eisenberg, M., “Queues with periodic service and changeover time,” Operations Research, v20, 440-451 (1972).

Eisenberg, M., “Two queues with changeover times,” *Operations Research*, v19, 386-401 (1971).

Eisenberg, M., “Multi-Queues with changeover times,” Diss. Massachusetts Institute of Technology (1967).

Garza, O. and Smunt, T. L., “Countering the negative impact of intercell flow in cellular manufacturing”, *Journal of Operations Management*, v10, n1, 92-118 (1991).

Gaver, D. P., Jr., “A comparison of queue disciplines when service orientation times occur,” *Naval Research Logistics Quarterly*, v10, 219-235 (1963).

Gravel, M., Price, W. and Gange, C., "An interactive tool for designing manufacturing cells for an assembly job-shop," *International Journal of Production Research*, v38, 2, 309-322 (2000).

Henriksen, J. O. and Crain, R. C., GPSS/H Reference Manual, Third Edition, Wolverine Software Corporation, Annandale, Virginia (1989).

Hyer, J. and Wemmerlöv, U., Reorganizing the Factory: Competing through Cellular Manufacturing, Productivity Press, Portland, Oregon (2002).

Pegden, C. D., Shannon, R. E., and Sadowski, R. P., Introduction to Simulation Using SIMAN, McGraw-Hill, Inc., New York (1990).

Jacobs, F. R. and Bragg, D. J., "Repetitive lots: flow-time reductions through sequencing and dynamic batch sizing," *Decision Sciences*, v19, n2, 281-284 (1988).

Jensen, J. B., Malhotra, M. K. and Philipoom, P. R., "Machine dedication and process flexibility in a group technology environment", *Journal of Operations Management*, v14, n1, 19-39 (1996).

Johnson, D. J., and Wemmerlöv, U., "On the relative performance of functional and cellular layouts – an analysis of the model-based comparative studies literature." *Production and Operations Management*, v5, n4, 309-334 (1996).

Kher, H. V. and Jensen, J. B., "Shop performance implications of using cells, partial cells, and remainder cells," *Decision Sciences*, v33, n2, 161-190 (2002).

Law, A. M. and Kelton, W. K., Simulation Modeling and Analysis, Second Edition, New York: McGraw-Hill (1991).

Marsh, R. F., Shafer, S. M. and Meredith, J. R., "A comparison of cellular manufacturing research presumptions with practice," *International Journal of Production Research*, v37, n14, 3119-3138 (1999).

Maxwell, W. L. "An investigation of multi-product, single-machine scheduling and inventory problems," Diss. Cornell University (1961).

Miller, L. W. "Alternating priorities in multi-class queues," Diss. Cornell University (1964).

Morris, J. S., "Forming cellular layouts in group technology: A simulation evaluation of alternative approaches", Diss. Univ. of Oklahoma (1988).

Morris, J. S. and Tersine, R. J., "A simulation analysis of factors influencing the attractiveness of group technology cellular layouts." *Management Science*, v36, n12, 1567-1578 (1990).

Morris, J. S. and Tersine, R. J., "A comparison of cell loading practices in group technology." *Journal of Manufacturing Operations Management*, v2, 299-313 (1989).

Olorunniwo, F. O., "A framework for measuring success of cellular manufacturing implementation." *International Journal of Production Research*, v35, n11, 3043-3061 (1997).

Pegden, C. D., Shannon, R. E. and Sadowski, R. P., Introduction to Simulation Using SIMAN, McGraw Hill, New York (1990).

Pine II, B. J. Mass Customization: The New Frontier in Business Competition. Harvard Business School Press, Boston (1993).

Rehman, A. and Diehl, M. B. "Rapid modeling helps focus setup reduction at Ingersoll." *Industrial Engineering*, November, 52-55 (1993).

Schriber, T. J., Simulation Using GPSS: John Wiley and Sons, New York (1974).

Seifoddini, H. and Djassemi, M., "Determination of a flexibility range for cellular manufacturing systems under product mix variations." *International Journal of Production Research*, v35, n12, 3349-3366 (1997).

Shafer S. M., and Meredith, J. R., "An empirically-based simulation study of functional versus cellular layouts with operations overlapping." *International Journal of Operations and Production Management*, v13, n2, 47-62 (1993).

Shafer, S. M. and Charnes, J. M., "Offsetting lower routeing flexibility in cellular manufacturing due to machine dedication." *International Journal of Production Research*, v35, n2, 551-657 (1997).

Shafer, S. M. and Charnes, J. M., "A simulation analyses of factors influencing loading practices in cellular manufacturing." *International Journal of Production Research*, v33, n1, 279-290 (1995).

Shafer, S. M. and Charnes, J. M., "Cellular versus functional layouts under a variety of shop operating conditions." *Decision Sciences*, v24, n3, 665-681 (1993).

Shambu, G., “The performance of hybrid cellular manufacturing systems: A computer simulation investigation.” Diss. State University of New York at Buffalo (1993).

Shambu, G. and Suresh, N. C., “Performance of hybrid cellular manufacturing systems: A computer simulation investigation,” *European Journal of Operational Research*, v120, 436-458 (2000).

Shunk, D. L., “The measurement of the effects of group technology by simulation.” Diss. Purdue Univ. (1976).

Singh, N. and Rajamani, D., Cellular Manufacturing Systems: Design, Planning, and Control. Chapman & Hall, London (1996).

Smunt, T. L., Buss, A. H. and Kropp, D. H. “Lot splitting in stochastic flow shop and job shop environments,” *Decision Sciences*, v27, n2, 215-238 (1996).

Suresh, N. C., “Partitioning work centers for group technology: Analytical extension and shop-level simulation investigation.” *Decision Sciences*, v23, n2, 267-290 (1992).

Suresh, N. C., “Partitioning work centers for group technology: insights from an analytical model.” *Decision Sciences*, v22, n4, 772-791 (1991).

Suresh, N. C. and Meredith, J. R., “Coping with the loss of pooling synergy in cellular manufacturing systems.” *Management Science*, v40, n4, 466-483 (1994).

Sykes, J. S., “Simplified analysis of an alternating-priority queueing model with setup times,” *Operations Research*, v18, n6, 1182-1192 (1970).

Takagi, H. “Priority queues with setup times,” *Operations Research*, v38, n4, 667-677 (1990).

Vakharia, A. J. and Wemmerlöv, U., “Designing a cellular manufacturing system: a materials flow approach based on operation sequences.” *IIE Transactions*, v22, n1, 84-97 (1990).

Wemmerlöv, U. and Hyer, N. L., “Cellular manufacturing in the U.S. industry: a survey of users,” *International Journal of Production Research*, v27, n9, 1511-1530 (1989).

Wemmerlöv, U. and Johnson, D. J., “Cellular manufacturing at 46 user plants: implementation experiences and performance improvements.” *International Journal of Production Research*, v35, n1, 29-49 (1997).

Wolff, R. W., Stochastic Modeling and the Theory of Queues, Prentice Hall, New Jersey (1989).

Yang, K. K. and Jacobs, F. R., “Comparison of make-to-order jobs shops with different machine layouts and production control systems,” *International Journal of Production Research*, v30, n6, 1269-1283 (1992).