Re-examining the Performance of MRP and Kanban Material Control Strategies for Multi-product Flexible Manufacturing Systems

Ananth Krishnamurthy
Department of Decision Sciences and Engineering Systems,
Rensselaer Polytechnic Institute,
110 8th Street, Troy, NY 12180.
Phone: 518-276 2958, Fax: 518-276 8227
Email: krisha@rpi.edu

Rajan Suri
Center for Quick Response Manufacturing,
University of Wisconsin-Madison,
1513 University Avenue, Madison, WI 53706-1572.
Telephone: 608-262 0921, Fax: 608-265 4017
Email: suri@engr.wisc.edu.

and

Mary Vernon,
Department of Computer Sciences,
University of Wisconsin-Madison,
1210 W Dayton Street, Madison, WI-53706.
Telephone: 608-262 7893, Fax: 608-262 9777
Email: vernon@cs.wisc.edu

Abstract

Material control schemes can be classified as push, pull, or hybrid strategies. This paper compares the performance of MRP (push) and kanban (pull) for a multi-stage, multi-product manufacturing system. Using simulation experiments we analyze system performance under different product mixes and observe that in certain environments with advance demand information kanban-based pull strategies can lead to significant inefficiencies. In these environments MRP-type push strategies yield better performance in terms of inventories and service levels. We also study the impact of design parameters such as safety lead time and safety stock policies on system performance and observe that for low and medium values of system loads, safety lead time policies yield better system performance than safety stock policies. These insights can be helpful in designing efficient MRP-type push strategies in multi-product environments.

Keywords: kanban, MRP, material control strategies, multi-product, multi-stage, pull, push.

(Submitted to *International Journal of Flexible Manufacturing Systems*, December 2000.
Revised September 2003, June 2004)

1. Introduction

Material planning and control strategies can be classified as push, pull, or hybrid strategies (Karmarkar, 1991). Material requirements planning (MRP) systems and kanban control systems are the most popular implementations of push and pull strategies respectively. The main distinction between push and pull strategies is based on how production orders are released to work stations in response to demands. In a *push* strategy production is initiated based on estimates of future demand. It is assumed that advance demand information is available, either in the form of actual orders, or forecasts, or a combination of both. Production orders are released into the shop floor by offsetting the due date of requirements by their corresponding planned production lead times (Orlicky 1975). In contrast, in a *pull* strategy, production is initiated in response to current demand. Demands are satisfied from inventory and the removal of items from the output inventory buffers to satisfy demand triggers production upstream to replenish these inventories. Material control strategies that combine features of push and pull are referred to as *hybrid* strategies. Recently, several hybrid strategies such as CONWIP (Spearman *et al.*, 1990) and POLCA (Suri, 1998) have been proposed.

In the last decade there has been considerable interest in the analysis of material control strategies for manufacturing systems. The successful implementation of kanban systems as well as analytical studies done on single product systems have led to the belief that the performance of pull systems and its variations are generally superior (Spearman and Zazanis, 1992; Womack and Jones, 1996). The success of pull strategies at certain original equipment manufacturers (OEMs) have further led to the belief that implementing pull in all portions of the supply chain would be beneficial. However, our experience with some companies struggling to implement kanban systems leads us to believe that pull strategies are fundamentally handicapped for manufacturing facilities that produce a number of different products with distinct demands and/or processing requirements, as well as for facilities that make highly engineered products in small batches (perhaps even one of a kind) for their customers (Suri, 1998).

We explain these ideas further using a typical supply chain scenario as an example. Consider a facility that manufactures a variety of transmissions for agricultural, marine, and military equipment. Figure 1 represents the schematic of such a facility. C_1, …, C_N could be assembly cells within this facility that are dedicated to assemble products for different market segments and S could be the common fabrication cell that supplies C_1, …, C_N with different components such as housings or other sub-assemblies required for final assembly. (S could also be an external supplier.) Due to the distinct nature of the market segments served by the final assembly cells (C_1, …, C_N) and the possible variety of products demanded, the customer demand patterns for the different products and the processing requirements for the different products at the manufacturing stations in fabrication cell S can be quite different. Although the demand patterns for the different products are quite different, assembly cells (C_1, …, C_N) often fix their assembly schedules several days in advance and share this information with the fabrication cell S.

The orders from the assembly cells therefore have fixed due dates or delivery dates. An important measure of performance of fabrication cell S then is its ability to meet these delivery dates for the different products and the corresponding inventory levels/costs.

Kanban (pull) and MRP (push) are two possible choices of material control strategy that fabrication cell S might adopt to meet its objectives. By considering the fundamental nature of push and pull strategies in the following paragraphs, we argue qualitatively that pull strategies need not perform better in these environments. In the remainder of the paper we provide quantitative results to support the qualitative arguments presented below.

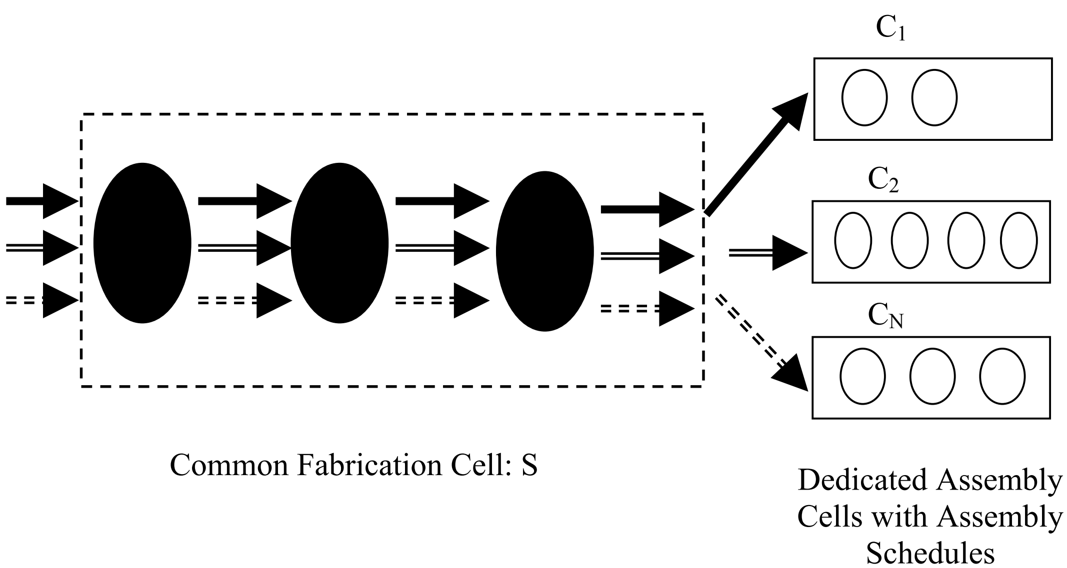


Figure 1. Motivating manufacturing example.

Since pull is essentially a *replenishment* strategy that was initially designed for manufacturing environments producing repetitive products with high demand volumes, it has some potential drawbacks for fabrication cell (or supplier S) operating in environment with multiple products. Pull requires that a minimum inventory of each product be maintained at the output of each workstation. When one unit (or container) of inventory is taken by the downstream workstation this would immediately signal the upstream workstation to begin work to replenish this quantity. If cell S is manufacturing a large number of product specifications with possibly distinct demands, this can lead to proliferation of work in progress (WIP) inventories at each stage of the process (see the example in Suri, 2000). Particularly, certain product environments could lead to situations where the time between demands for some products is greater than the average their flow times. In these situations, the pull strategy could lead to inventory replenishments well in advance of their requirements, resulting in excess WIP inventories. Ignoring this possibility and implementing pull material control strategies in such an environment could lead to inefficient system performance. Pull strategies also exercise rigid controls on production schedules by enforcing *takt* times and *level scheduling* (Womack and Jones, 1996). These involve optimizing or standardizing tasks and freezing production schedules. However, at fabrication cell S, where the products could be possibly highly customized or demands highly variable, setting *takt* times could be impractical. Frequent revision of *takt* times could be required for fabrication cell S to respond to the varied requirements of its customers efficiently. Hence, pull might not be the best material control strategy for such environments.

In a push strategy, the production triggers are based on the *release times* of jobs at each station in the system. These release times are determined by backward scheduling the due dates of orders from the assembly cells using the planned production lead times at each station (Vollmann et al., 1991). To account for the difference between the planned lead time and the actual flow time of an order, push strategies often incorporate safety stocks or safety lead times. Incorporating safety lead times involves inflating the lead time estimates, while incorporating safety stocks involves increasing the target inventory levels in the system. Although safety stocks and safety lead times are intended to improve performance it has been shown that due to the uncertainties in customer demands and errors in estimates of planned lead times, push strategies may result in excessive inventories (Hopp and Spearman, 1996). On the other hand, in an environment such as the one described for fabrication cell S, the cell might have reliable information about its customer demands and average lead times for its products. Furthermore, several manufacturers today routinely track and monitor flow times at their facilities (Ericksen and Suri, 2001). In view of this available information, facility S might benefit from adopting a push material control strategy that explicitly considers future requirements while triggering production releases. Ignoring the available information on future requirements and adopting a pure pull strategy that merely replenishes consumed inventories might prove detrimental, especially if the number or diversity of the products manufactured in cell S is quite high.

Recently a few researchers have provided qualitative arguments similar to the above. For example, see Hopp and Spearman (1996) and Suri (2000). However, there is a lack of quantitative studies that analyze the performance of material control strategies in manufacturing environments with multiple products and diverse product mixes. Therefore, the goal of this paper is to conduct quantitative comparisons of the performance of push and pull strategies by modeling fabrication cell S as a multi-stage manufacturing system producing multiple diverse products. We use simulation for the comparisons, since exact analytical models of multiple products in a general manufacturing setting operating under push and pull are not available. The simulation studies performed as part of this research are an initial step towards getting quantitative insights on the performance of push and pull in multi-product environments. In fact, the simulation models needed for the comparisons become computationally prohibitive for systems manufacturing a large number of products. However, under abstract but equitable assumptions that are necessary to make these computations manageable we construct simulation models that yield insights that can be generalized and applied to a wide range of real systems. Using these simulation models we compare the performance of MRP (push) and kanban (pull) systems for different loads and product mixes.

Our key observation is that in multi-product environments the performance of kanban (pull) deteriorates significantly. We consider several product mixes in which the products could differ in terms of their demand rates and/or service requirements. For a moderate number of products with homogeneous demand rates and service requirements and even more so for heterogeneous demand rates, we observe that push strategy outperforms pull in terms of service levels and average inventories. This is different from the performance of push and pull in environments with low product variety (Spearman and Zazanis, 1992; Buzacott and Shanthikumar, 1993; Hopp and Spearman, 1996). These performance results have important implications for companies increasing their number of product offerings to their customers as part of their corporate strategy. It might be more efficient for them to adopt push strategies in portions of the manufacturing facilities where reasonably accurate information about future requirements and production lead times are available. Our simulation studies also provide important insights into the design of pull and push systems. In particular, we observe that system performance under the pull strategy is very sensitive to design parameters such as the allocation of kanbans. Improper allocation of kanbans can result in low service levels and high inventories. On the contrary, our studies show that system performance under the push strategy is more robust to the choice of design parameters such as safety lead times and safety stocks. Our experiments also include an evaluation of whether safety lead times or safety stocks are preferable in these environments.

The comparison of push and pull strategies conducted in this research primarily focuses on the flow control aspect of MRP and kanban strategies. Several other studies have focused on other aspects of MRP and kanban control and their variations. For instance, Suri and de Treville (1986) demonstrate that although the rigid control of pull is susceptible to disruptions in production, such disruptions provide opportunities to identify bottlenecks and improve system performance. In contrast, MRP-type push strategies in the presence of imperfect demand information can lead to dysfunctional behavior (Suri, 1998). We will not consider these behavioral aspects of pull or push in this research. Further, we recognize that a comprehensive comparison of push and pull strategies would require considering manufacturing systems with different topologies (assembly systems, job shops etc.), exploring the influence of factors such as forecast inaccuracies, errors in MRP lead times, and analyzing the performance of the different variants of pull strategies proposed in the literature (Liberopoulos and Dallery, 2000). We consider these as important areas for extending the current research. Consistent in spirit with other modeling efforts including Karmarkar (1987), Buzacott et al. (1992), and more recently Karaesmen et al. (2002), our aim here is to design a simulation model simple enough to capture the basic impact of push and pull strategies on the dynamics of the manufacturing system.

The remainder of the paper is organized as follows. Section 2 provides a review of the relevant literature. Section 3 provides the system description and discusses the setup of the simulation experiments. Section 4 discusses the insights obtained from simulation on the performance of push and pull in multi-product systems with a homogeneous product mix. Section 5 provides insights from similar comparisons for multi-product systems with heterogeneous product mixes. Section 6 compares alternative designs for push strategies in multi-product environments. In particular the performance of push strategies operating under several safety stock and safety lead time policies are compared. Section 7 provides the conclusions.

2. Literature review

In this section we review the relevant literature on push and pull strategies. The review is structured as follows. We first review the literature that discusses the key modeling issues in push and pull systems. Next we review the literature on performance comparisons of push and pull systems.

2.1. Push systems

Considerable research has focused on the performance of MRP (push) strategies (see Orlicky, 1975; Vollmann et al., 1991; Buzacott and Shanthikumar, 1993; and the references therein). From these efforts we can conclude that there are three main issues in modeling push systems: (1) estimating release lead times for MRP, (2) modeling future requirements for the different products, and (3) determining the safety lead times and/or safety stocks required to guarantee the required service levels for the different products. Regarding the first issue, Buzacott and Shanthikumar (1994) and Karaesmen et al. (2002) analyze single product systems operating under MRP (push) policies under various scenarios. Both these studies assume that accurate estimates of lead times are available. With respect to determining future customer requirements, Buzacott and Shanthikumar (1994) consider two situations (a) future demands over the lead times are known exactly, and (b) only the mean demand rate is known, while Karaesmen et al. (2002) only consider the former environment where future requirements over the lead times are known exactly. For the comparisons of multi-product systems in this paper, we also focus on the first environment and assume that accurate estimates of release lead times are available. As noted in the previous section, there are many practical situations where these assumptions hold. Comparisons in other situations are left for future research.

Regarding safety stock and safety lead times, Lambrecht et al., (1984), Buzacott et al. (1992), and Buzacott and Shanthikumar (1994) report several key results for single product systems. First, they report that the service delay experienced by a customer order decreases by increasing the safety stock at any stage. Second, if there is a limit on the total safety stock, then service delays are minimized by having all the safety stock at the final stage. Third, customer service delays can be reduced by inflating the values of the planned lead times at each station. Finally, for the case wherein complete information about future customer demands is available, safety lead times are always preferable to safety stocks. Since to our knowledge, it has not been shown that these insights also hold for multi-product manufacturing systems, our simulation experiments consider several settings for safety stocks and safety lead times, rather than safety lead times alone.

2.2. Pull systems

Pull (kanban) strategies have been the subject of numerous studies by researchers (see Uzsoy and Martin-Vega, 1990; Berkley, 1992; Liberopoulos and Dallery, 1997; and the references therein). These works review the different deterministic, stochastic, and simulation models used for the analysis of kanban and generalized pull systems. From these efforts we can conclude that the main issues in modeling kanban systems are (a) determining the number of kanbans for each product and (b) their allocation among the different stages of the manufacturing system. However it appears that determining the optimal allocation of kanbans at the different stages is not easy. Tayur (1992), Muckstadt and Tayur (1995), and Gstettner and Kuhn (1996) study the impact of kanban allocation patterns on system performance and report optimal allocations for serial production systems manufacturing a single product. Similar studies for multi-product systems are not available. In addition, Buzacott and Shanthikumar (1993) qualitatively argue that the allocation of kanbans for one product could impact the performance of other products sharing the same production resource. Consequently, in our simulation experiments we consider a wide range of kanban allocations for the different products.

2.3. Comparison of push and pull systems

Spearman and Zazanis (1992) compare the performance of pull and push strategies using simple queueing models and show that for serial lines manufacturing a single product pull strategy always results in less congestion and WIP. Buzacott (1989) and Buzacottt and Shanthikumar (1993) conduct similar studies and report that, for systems operating under kanban strategies the service delay experienced by a customer order decreases with increase in the total number of kanbans in the system and for a system operating under the push strategy forecast inaccuracies could result in significant deterioration of system performance. Analytical studies that compare the performance of push and pull for multi-product systems are limited, because analytical models that explicitly model multiple products in a general manufacturing setting operating under push and pull are hard to solve exactly. Zhou et al. (2000) use simulation to compare the performance of push, pull, and hybrid push-pull strategies for manufacturing systems with multiple products. They observe that in certain environments hybrid push-pull strategies outperform both push and pull strategies. Our research also uses simulation to compare the performance of material control strategies in multi-product environments, but we examine the impact of product mix diversity on system performance, and restrict our focus to only push and pull strategies.

3. System description and setup of simulation experiments

In this section we provide additional details regarding the system being analyzed. First, we describe the system structure, and discuss the assumptions regarding customer demands and service requirements for the different products. Subsequently, we describe how the operation of push and pull strategies are modeled in the simulation experiments. We conclude this section with a discussion of the experimental design and the measures that are computed in the simulations.

3.1. System description

To conduct quantitative comparisons of the performance of push and pull strategies at a manufacturing cell (or supplier) S supplying diverse products to multiple assembly cells within an OEM facility, we model cell S as a serial production line with three stages (j=1,2,3) manufacturing 8 different products. See Figure 2.

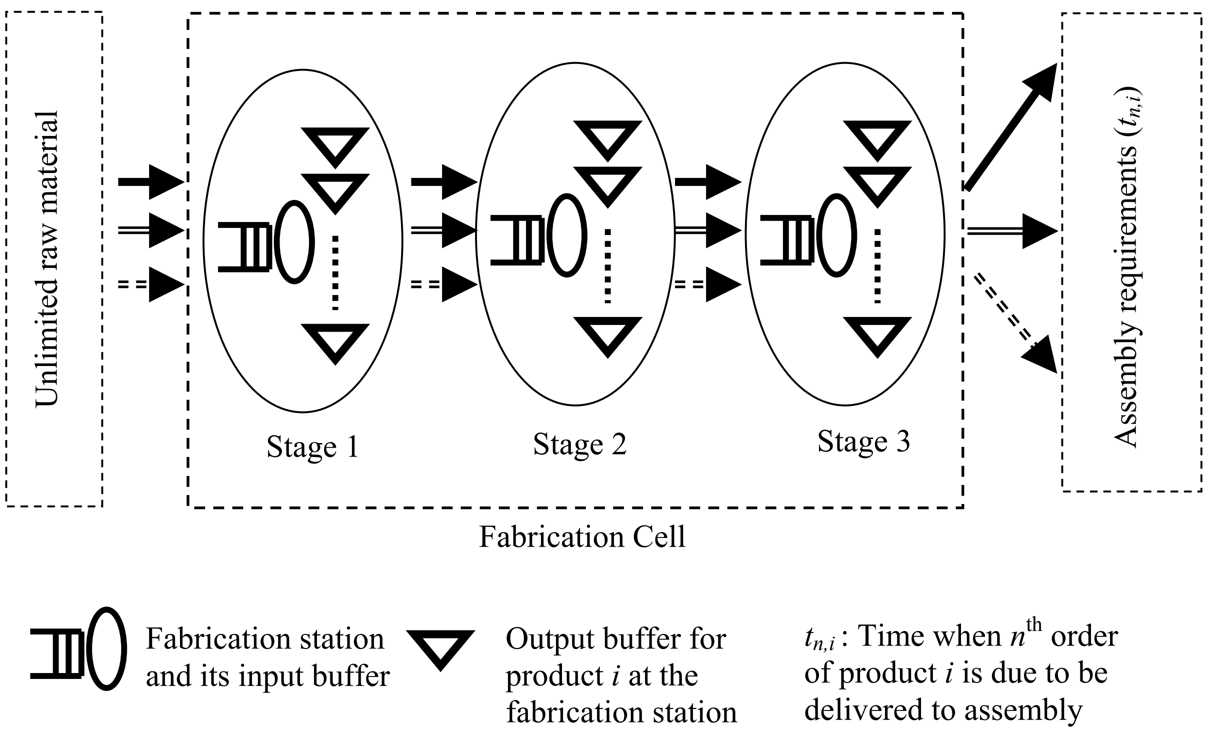


Figure 2. System description.

For simplicity we assume that each stage consists of a single machine and its input and output buffers. We consider a three stage line to keep the number of machine parameters that need to be varied in our simulations manageable while still capturing the effects of initial stage, intermediate stage, and final stage processing times on the performance of the control strategies. We assume that machine 1 always has a sufficient supply of the raw materials required for the different products. The output product of one machine becomes the input product to the subsequent machine and at the output of the last machine there is a finished goods inventory buffer that stores final products before they are delivered to the respective assembly cells C_1, …, C_N. In real systems, the number of different products being manufactured by cell S could be very large. The purpose of modeling multiple products for our simulation experiments is to analyze the impact of product diversity on the performance of material control strategies at cell S. Choosing a small number such as 2 products would be inadequate for a meaningful study while a large number of products would significantly increase the computational burden. Our preliminary experiments indicated that modeling a serial production line with 8 different products keeps the computations reasonable and yet provides the key insights.

The production activity at a machine typically consists of tasks such as fabrication, finishing operations, inspection, rework, or test. These operations could be manual or automated. To model the shape of the service time distribution observed in previous studies (Knott and Sury, 1987), we assume that the processing time for product i, i=1, …, 8 at machine j, j = 1, 2, 3 has a shifted beta distribution with mean S_{ij}. Recall that a random variable X, with a shifted beta distribution, has a density function $f(x,\alpha,\beta)$ given by

$$f(x,\alpha,\beta)=\frac{1}{B(\alpha,\beta)}\frac{(x-a)^{\alpha-1}(b-x)^{\beta-1}}{(b-a)^{\alpha+\beta+1}}, a\le x\le b,\ \alpha>0,\beta>0 \qquad (1)$$

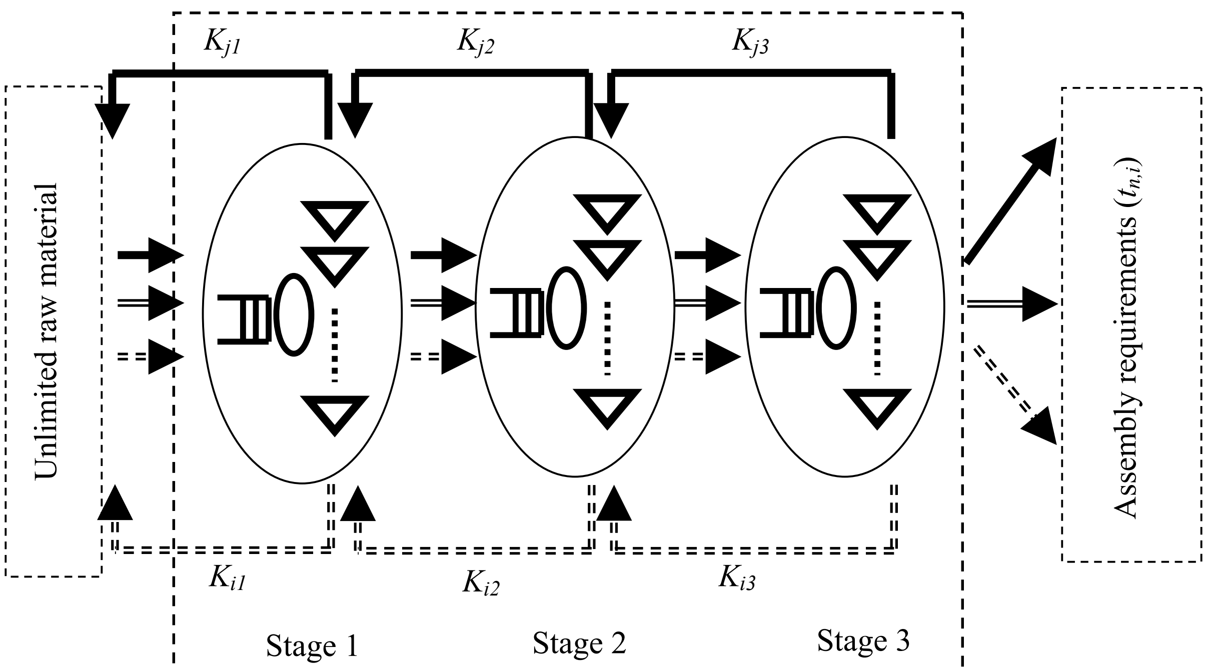
where a and b define the domain of X, and $B(\alpha,\beta)$ is the beta function defined by shape parameters α and β (Canavos, 1984). The expressions for mean, variance and coefficient of variation (CV) of these distributions are given in the Appendix. We assume that the setup time for each product is included in its processing time at each machine and that cell S follows a lot-for-lot policy (Vollmann et al., 1991).

To model the demand process, we assume that based on the information about final customer demands, assembly cells C_1, …, C_N develop their own production plan and place specific orders for different products with cell S. Usually each such order has a specific due date (or delivery date) to the assembly cell. As mentioned in Section 1, there are many situations where cell S might have such information regarding the orders. Let $t_{n,i}$ represent the time when the n^{th} order for product i is due to be delivered to assembly. For modeling purposes, we assume that the time between the due dates of successive requests for product i (i.e. $t_{n,i}$ - $t_{n-1,i}$) is exponentially distributed with mean $1/D_i$, i=1, …, 8. This would imply that the demand for product i occurs at a rate of D_i. We assume beta distributed service times and exponentially distributed times between assembly requests in order to model the variability observed in real systems.

In the above setting, depending upon whether cell S operates under a push or pull strategy, production would be triggered differently at the various workstations to satisfy the assembly requirements. Next, we give details of how these strategies are modeled in our simulations.

3.2. Simulation of pull strategy

When the multi-stage manufacturing line at cell S operates under a pull strategy, there are a fixed number of kanbans associated with each product in each stage. These kanbans circulate within a loop (see Figure 3), transmitting information regarding inventory consumption at the downstream stage and triggering production for inventory replenishment.



Fabrication Cell

Figure 3. System operating under the pull strategy.

More precisely, the system operates as follows. Initially, the output buffer at machine j, j = 1,2,3 has inventory of each product i, i=1,…, 8. Each unit in the output buffer has attached to it a product specific kanban card. Since there are K_{ij} kanban cards for product i at machine j, the initial inventory of product i at machine j equals K_{ij}. In the pull strategy, production is triggered when a customer request is satisfied by removing a product from inventory in the finished goods buffer at machine 3. (Orders not satisfied on the due date are backlogged until they are met.) When a product is removed from finished goods inventory at machine 3 to satisfy an order requirement, the kanban card attached to the product is released, triggering production of the corresponding product at machine 3. For production, machine 3 withdraws the corresponding inventory from the output buffer of machine 2, which in turn releases a card triggering production at machine 2 and so on back to machine 1. In this manner information about inventory consumption at the end of the line is transmitted to the rest of the line via the product-specific kanban cards.

As stated in Section 2, the major issue in designing a pull kanban system is determining the number of kanbans for each product at each stage of the manufacturing system. Since each product is attached to a kanban card, the number of kanban cards also bound the inventory at each machine in the line. Changing the number of kanban cards for each product at each machine not only impacts the rate of transmission of the production triggers and the average inventory at each machine, but also impacts the service levels. In the simulation experiments, we consider a wide range of kanban allocations for each product as discussed in Section 3.4.

3.3. Simulation of push strategy

To model the push strategy we associate with each product manufactured at cell S a parameter L_i. The parameter L_i called the release lead time for product i, is used to determine the timing of the material release into the manufacturing system. We assume that when the assembly cells place orders for product i with cell S they do so at least L_i time units before the due date. This assumption is justified if the downstream assembly cells fix their assembly schedules sufficiently in advance.

We simulate the push system as follows. Let $t_{n,i}$ denote the due date of the n^{th} order for product i at assembly. In the push strategy, for each such order, the corresponding raw material is released for production at stage 1 exactly L_i units of time in advance. (See Figure 4.) Upon release, the order is processed in turn at machines 1, 2, and 3, respectively. If processing is completed at machine 3 in advance of the due date, the product waits in the output buffer of machine 3. If on the due date, the order requirement for a given product can be satisfied from inventory at the output buffer of machine 3, the product is removed from the output buffer and the order is satisfied immediately. However because of congestion delays experienced at the different machines, the completion time of production at machine 3 might not be in advance of the due date of the order. In such a situation, on the due date the order is satisfied from safety stock in the finished goods buffer, if any. When the product is eventually completed at machine 3, it is used to replenish the safety stock consumed. Orders not satisfied on the due date are backlogged until they are met.

As mentioned in section 2, the major issues in designing a push strategy include estimating release lead times L_i, and determining the safety lead times and/or safety stocks required to guarantee the required service levels for the different products. Regarding the release lead times, we assume that for a given demand scenario, the planning system obtains an accurate estimate of F_i, the average flow time of product i. (Although in practice such estimates are themselves subject to error, we do not wish to complicate our experimental design space with this additional factor.) For our experiments the estimate of F_i is obtained from initial simulation runs for each scenario. When safety lead times are used, the release lead time L_i for each product is set higher than its expected product flow time F_i by an amount known as the safety lead-time. Safety stock for a given product is the inventory target in the output buffer when there is no unmet demand and no lots delayed in the manufacturing process beyond the actual flow time (Buzacott and Shanthikumar, 1994).

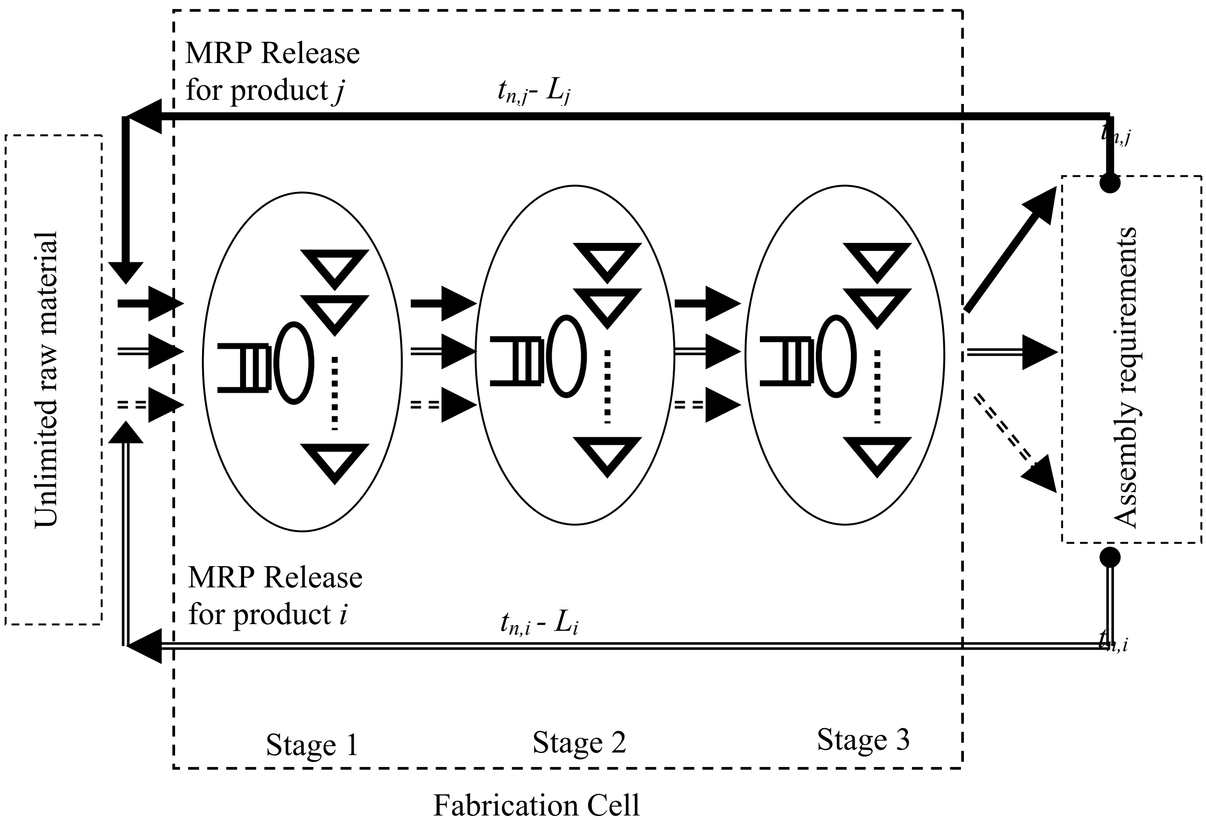


Figure 4. System operating under the push strategy.

We are not aware of any literature to date that shows when safety lead time would be preferable to safety stock in multi-stage multi-product manufacturing systems. To gain insight into this question, we study the safety stock and safety lead time policies independently. Specifically, we run two sets of simulation experiments for the push strategy. One set uses only safety stocks and the other uses only safety lead times. For the experiments with safety stocks, the release lead time L_i is set equal to F_i. Next, the total amount of safety stock (S_i^S) for product i is set to a fixed percentage of the average demand over its flow time F_i, and this percentage is varied over different experiments. In addition, we assume all the safety stock is located at the output buffer of the last machine. This implies that for the safety stock policy, the initial inventory at the buffers following machines 1 and 2 is zero while the initial final inventory at the output buffer of machine 3 is the specified safety stock, S_i^S. For the experiments with safety lead times, the total amount of safety lead time (S_i^L) for a product i is set to a percentage of its average flow time, F_i, and this percentage is varied over different experiments. Additionally, all the safety lead time for the product is added to its average flow time so as to advance the release of raw materials to the first machine by that amount. Therefore, for the safety lead-time policy, the initial inventory at all the buffers is zero and the release lead time L_i is set equal to $F_i + S_i^L$. The specific percentages for safety lead times and safety stocks used in our simulation experiments are given in Section 3.4. For convenience the main parameters of our simulation experiments are summarized in Table 1.

Table 1. List of main parameters used in simulation experiments.

Parameter	***Definition***
D_i	Mean arrival rate of requirements for product i
S_{ij}	Mean processing time for product i at machine j
$t_{n,i}$	Due date of n^{th} order for product i at assembly
L_i	Release lead time for product i
F_i	Average flow time for product i
S_i^L	Safety lead time for product i
S_i^S	Safety stock for product i
K_{ij}	The number of kanban cards of product i at machine j

3.4. Design of the simulation experiments

Since the performance of push and pull strategies vary for different values of safety stock and safety lead times and for different settings of kanban cards, we run simulation experiments for a wide range of parameter settings for each strategy. A more comprehensive approach would be to run simulation experiments to determine the optimal values of safety stock, safety lead times, and allocation of kanban cards. However, this entails excessive computational burden. Instead, we choose a wide range and suitably fine grained set of parameter values that significantly reduces the computational burden, captures the near optimal points in the design space, and also reveals the sensitivity of system performance to sub-optimal parameter settings. For the push system, we run experiments with safety lead times for each product being set at 0, 20%, 40%, 60%, 80%, 100%, and 200% of the average flow time, and with safety stocks for each product being set at 0, 20%, 40%, 60%, 80%, 100%, and 200% of the average demand over the flow time.

For the pull system, we consider several kanban allocations for each product. A particular allocation of kanbans for product i consists of a triple (K_{i1}, K_{i2}, K_{i3}). For the pull system simulations, we run experiments with kanban allocations for each product ranging across the following set of triples:

$$
\begin{aligned}
A = \{&(1,1,1), (1,1,2), (1,2,1), (1,2,2), (2,1,1), (2,1,2), (2,2,1), (2,2,2),\\
&(2,2,2), (2,2,4), (2,4,2), (2,4,4), (4,2,2), (4,2,4), (4,4,2), (4,4,4),\\
&(2,2,2), (2,2,8), (2,8,2), (2,8,8), (8,2,2), (8,2,8), (8,8,2), (8,8,8)\}
\end{aligned}
$$

Our main interest is in comparing the trade-offs between the service levels, average backorder delays, and average total system inventory for the two strategies for optimal and sub-optimal settings of safety lead times, safety stocks, and/or kanban allocations. Service level is defined as the proportion of requirements that are met on the due date. In computing the average backorder delay we include the observations where the customer requirements are met with no delay. In computing the inventory in the system at any time instant, we assume that any part that has been released for processing at the first station but has not yet been shipped to the customer counts towards inventory of the system at that time. Note that by this definition, any initial inventory present in the buffers at the start of the simulation also counts towards inventory in the system. In addition to computing service levels, average backorder delays, and average inventories, we also compute the average flow times for each product. The flow time is defined as the time from when a product is released to the input queue of machine 1 until the time it finishes processing at machine 3. These flow times provide a perspective for the average backorder delays experienced by customer demands. We compare the magnitude of the average backorder delay to the average flow time of the product.

For each simulation experiment the service level, average backorder delay, and average inventory levels are plotted on a three-dimensional figure and the points on the efficient frontier are used for comparison of push and pull strategies. For the simulation experiments of the push strategy we observe that the range of safety stock and safety lead times considered allows us to determine the efficient frontier for the demand settings considered in the experiments. Similarly for the simulation experiments of the pull strategy we observed that the set A of kanban allocations allows us to determine the efficient frontier for all parameter settings considered in the experiments. Therefore, we decided not to consider any additional kanban allocations or levels of safety stocks and safety lead times.

The simulation study is carried out using the PROMODEL® discrete event simulation tool (*www.promodel.com*). For each parameter setting, the simulation results are the average of six independent runs each of which represent the production of 66,000 customer orders. In each run, the statistics corresponding to the initial 6,000 orders were discarded to account for transient start-up effects. For each experiment, the inventory at each machine was considered for computing the average inventory in the system. For each experiment, the 95% confidence intervals were found to be within 3% of the mean values. This range is small as compared to the variation in performance along the efficiency frontier for the different push and pull system settings. The confidence intervals are omitted in the figures to improve readability.

4. Multi-product systems with homogeneous product mix

The first set of simulation experiments compare the impact of multiple products on the performance of push and pull strategies in a homogeneous environment as now explained. The values of S_{ij} and D_i are chosen assuming that the cell manufactures homogeneous products and is well balanced. Such a setting would be considered the ideal for pull strategy and one might expect it to perform the best. For the simulation experiments of both push and pull we set the mean processing times

$$S_{ij} = 1, \text{ for all } i \text{ and } j. \tag{2}$$

The service times have a shifted beta distribution with density function given by equation (1) with $\alpha = 0.75$, $\beta = 1.5$, $a = 0.5$, and $b = 2$. These values ensure that the service time distribution has an L-shape with mean 1 and CV = 0.39. The choice of shape of service time distribution and the value of CV is supported by studies reported in Knott and Sury (1987). Also, setting $a = 0.5$ and $b = 2$ ensures that the values of the service times are bounded above and below the mean by a factor of 2.

To set the mean demand rates, D_i, some discussion is required. Note that the performance measures we are primarily interested in are the tradeoff between service level and the average inventory in the system. This tradeoff obviously depends on the system load, i.e., throughput of the system. Therefore we explore this tradeoff at different values of system throughput. Since the throughput is governed by the mean arrival rates of demands, we conduct simulation experiments for different demand rates and observe the service levels and the corresponding average inventory in the system. Specifically, we set $D = \sum_{i=1}^{8} D_i$ equal to 0.5, 0.725, and 0.95, corresponding to low, medium, and high levels of system load respectively. This would correspond to throughput values of 0.5, 0.725, and 0.95, respectively. Further, since we assume for this first set of experiments that the line manufactures homogeneous products and is well balanced, we also assume that all products have identical demands. Therefore:

$$D_i = \frac{D}{8},\ i = 1,\ldots,8 \qquad (3)$$

Since all stations have a mean processing time $S_{ij} = 1$, for all i and j, equation (3) implies that when $D = 0.5$, 0.725, and 0.95, each of the three machines has a utilization of 50%, 72.5%, and 95%, respectively. Consequently, there is no unique bottleneck in the system. Note that the maximum total throughput of the system is one product per unit time. Since the average demands and service times for all the products are identical in this setting we refer to this as the *homogeneous product mix* case (see equations (2) and (3)). The simulation experiments are carried out for the different settings of safety stock and safety lead-times (for the push strategy) and kanban allocations (for the pull strategy) discussed in Section 3.4. Next we discuss the results and insights obtained from these experiments.

4.1. Results and insights

In Figure 5 we plot the average total inventory versus (1) service level and (2) the average backorder delay of a customer order. Figures 5(a) and 5(b) correspond to the case when D=0.5, Figures 5(c) and 5(d) correspond to D=0.725, while Figures 5(e) and 5(f) correspond to D=0.95. The graphs in Figure 5 also report the average flow times in each case. For certain kanban allocations, average backorder delays were greater than 50% of the average flow times. These are clearly inefficient kanban allocations and an efficient design of a pull strategy would not operate with such kanban allocations. To focus on the more relevant points of the design space, we omit displaying these points from the graphs comparing performance of push and pull strategies. From the graphs several observations can be made.

- Most importantly we observe that, contrary to the popular belief about pull, in this case, the total inventory required to meet any customer service level is *higher in the pull strategy* than in the push strategy for most observations on the efficient frontier.
- At low and medium values of system load, high service levels and reasonably small backorder delays are achievable for both push and pull strategies. As seen from the graphs, the average backorder delay rarely exceeds 25% of the average flow times. However, because the pull strategy is a replenishment strategy, a minimum inventory needs to be stored for the strategy to operate. Figures 5(a)-(d) indicate that at low to medium loads, this minimum inventory can be significantly higher than the inventory in a push strategy.
- When the system load is high (D=0.95), we observe from Figure 5(f) that for certain kanban allocations, low service levels and high backorder delays are observed despite having large inventories. Therefore, while operating under a pull strategy, it is important to choose the kanban allocation that guarantees the required service levels. However determining the optimal kanban allocation for each product is not an easy task. Furthermore, as can be seen from figure 5(f), a

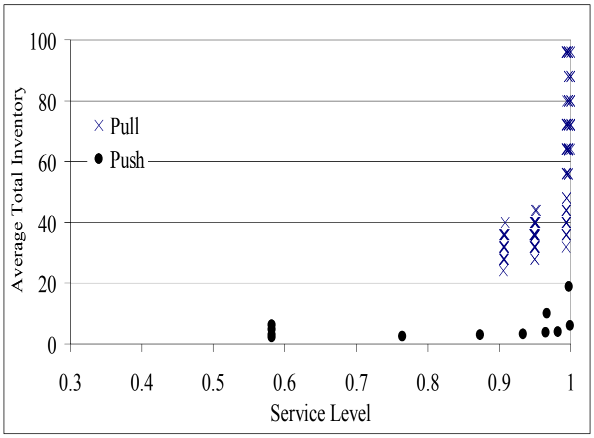
kanban allocation that guarantees a high service level might also result in high inventories.

- From Figure 5, we also observe that while certain kanban allocations yield highly inefficient system performance, the results of push strategies for different safety lead time and safety stock policies lie close to the efficient frontier. This indicates that in the cases considered here, push systems provide more robust performance than pull. We find this insight particularly interesting because in the single product case Hopp and Spearman (1996) have noted that pull is a more robust control system than push. Our study shows that in certain situations, this conclusion is reversed for the case of multiple products.

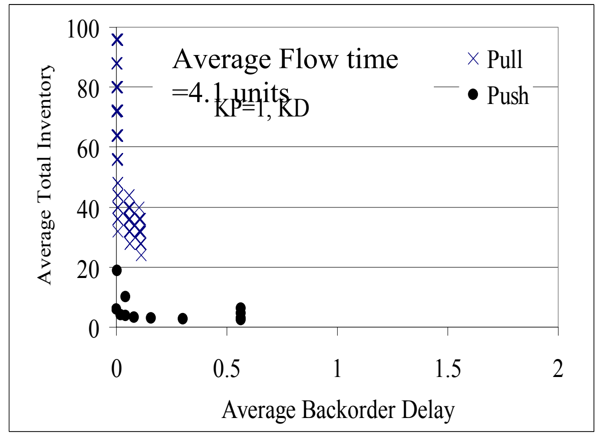
Although Figure 5 allows us to compare the tradeoff between inventory, service level, and the average delay in meeting customer orders, it is not clear from the figure how service level relates to the delay experienced by a particular order. Specifically, are there certain designs of push or pull systems wherein the service levels are high, but the average backorder delay is also high? To answer this question, in Figure 6 we plot the service level, the average backorder delay, and the total inventory on a three-dimensional figure. To put the average backorder delay experienced by a customer in perspective, the figure also provides the average flow time for the products. Note that the lower left hand corner in these graphs represents the efficient region of the system design space.

As before, we observe that for low and medium values of system loads (D=0.5 or 0.725; Figures 6a, 6b), both push and pull strategies yield high customer service levels with low backorder delays. However, for the pull strategy, the average system inventory is higher than that with the push strategy. In addition we observe that at higher system loads (D=0.95, Figure 6(c)), kanban allocations that result in poor service levels also result in large backorder delays. Furthermore, we observe that although certain kanban allocations might guarantee high service levels, the corresponding backorder delays could be significantly different for each of those allocations.

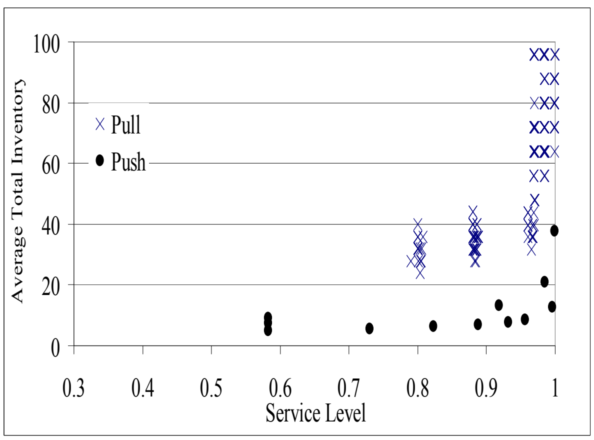
One of the reasons push yields less inventory than pull in a multi-product system is the following. Pull strategies require kanbans and hence inventory for every product at every stage of manufacturing. Correspondingly, there is a minimum average work in process inventory for each product in the system irrespective of the order patterns. We call this the *resident WIP*. A system operating under pull has no option but to carry the resident WIP for each product. In contrast, in the push strategy, the minimum WIP level is zero. When the MRP release is based on firm customer orders and accurate estimates of average flow times, any raw material released into the system is likely to be used to satisfy customer demand relatively soon after manufacturing is complete. Thus, the system operating under the push strategy can be leaner, i.e., have less wasteful inventory, than that operating under a pull strategy. Consider the results in Figure 5. At 72.5% load, the average time between customer orders for each product is 8/0.725, or 11.1 units of time, which is larger than the average time to manufacture the product (i.e. 5.8 units of time). This leads to excess inventory even for small numbers of kanbans. For higher values of system load (e.g. 0.95), the average time between customer orders is lower than the average flow time, but even in this situation the push strategy releases new raw material in a manner that is better timed with respect to the order due date, and thus performs better than the pull system. For any given system, the increased product variety can cause the average time between unit demands for the different products to be significantly greater than the average production flow times, in which case the pull strategy will initiate inventory replenishments well in advance of their requirements, resulting in excess WIP inventories. Buzacott and Shanthikumar (1993) use analytical models of single product systems to conduct similar comparisons of push and pull and obtain similar qualitative insights. They observe that in the presence of reliable information about demands in advance, push systems would have less inventory than pull for the same throughput requirements. Our simulation results indicate that this difference can be even more significant in certain multi-product systems due to resident WIP for each product.



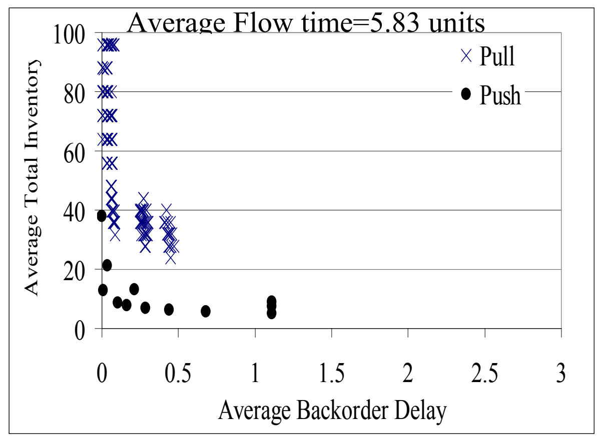
(a) D=0.5



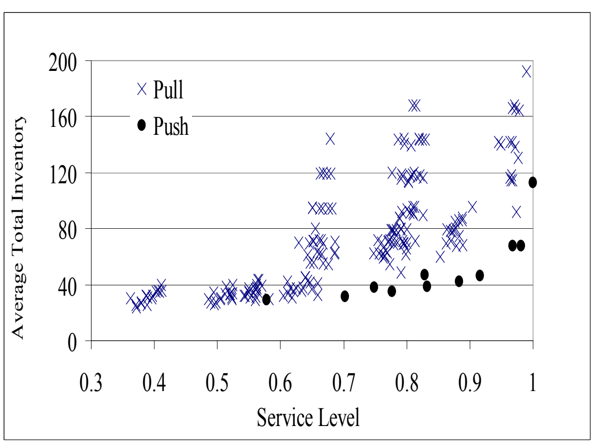
(b) D=0.5



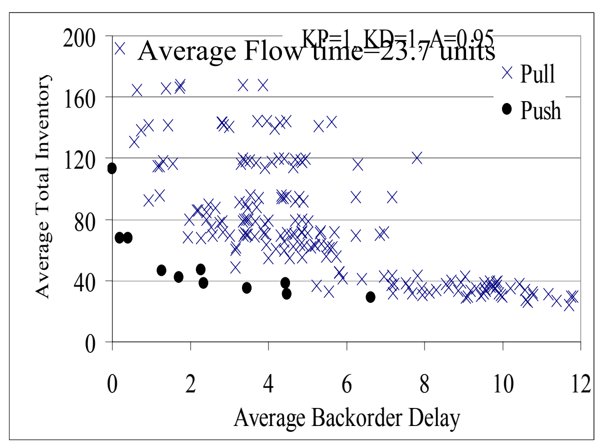
(c) D=0.725



(d) D=0.725



(e) D=0.95



(f) D=0.95

*Figure 5.*Comparison of push and pull for multi- product system with homogeneous product mix.

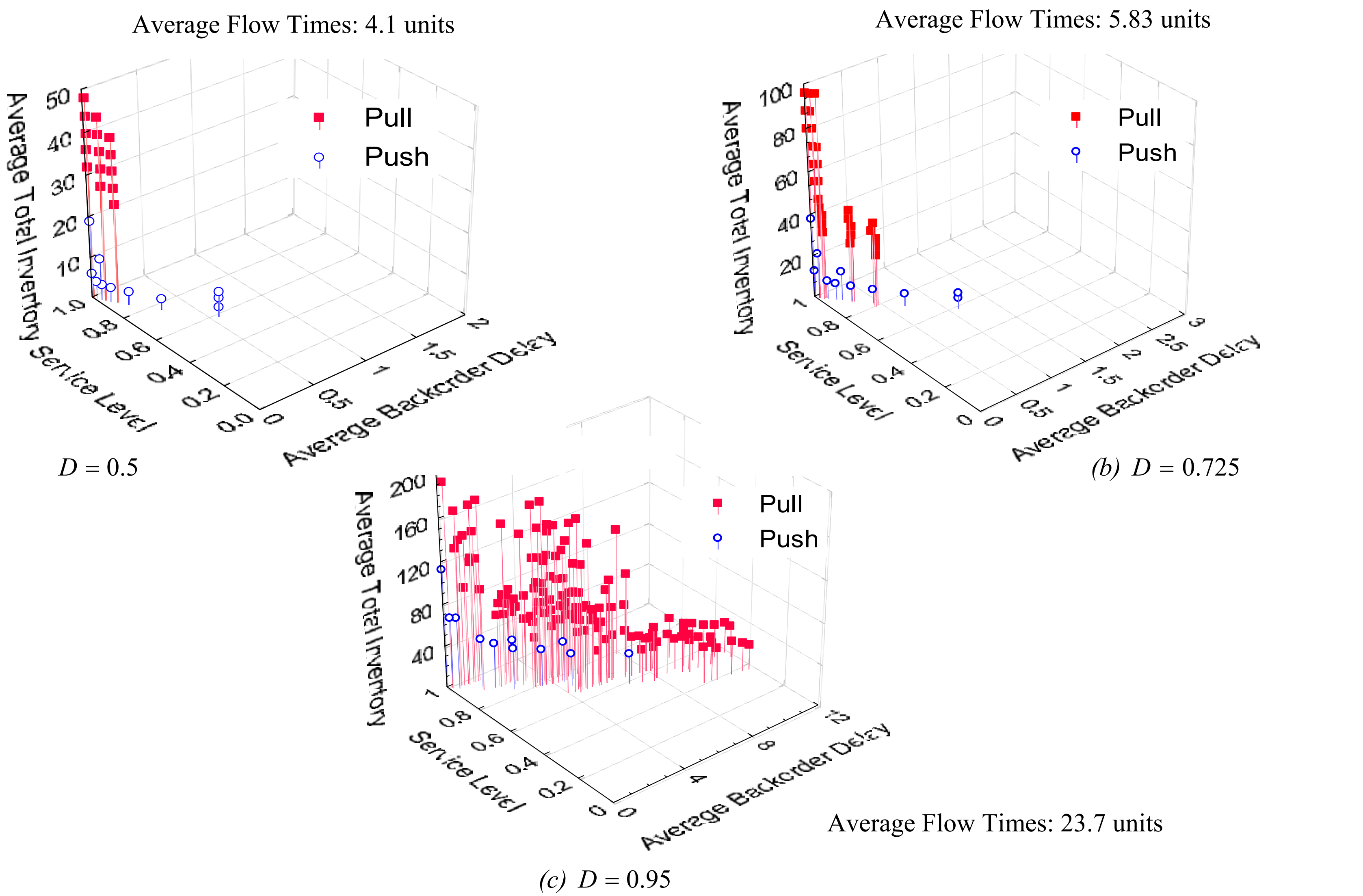


Figure 6. Performance tradeoffs with push and pull for multi-product systems with homogeneous product mix.

The impact of resident WIP observed from our simulation experiments also illustrates an important issue to be kept in mind while using analytical approaches based on product aggregation for the performance comparison of push and pull systems in multi-product environments. Although the manufacturing system with homogeneous product mix analyzed above appears to be amenable for such an analysis using product aggregation, it should be noted that if the approach does not incorporate the impact of product specific kanban control at each stage, the resulting analytical models might significantly underestimate the impact of resident WIP in multi-stage systems operating under pull strategies. Further the performance of pull strategies can be very sensitive to the kanban allocation patterns for the different products. These detailed insights cannot be ascertained from aggregate models.

5. Multi-product systems with heterogeneous product mix

In this section, we further examine the insights obtained in the previous section by comparing the performance of push and pull in multi-product environments with heterogeneous product mixes. In the previous section we assumed that the line manufactures products that are similar in terms of the rates for order arrivals and service times required at each machine. However, in practical systems, this is seldom the case. Product heterogeneity results if the products manufactured by department S for assembly departments C_1, …, C_N differ in terms of the arrival rate of customer orders, processing times at the different machines, or both. In the next set of experiments we investigate the performance of push and pull strategies in such situations.

For simplicity, in our simulation experiments we assume that the 8 products are divided into 2 groups, namely, Type 1 (i= 1, …, 4) and Type 2 (j =5, …, 8) such that,

$$D_i = \lambda_1, \text{ for } i = 1,...,4$$

$$D_i = \lambda_2, \text{ for } i = 5,...,8$$

$$S_{ij} = \tau_1, \text{ for } i = 1,...,4 \text{ and for all } j$$

$$S_{ij} = \tau_2, \text{ for } i = 5,...,8 \text{ and for all } j$$

As before, we still assume that service times have a shifted beta distribution with parameters $\alpha = 0.75$, $\beta = 1.5$, and coefficient of variation (CV) equal to 0.39, and $D = \sum_i D_i$:

$$4(\lambda_1 + \lambda_2) = D. \tag{4}$$

Additionally, for the sake of comparison with our simulation experiments in the previous section, we set the overal average service time at each machine to be equal to 1. That is,

$$\frac{4(\lambda_1\tau_1 + \lambda_2\tau_2)}{D} = 1. \tag{5}$$

In our experiments, we set D=0.5, 0.725, and 0.95, corresponding to low, medium and high levels of system load, respectively.

Equations (4) and (5) ensure that although the product mix is heterogeneous, the total load on the system remains the same as in the case of homogeneous system. Additionally, equation (5) also ensures that no new bottlenecks are introduced by product heterogeneity into the system. We do this so as not to confound the insights from our experiments with the impact of production bottlenecks.

We define two parameters, the *processing time factor*, k_P, and the *demand factor*, k_D, according to

$$k_P = \frac{\tau_1}{\tau_2} \text{ and } k_D = \frac{\lambda_1}{\lambda_2}. \tag{6}$$

Using equations (4), (5), and (6), we can express λ_1, λ_2, τ_1, and τ_2 in terms of k_P, k_D, and D, i.e.,

$$\lambda_1 = \frac{D}{4}\frac{k_D}{(1+k_D)},\ \lambda_2 = \frac{D}{4}\frac{1}{(1+k_D)},\ \tau_1 = \frac{k_P(1+k_D)}{1+k_P k_D}, \text{ and } \tau_2 = \frac{1+k_D}{1+k_P k_D}.$$

Note that k_P and k_D are the relative ratios of the mean processing times and demands of the two groups of products. By setting different values for k_P and k_D, we can capture a wide range of product mixes. In addition, if for a given product mix we have $k_P k_D$=1, the utilization of each machine by each product is the same.

To study the impact of heterogeneous product mix, we choose values of k_P=1/5, 1, and 5 and k_D=1/5, 1, and 5. Note that after taking symmetry into account there are only four distinct scenarios out of the nine possible combinations of these parameter values, namely where (k_P, k_D) equal (1/5,1), (1/5, 5), (1,5), and (5, 5), respectively. The other four cases are symmetric versions of these cases. For example, the cases (1, 1/5) and (1, 5) yield identical results since in both the cases the service rate for all the products are identical while the demand rate for one set of products is five times that of the other. Similarly, in the cases (1/5, 1/5) and (5, 5) the demand and the service rate for one set of products is five times the corresponding value of the other. Finally, note that, our initial system with homogeneous product mix corresponds to the ninth case (1, 1).

The selected values of k_P and k_D are designed to reflect the kinds of heterogeneous product mixes that occur in practice. For example, the case (1, 5) corresponds to the situation where the demand for one set of products is low while the demands for the other set of products is high. This might correspond to the situation where some products have a higher sales volume than others. Similarly the case when (k_P, k_D) equals (5,1) could correspond to the situation when the orders from certain assembly departments are more complex (require intricate machining, special tolerances) than those from the others. The other situations wherein (k_P, k_D) equal (1/5, 5) or (5, 5) could be combinations of the above possibilities. We could consider other combinations of values of k_P and k_D, but we restrict our experiments to these specific values for simplicity and to obtain the most important initial insights.

5.1. Results and insights

Figures 7 to 10 contain three-dimensional plots of service levels, average backorder delays, and average inventory levels obtained for system configurations with various heterogeneous product mixes and different system loads. The figures indicate that for heterogeneous product mixes, push strategies again guarantee better performance with less inventory than pull strategies. Several additional insights are also obtained. These are described below.

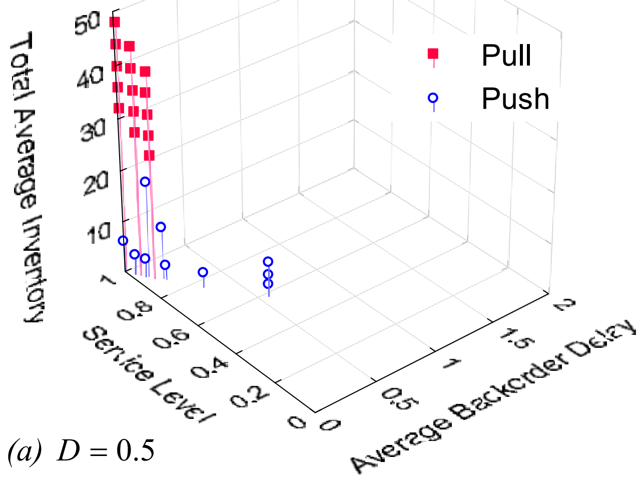
- From Figures 7, 8, and 10 we see that the dominance of push over pull is quite significant for all values of system load. These figures correspond to the cases where (k_P, k_D) equal (1, 5), (5, 5), and (1/5, 5), respectively. This implies that pull strategies are handicapped for product mixes with heterogeneous demands. In Figure 9, the product mix has homogeneous demands but heterogeneous mean processing times at each machine, and push is more effective than pull at low to moderate loads while having equal performance with less sensitivity to sub-optimal parameter values at high load.

- From Figures 9 and 10 we observe that higher average delays experienced by customers than for the corresponding system configurations in Figures 7 and 8. These correspond to the cases where (k_P, k_D) equal (1/5, 1), (1/5, 5), respectively. For (1/5, 1), when the manufacturing lines work on products with significantly different mean processing time, the average flow times are higher even though the load on each machine is still the same. This effect becomes even more pronounced when the demands are heterogeneous, as for (1/5, 5). Correspondingly, the average backorder delays experienced also get higher. In these situations a push strategy yields better service levels with lower average inventory than a pull strategy. Merely stocking inventory via additional kanbans is not the best way to meet service level requirements: incorporating a look-ahead feature yields significantly improved system performance.

The results from our simulation experiments agree qualitatively with the insights on the performance of push and pull strategies for multi-stage manufacturing systems in Buzacott and Shanthikumar (1993). Although their comparisons are restricted to single product systems some qualitative comparisons can be made on aggregate performance measures. Buzacott and Shanthikumar (1993) show that in multi-stage systems manufacturing a single product, if reliable information on demand and lead times are available, push systems have less inventory than pull systems. Our simulation results show that the difference in inventory levels required under push and pull is significantly greater in multi-product systems. Further, the results in Figures 7-10 indicate that overall service levels could deteriorate significantly under pull if kanban allocations are not chosen carefully.

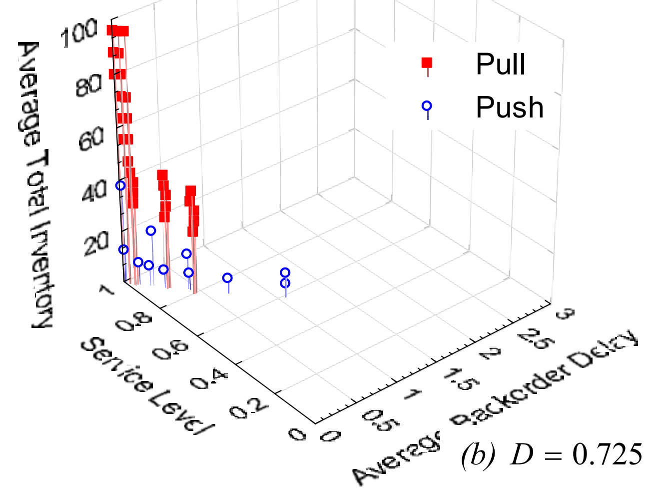
Buzacott and Shanthikumar (1993) also make observations on the impact of variability in the form of product mix diversity on flow times for the different products in a job shop. They show that in general product mix diversity results in an increase in the flow times of the different products. In our experiments too we observe that heterogeneity in product mix due to different processing times and demands for products increase the average flow time in the system which necessitates additional inventory in the system to meet the required throughput, regardless of whether the line operates under push or pull strategies. However, our studies show that in these environments, achieving reasonable service levels requires excessive amounts of inventory if the line operates under a pull strategy. This is because a pull strategy fails to incorporate valuable information on future demands when triggering production.

Average Flow Times:
4.1 units for Type 1, 4.1 units for Type 2

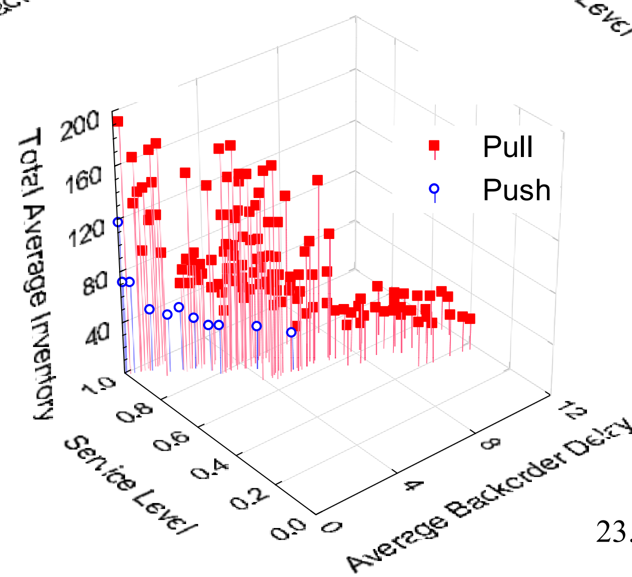


(a) $D = 0.5$

Average Flow Times:
5.8 units for Type 1, 5.8 units for Type 2



(b) $D = 0.725$

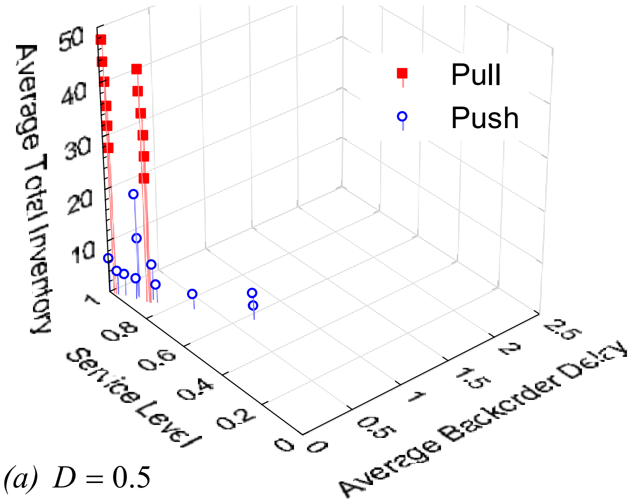


Average Flow Times:
23.7 units for Type 1, 23.7 units for Type 2

(c) $D = 0.95$

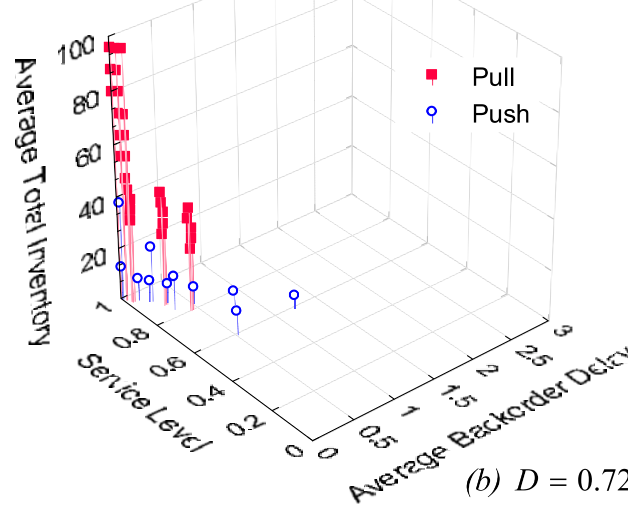
Figure 7. Performance tradeoffs with push and pull for multiple products with $k_P = 1, k_D = 5$.

Average Flow Times:
4.7 units for Type 1, 3.0 units for Type 2

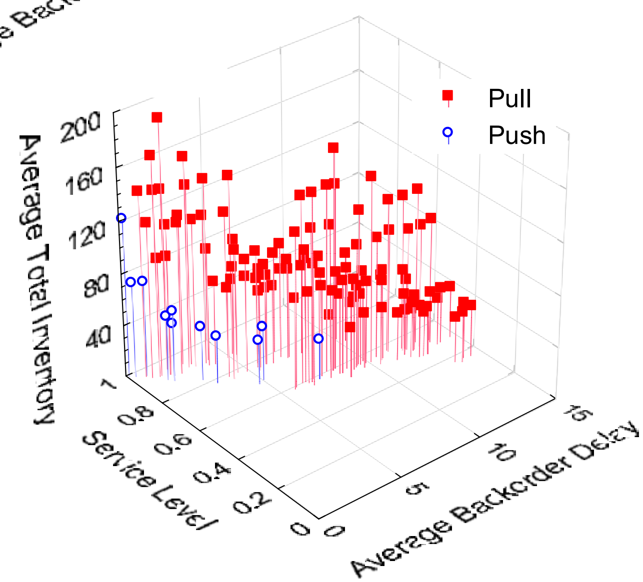


(a) $D = 0.5$

Average Flow Times:
6.6 units for Type 1, 5.4 units for Type 2



(b) $D = 0.725$



Average Flow Times:
25.8 units for Type 1, 25.0 units for Type 2

(c) $D = 0.95$

Figure 8. Performance tradeoffs with push and pull for multiple products with $k_P = 5, k_D = 5$.

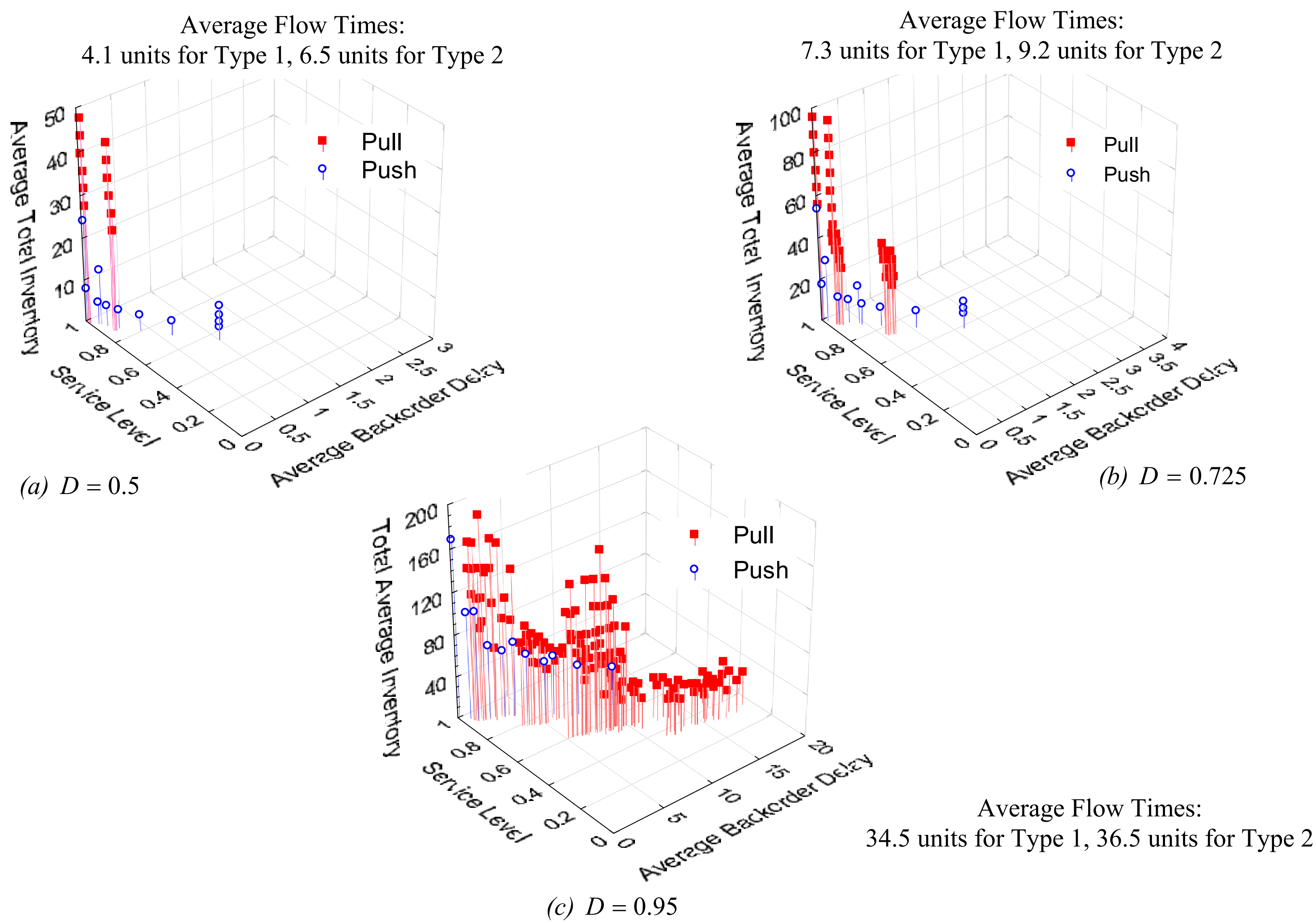
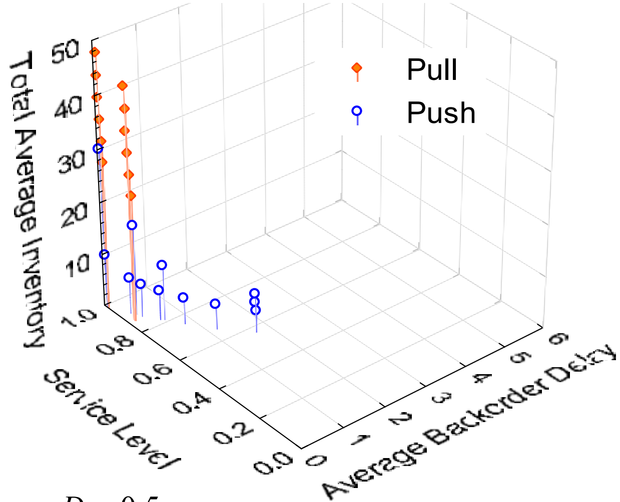


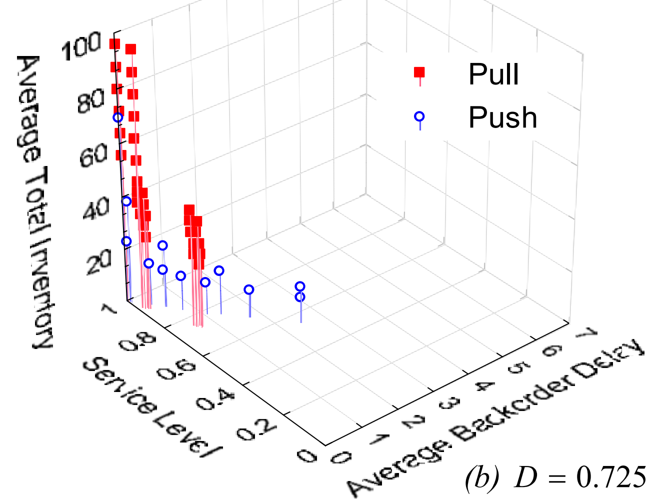
Figure 9. Performance tradeoffs with push and pull for multiple products with $k_P = 1/5, k_D = 1$.

Average Flow Times:
5.9 units for Type 1, 10.9 units for Type 2

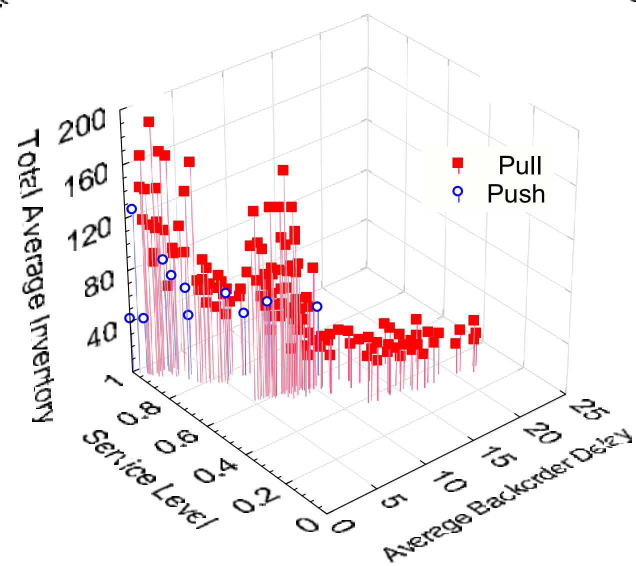


$D = 0.5$

Average Flow Times:
10.0 units for Type 1, 13.9 units for Type 2



(b) $D = 0.725$



Average Flow Times:
44.2 units for Type 1, 47.0 units for Type 2

(c) $D = 0.95$

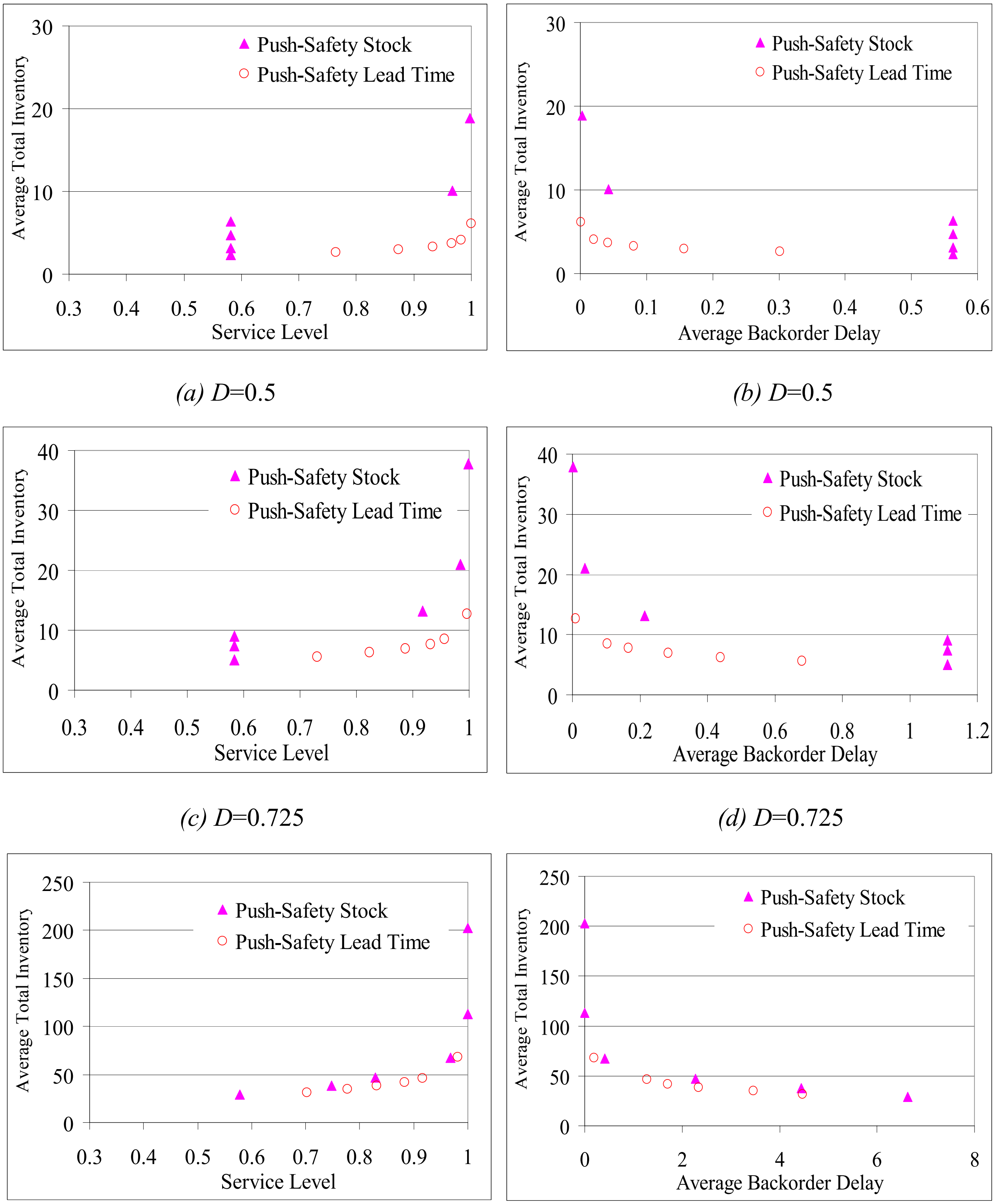
Figure 10. Performance tradeoffs with push and pull for multiple products with $k_P = 1/5, k_D = 5$.

6. On the use of safety stock versus safety lead times

In this section we use the results from our simulation experiments, to evaluate whether safety lead times or safety stock policies yield better system performance in multi-product push systems. In Section 2 we noted that, to our knowledge, there are no results in the literature that address this question.

Figure 11 compares the performance of the push strategy under both policies for the homogeneous product mix case (k_P=1, k_D=1) and Figure 12 compares the performance of the push system under both policies for one of the heterogeneous product mixes, specifically the system with k_P=1/5 and k_D=5. In both of these figures, we plot the average total inventory in the system against service level and average backorder delay. We observe that for low and medium values of system loads, i.e., D=0.5 and 0.725, the safety lead time policy provides better system performance than the safety stock policy. At high system loads the performance of the safety lead-time and safety stock policies are comparable. Furthermore, we observed similar insights for all product mixes considered in our study.

One possible explanation for this behavior is as follows. Introducing safety lead times or safety stock policies in the push strategy are both attempts to increase customer service levels by increasing the availability of finished goods at the last buffer. However their impact is significantly different only when the system load and average flow times are moderate. Although push systems with safety lead times release inventory into the system earlier than necessary, it is always against a future order due date. In the safety stock policy there is no such guarantee for the additional stock and therefore, the safety stock policy could result in inventories that are held for significant periods before they are used to satisfy orders. Based on these observations we conjecture that even in the case of multi-product systems, that have accurate information about average flow times and order due dates, the safety lead time policy is preferable to the safety stock policy.

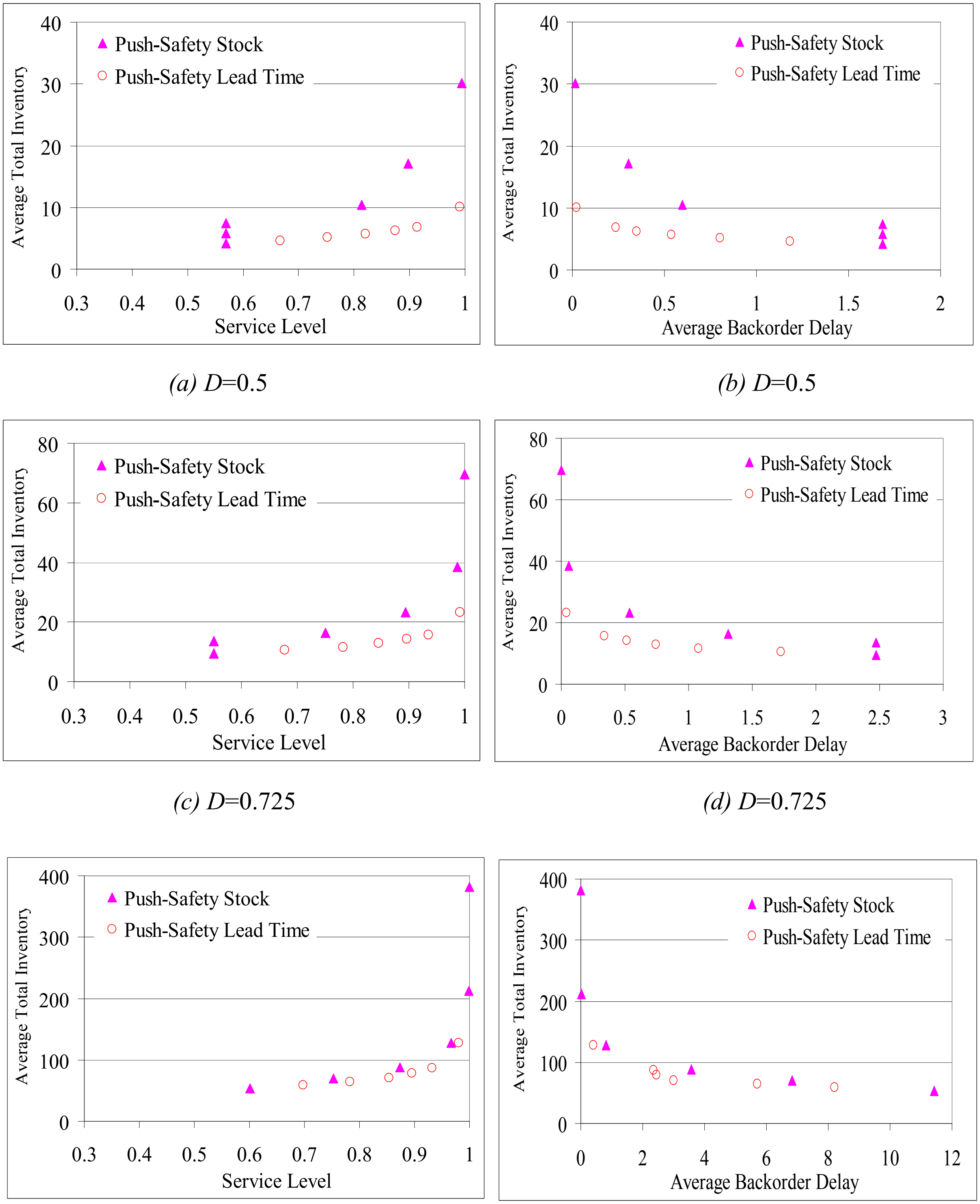


(a) D=0.5 *(b) D*=0.5

(c) D=0.725 *(d) D*=0.725

(e) D=0.95 *(f) D*=0.95

Figure 11. Safety stock versus safety lead time for multiple products with k_P=1, k_D=1.



(a) D=0.5 *(b) D*=0.5

(c) D=0.725 *(d) D*=0.725

(e) D=0.95 *(f) D*=0.95

Figure 12. Safety stock versus safety lead-time for multiple products with $k_P = 1/5, k_D = 5$.

7. Summary and conclusions

Manufacturing environments have undergone considerable changes in recent years and therefore it is important to understand the performance of material control strategies in different environments. In this paper we have compared the performance of MRP (push) and kanban (pull) strategies in a manufacturing setting where a fabrication cell S supplies different products to several assembly cells. Those comparisons assume that the assembly cells fix their assembly schedules in advance and share this information with their supplier cell S. For different system loads and product mixes, we compared the average total inventory at cell S to guarantee certain service levels under push and pull systems. We find that for this environment push outperforms pull in terms of service levels and average inventories. Further, in the pull strategy, if the kanban allocations are not set carefully, despite having high inventories the system could result in large average backorder delays and poor service levels.

The inferior performance of pull strategies in these environments has important implications not only for manufacturing planners but also for supply chain managers. That is, the manufacturing system considered in our study can also be viewed in the context of a supply chain. From a supply chain perspective, assembly facilities could correspond to different OEMs and facility S could be their common supplier. For example, S may be a contract supplier that manufactures and supplies particular components (fenders, hoods) to different assembly facilities in the auto industry. The successful implementation of pull systems for manufacturing systems with low product variety and stable demands has led to the belief that implementing pull strategies in all areas of the manufacturing system or across all partners in the supply chain would be mutually beneficial. However as our results have shown, pull strategies can perform poorly in certain multi-product environments. Manufacturing departments or suppliers that provide multiple products to assembly lines that share their assembly schedules can achieve better system performance by adopting a push strategy that explicitly considers future requirements when triggering production releases. Ignoring the available information on future requirements and adopting a pure pull strategy that merely replenishes consumed inventories may prove detrimental.

Using the results of our simulation experiments we also compared the performance of a push strategy operating under safety stock and safety lead time policies. Our experiments indicate that, as in the case of single product systems, when reliable information about future requirements is available, a push strategy operating with safety lead times yields better system performance than one operating with safety stocks.

Reliable information about future demands and flow times is not always available in practice. However, the results for such environments are useful because there are several practical situations where reliable information about future demands and average flow times are available. Recent studies that demonstrate the positive impact of sharing information about future requirements have motivated assembly departments to share schedules with departments supplying their components. Information sharing is also becoming popular among partners in a supply chain (Karaesmen et al., 2002). Furthermore, the increasing pressure to reduce lead times and stay competitive has forced many manufacturers to measure and monitor flow times for their products. In some cases flow times have become key metrics in evaluating supplier performance (Ericksen and Suri, 2001). Nevertheless, it would be interesting to study the performance of push and pull strategies when estimates of future demands and flow times are inaccurate and compare them against the insights presented in this paper.

There are several interesting directions for future research. The performance comparison of push and pull should be investigated under situations where reliable estimates of average flow times are not available, or when only partial information about future orders is available, or when planned lead times are larger than the horizon over which accurate information about demand is available. It would also be interesting to investigate which kanban allocations yield robust system performance in multi-product environments. In our experiments, we assumed that there is no dominant bottleneck so as not to confound the initial insights obtained. Testing the robustness of the performance of push and pull in the presence of bottlenecks is another area for further research. In this research our comparisons focused only on pure push (MRP-type) and pure pull (kanban-based) strategies. As mentioned before, several variations of push and pull strategies have been proposed (Liberopoulos and Dallery, 2000). In addition, hybrid strategies that combine different features of push and pull have also been proposed in the literature. The CONWIP system proposed in Spearman et al. (1990) and the POLCA system proposed in Suri (1998) are examples of such strategies. More recently, several researchers have proposed material control strategies that integrate advance demand information with pull control (Karaesmen et al., 2003). The qualitative arguments provided in favor of these strategies as well as the success of recent practical implementations of some of these strategies indicate that the performance of such strategies needs to be studied in greater detail. While simulation studies that address the different issues mentioned above would be useful, creating simulation models and designing experiments to test the impact of different parameters might become very complicated. On the other hand, queueing models have reasonable computational requirements and can be very useful in providing prescriptive insights. Developing analytical queueing models that analyze performance of push, pull, and hybrid strategies would also prove useful in understanding how different strategies perform under different settings.

Acknowledgements

The authors would like to thank the editors and the two anonymous referees for their careful comments and suggestions that helped to vastly improve this paper. The authors would like to acknowledge the National Science Foundation (Grant Number EIA-0127857), Rensselaer Polytechnic Institute, and the Center for Quick Response Manufacturing at the University of Wisconsin-Madison for funding this research.

Appendix

A random variable *Y* with a standard beta distribution has a density function $f(y,\alpha,\beta)$ given by:

$$f(y,\alpha,\beta) = \frac{1}{B(\alpha,\beta)}(y)^{\alpha-1}(1-y)^{\beta-1}, 0 \le y \le 1, \alpha > 0, \beta > 0$$

where $B(\alpha,\beta)$ is the beta function defined by the shape parameters α and β. For $\alpha < 1$ and $\beta \ge 1$, $f(y,\alpha,\beta)$ is L shaped while for $\alpha \ge 1$ and $\beta < 1$ $f(y,\alpha,\beta)$ is J shaped. For further details refer to Canavos (1984). The mean $E(Y)$, variance $Var(Y)$, and coefficient of variation $CV(Y)$ of the random variable *y* are given by:

$$E(Y) = \frac{\alpha}{\alpha+\beta},$$

$$Var(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$CV(Y) = \sqrt{\frac{\beta}{\alpha(\alpha+\beta+1)}}.$$

If *X* is a random variable with a shifted beta distribution, having a density function *f(x)*, on the domain [*a, b*], then

$$f(x,\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\frac{(x-a)^{\alpha-1}(b-x)^{\beta-1}}{(b-a)^{\alpha+\beta+1}}, a \le x \le b, \alpha > 0, \beta > 0$$

$$E(X) = \frac{\alpha b + \beta a}{\alpha+\beta},$$

$$Var(X) = \frac{(b-a)^2\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$CV(X) = \sqrt{\frac{(b-a)^2\alpha\beta}{(\alpha b+\beta a)(\alpha+\beta+1)}}.$$

Note: *X* can be obtained from the standard beta random variable *Y* by the transformation $X = a + (b - a)Y$.

References

Berkley, B.J., “A Review of Kanban Production Control Research Literature,” *Production and Operations Management*, Vol. 1, No. 4, pp. 393-411 (1992).

Buzacott, J.A., “Queueing Models of Kanban and MRP Controlled Production Systems,” *Engineering Costs and Production Economics*, Vol. 17, pp. 3-20 (1989).

Buzacott, J.A., Price, S.M., and Shanthikumar, J.G., “Service Level in Multistage MRP and Base Stock Controlled Production Systems,” in *New Directions for Operations Research*, G. Fandel, T. Gulledge, and A. Jones, Editors, Springer-Verlag, New York, NY pp. 445-463 (1992).

Buzacott, J.A. and Shanthikumar, J.G., *Stochastic Models of Manufacturing Systems*, Prentice Hall, Englewood Cliffs, NJ (1993).

Buzacott, J.A. and Shanthikumar, J.G., “Safety Stock Versus Safety Times in MRP Controlled Production Systems,” *Management Science*, Vol. 40, No. 12, pp. 1678-1689 (1994).

Canavos, G.C., *Applied Probability and Statistical Methods*, Little Brown and Company, Boston, MA (1984).

Ericksen, P.D. and R. Suri, “Managing the Extended Enterprise,” *Purchasing Today*, Vol.12, No. 2, pp. 58-63 (2001).

Gstettner, S. and Kuhn, H., “Analysis of Production Control Systems: Kanban and CONWIP,” *International Journal of Production Research*, Vol. 34, No. 11, pp. 3253-3273 (1996).

Hopp, W. and Spearman, M., *Factory Physics: Foundations of Factory Management*, Irwin/McGraw Hill, Chicago, IL (1996).

Karaesmen, F., Buzacott, J.A., and Dallery, Y., “Integrating Advance Order Information in Production Control,” *IIE Transactions*, Vol. 34, pp. 649-662 (2002).

Karaesmen, F., Liberopoulos, G., and Dallery, Y., “Production/Inventory Control with Advance Demand Information,” in *Stochastic Modeling and Optimization of Manufacturing Systems and Supply Chains*, J.G. Shanthikumar, D.D. Yao and W.H.M. Zijm, Editors, International Series in Operations Research and Management Science, Kluwer Academic Publishers, Dordrecht, Netherlands, pp.243-270 (2003).

Karmarkar, U.S., “Push, Pull and Hybrid Control Schemes,” *Tijdschrift voor Economie en Management*, Vol. 26, pp. 345-363 (1991).

Karmarkar, U.S., "Lot Sizes, Lead Times and In-Process Inventories," *Management Science*, Vol. 33, No. 3, pp. 409-418 (1987).

Knott, K. and Sury, R.J., "A Study of Work-time Distributions on Unpaced Tasks," *IIE Transactions*, Vol. 19, pp. 50-55 (1987).

Lambrecht, M.R., Muckstadt, J.A., and Luyten, R., "Protective Stocks in Multi-Stage Production Systems," *International Journal of Production Research*, Vol. 22, pp. 1001-1025 (1984).

Liberopoulos, G. and Dallery, Y., "A Unified Framework for Pull Control Mechanisms in Multistage Manufacturing Systems," *Annals of Operations Research*, Vol. 93, pp. 325-355 (2000).

Muckstadt, J.A. and Tayur, S.R., "A Comparison of Alternative Kanban Control Mechanisms-II. Experimental Results," *IIE Transactions*, Vol. 27, No. 2, pp. 151-161 (1995).

Orlicky, J., *Materials Requirement Planning*, McGraw Hill, New York, NY (1975).

Spearman, M.L., Woodruff, D.L., and Hopp, W.J., "CONWIP: A Pull Alternative to Kanban," *International Journal of Production Research*, Vol. 28, No. 5, pp. 879-894 (1990).

Spearman, M.L. and Zazanis, M.A., "Push and Pull Production Systems: Issues and Comparisons," *Operations Research*, Vol. 40, No. 3, pp. 521-532 (1992).

Suri, R., *Quick Response Manufacturing: A Companywide Approach to Reducing Lead Times*, Productivity Press, Portland, OR (1998).

Suri, R., "Quick Response Manufacturing: A Competitive Strategy for the 21st Century" in *Proceedings of the Quick Response Manufacturing 2000 Conference*, Rajan Suri (Editor), Society of Manufacturing Engineers, Dearborn, MI, pp. 1-27 (2000).

Suri, R. and de Treville, S., "Getting from 'Just-in-Case' to 'Just-in-Time': Insights from a Simple Model," *Journal of Operations Management*, Vol. 6, No. 3, pp. 295-304 (1986).

Tayur, S.R, "Properties of Serial Kanban Lines," *Queueing Systems*, Vol. 12, pp. 297-318 (1992).

Uzsoy, R. and Martin-Vega, L.A., "Modeling Kanban-based Demand-Pull Systems: A Survey and Critique," *Manufacturing Review*, Vol. 3, No. 3, pp. 155-160 (1990).

Vollmann, T.E., Berry, W.L., and Whybark, D.C., *Manufacturing Planning and Control Systems*, Dow Jones-Irwin, Homewood, IL (1991).

Womack, J.P. and Jones, D.T., *Lean Thinking: Banish Waste and Create Wealth in your Corporation*, Simon and Schuster, New York, NY (1996).

Zhou, X., Luh, P.B., and Tomastik, R.N., "The Performance of a New Material Control and Replenishment System: A Simulation and Comparative Study," in *Proceedings of the Quick Response Manufacturing 2000 Conference*, Rajan Suri, Editor, Society of Manufacturing Engineers, Dearborn, MI, pp. 805-826 (2000).