Design and Control of Agile Automated CONWIP Production Lines

Wallace J. Hopp,[1] Seyed M.R. Iravani,[2] Biying Shou,[3] Robert Lien[2]

[1] *Ross School of Business, University of Michigan, Ann Arbor, Michigan*

[2] *Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois*

[3] *Department of Management Sciences, City University of Hong Kong, Hong Kong S.A.R., China*

Received 22 August 2006; revised 17 March 2008; accepted 9 August 2008
DOI 10.1002/nav.20325
Published online 16 December 2008 in Wiley InterScience (www.interscience.wiley.com).

Abstract: In this article, we study the design and control of manufacturing cells with a mix of manual and automated equipment, operating under a CONWIP pull protocol, and staffed by a single agile (cross-trained) worker. For a three-station line with one automated station, we fully characterize the structure of the optimal control policy for the worker and show that it is a static priority policy. Using analytical models and extensive simulation experiments, we also evaluate the effectiveness of practical heuristic control policies and provide managerial insights on automation configuration design of the line. This characterization of the worker control policy enables us to develop managerial insights into the design issues of how best to locate and concentrate automation in the line. Finally, we show that, in addition to ease of control and greater design flexibility, the CONWIP protocol also offers higher efficiency and robustness than does the push protocol. Naval Research Logistics 56: 42–56, 2009

Keywords: workforce agility; CONWIP; Markov decision processes; automation

1. INTRODUCTION

The ability to provide high levels of (i) *efficiency*, as pursued through lean manufacturing, business process reengineering, and other methods for making more with less, and (ii) *responsiveness*, via time-based competition, agile manufacturing, and other methods for meeting diverse customer requirements in a prompt and personalized manner, has become a defining characteristic of competitiveness in the manufacturing sector. To achieve these objectives, firms have made extensive use of automated machinery (for efficiency) and *workforce agility* (for responsiveness). In cells with automated machinery, processing a job at a machine may not require the presence of a worker during the entire operation, and thus the agile (i.e., cross-trained) worker can operate another machine while the automated machine is running. We refer to systems with these features as *agile automated production* (AAP) environments.

Regardless of the nature of the manufacturing environment, there are two major ways to release jobs into production lines: push protocol and pull protocol. Under a push protocol, new jobs are pushed into the production system according to a predetermined schedule. In contrast, under a pull protocol, the release of new jobs depends on the status of the production system. Under a CONWIP (Constant Work In Process) pull protocol, a new job is released to the beginning of the line each time a job departs from the end of the line (Hopp and Spearman [12]).

Although AAP systems have become common in industry, they have received much less attention in the research literature than have traditional manual production lines and fully automated lines in which labor is not explicitly considered. As a result, we do not yet have a well-defined set of design and control principles for such systems. In this article, we attempt to partially fill this void by applying an analytic modeling approach to analyze AAP cells that are staffed by a single worker. Specifically, we focus on *control* and *design* issues in one-worker cells that are run under a CONWIP protocol and examine the following questions faced by operation managers:

1. **Control:** *How should the agile worker choose what to work on in an AAP cell?* In systems with cross-trained labor there will be times when the worker has a choice of which station to staff. We are interested in how sensitive performance is to the policy the worker follows and what an optimal policy looks like.

Additional Supporting Information may be found in the online version of this article.

Correspondence to: S.M.R. Iravani (iravani@iems.northwestern.edu)

2. **Design:** *Where in the cell is automation most effective?* In practice, managers' decisions regarding what operations to automate hinge on a variety of factors, including technological constraints, financial constraints, safety concerns, quality issues, as well as operational efficiency. We are interested in whether or when the position of an automated machine (i.e., in the front, middle, or back of the line) makes a difference in operational performance.
3. **Design:** *Should automation be concentrated or distributed?* It is often the case that operations are only partially automated. For example, a CNC (computer and numerically controlled) machine may automate processing but leave loading and unloading as manual operations that require an operator. Additional investment (e.g., purchase of automation to load and unload parts) could increase the level of automation at a given station. But is highly automating a single station preferable to partially automating multiple stations? We are interested in how automation should be allocated across a line.
4. **Efficiency and Robustness:** *Does superior performance (i.e., efficiency and robustness) of traditional CONWIP lines over push lines still hold in lines with agile workforce and automated equipment?* Many studies have been done to show the operational benefits of pull relative to push, e.g., Spearman and Zazanis [22], Roderick et al. [21]. These studies are all for conventional production lines (i.e., with neither automated machinery nor workforce agility). A natural question to ask, therefore, is how the relative performance of pull (in our case, CONWIP) and push is affected by the introduction of automated equipment and workforce agility.

Note that answers to the "Design" and "Efficiency and Robustness" questions clearly depend on the answer to the "Control" question. That is, we cannot analyze the impacts of design changes or compare the efficiency and robustness of AAP CONWIP and AAP push lines without specifying how the worker will behave. Therefore, in this article, after presenting a brief literature review in Section 2, we first focus on optimal and suboptimal (but commonly used) control policies in Section 3. We then investigate the design questions in Section 4. Finally, in Section 5, we compare and contrast our control and design principles for AAP CONWIP lines with those for AAP push lines in order to analyze the efficiency and robustness of AAP CONWIP lines. We conclude the article in Section 6.

2. LITERATURE REVIEW

A significant amount of literature exists on production systems with cross-trained workers. Askin and Strada [3] and Hopp and Van Oyen [13] provided surveys of these studies in the context of manufacturing cells and workforce agility, respectively. Ostalaza et al. [18], McClain et al. [15, 16], Zavadlav et al. [26], and Gel et al. [10] examined worksharing in a variety of situations. Bartholdi and Eisenstein [4] and Bartholdi et al. [5] studied bucket brigade lines. Farrar [9], Iravani et al. [14], Sennott et al. [23], Duenyas et al. [7], Ahn et al. [1], Andradottir et al. [2], and Van Oyen et al. [25] investigated the optimal assignment of flexible labor in tandem lines.

Although these articles yield many useful insights into the subject of workforce agility, all of them have focused on production environments without automated machinery, where processing a job at a workstation requires the presence of a worker during the entire operation. But, as we noted earlier, a key feature of AAP environments is that automated machinery permits workers to staff other stations while an automated station is processing a job. This opportunity can be significant, as indicated by Nakade et al. [17] who presented industrial examples where the percent of automated processing time (which does not require worker presence) to total processing time was as high as 80%.

There have been some efforts to explicitly model automation in agile production systems. For example, Nakade et al. [17] and Ohno and Nakade [19] analyzed serial AAP lines in which cross-trained workers visit their assigned workstations according to a cyclic policy. (See Section 3.3 for a detailed description of a cyclic policy). They obtained performance measures, such as cycle time and worker waiting time, under this policy. Desruelle and Steudel [8] investigated a similar system from a work cell design perspective. By modeling the work cell as two interacting queuing networks, an open part/machine network and a closed machine/operator network, they evaluated machine utilization and waiting times for the operator.

Hopp et al. [11] considered AAP cells operating under a *push* protocol and showed that the capacity of production lines with automated machines can be significantly lower than the rate of the bottleneck. They also showed that automation is more effective when placed toward the end of a push line rather than toward the front, and that the automation level increases the priority workers should give to a station when selecting a work location.

Because cellular manufacturing environments make use of *pull* mechanisms to promote efficiency, it is important to understand AAP cells in pull environments. The literature on pull is ample (for a review, see Uzsoy and Martin-Vega [24]). However, these articles are all for conventional production lines with neither automated machinery nor workforce agility. Hence, in this article, we focus on AAP cells that run under CONWIP pull protocol. Our main objective is to provide insights into the design and control principles of AAP CONWIP lines.

3. CONTROL OF AGILE AUTOMATED CONWIP LINE

As we mentioned, we cannot address the design issues in an AAP CONWIP line without specifying how the cell will be controlled. Specifically, we first need to find the most efficient (i.e., optimal) way to allocate the worker's effort among stations. To study the structure of the optimal worker control policy in CONWIP lines, we consider a simple three-station line that contains a single machine with automatic processing times, two machines with manual processing times, and one cross-trained worker. We call the station with the automated machine, the automated station, and the other stations, the manual stations. The line operates under a CONWIP protocol, where a new job is released to the beginning of line each time a job departs from the end of the line. We assume that the automated station is placed at the front of the line, whereas the two manual stations are put in the second and third positions in the line. (Note that this assumption about the position of the automated station is without loss of generality because of the closed-loop property of CONWIP lines.) For simplicity, we assume that the operation performed at the automated station includes a manual loading time and an automatic processing time. The automated machine requires the worker to be present only during the manual loading operations, whereas machines at manual stations require the worker to be present during the entire operation.

3.1. MDP Formulation

To construct a model, we assume that loading the automated machine requires an exponential amount of time with mean $1/l_1$. Although in practice the actual automatic processing times themselves may be close to deterministic, occasional interruptions (e.g., failures, adjustments, cleanings, material outages, etc.) will sometimes inflate the effective processing time. We approximate this behavior by representing the (effective) automatic processing times as exponential random variables with mean $1/\mu_1$. We also model the manual process times at stations $i = 2, 3$ as exponential with mean $1/l_i$.

The above assumptions allow us to formulate the problem of finding the operating policy that maximizes the throughput of the line as a Markov decision process (MDP). We define:

- System State is (n_1, n_2, s), where n_1 and n_2 are the WIP levels (including jobs in process) at the first and the second stations, respectively, and s refers to the status of the automated machine at Station 1. Specifically, $s = 1$ implies that the automated machine is performing automatic processing, whereas $s = 0$ implies it is not processing a job.
- Decision Epochs consist of job loading completion epochs at Station 1, machine processing completion epochs at Station 1, job processing completion epochs at Station 2, and job processing completion epochs at Station 3.
- Action Space includes the followings: (i) idling, (ii) loading the automated machine at Station 1 (if the machine is idle and the station is nonempty), (iii) processing a job at Station 2 (if there is a job at Station 2), and (iv) processing a job at Station 3 (if there is a job at Station 3).

Note that state of the system does not need to track the number of jobs at Station 3 because, under CONWIP, there is a constant number of jobs in the system, and so the number of jobs at Station 3 is determined by the number of jobs at stations 1 and 2. Denoting the WIP level in the CONWIP line by W and assuming that the worker can preempt a task to switch between stations, allows us to express the optimality equation for the MDP with the objective of maximizing the average throughput rate as:

$$g + V(n_1,n_2,0) = \frac{\mu_1}{\Lambda} V(n_1,n_2,0) + \frac{1}{\Lambda}\max\begin{cases} (l_1+l_2+l_3)V(n_1,n_2,0); \quad \text{idling} \\ \mathbf{I_1}[l_1 V(n_1,n_2,1) + (l_2+l_3)V(n_1,n_2,0)]; \\ \quad \text{loading at Station 1} \\ \mathbf{I_2}[l_2 V(n_1,n_2-1,0) + (l_1+l_3)V(n_1,n_2,0)]; \\ \quad \text{processing at Station 2} \\ \mathbf{I_3}[l_3 V(n_1+1,n_2,0) + (l_1+l_2)V(n_1,n_2,0)]; \\ \quad \text{processing at Station 3} \end{cases} \tag{1}$$

$$\begin{aligned} g + V(n_1,n_2,1) &= \frac{\mu_1}{\Lambda}[1 + V(n_1-1,n_2+1,0)] + \frac{l_1}{\Lambda}V(n_1,n_2,1) \\ &\quad + \frac{1}{\Lambda}\max\begin{cases} (l_2+l_3)V(n_1,n_2,1); \quad \text{idling} \\ \mathbf{I_2}[l_2 V(n_1,n_2-1,1) + l_3 V(n_1,n_2,1)]; \\ \quad \text{processing at Station 2} \\ \mathbf{I_3}[l_3 V(n_1+1,n_2,1) + l_2 V(n_1,n_2,1)]; \\ \quad \text{processing at Station 3} \end{cases} \end{aligned} \tag{2}$$

where $\Lambda = l_1 + l_2 + l_3 + \mu_1$, and for $i = 1, 2$,

$$\mathbf{I_i} = \begin{cases} 0; & \text{if } n_i = 0 \\ 1; & \text{otherwise} \end{cases} \quad \text{and} \quad \mathbf{I_3} = \begin{cases} 0; & \text{if } n_1 + n_2 = W \\ 1; & \text{otherwise} \end{cases}$$

Note that assuming exponential distribution for operation times helps us formulate the worker control problem as an

MDP model and gain insights into simple lines. However, later in this article we will also study more general lines in which operation times are not exponentially distributed.

We would like to emphasize that considering preemption in our model is not a restrictive assumption. Manual operations are often simple operations that can be interrupted and resumed later. On the other hand, because automated machines (e.g., numerically controlled (NC) machines) are expensive resources, it is not economical to keep those machines underutilized. Therefore, in order to keep the automated machines highly utilized, workers often preempt their tasks to attend stations with automated machines. In fact, this is what the optimal solution of our MDP model recommends (see Theorem 1). Furthermore, this article also considers policies that do not allow preemption. For example, in Sections 4 and 5, we investigate the design and control issues in lines under the cyclic policy, which is commonly used in practice and does not preempt tasks.

3.2. Structure of the Optimal Control Policy

Because the MDP problem has finite action and finite state spaces, there exists a stationary optimal policy with a constant gain [20]. Theorem 1 characterizes the structure of the optimal worker control policy in a three-station AAP CONWIP line. The proofs are presented in Supporting Information Appendix II.

THEOREM 1: For the three-station CONWIP line with an automated Station 1 and one agile worker

(i) The optimal policy is non-idling.[1]
(ii) When the automated machine in non-empty Station 1 is not processing a job, the optimal policy is to always load that machine.
(iii) When the automated machine in Station 1 is either starved[2] or processing a job, the optimal policy is to process a job, if there is any, at Station 3 which directly feeds the automated machine.

Theorem 1 implies that the optimal dynamic control policy is in fact a simple static priority policy that does not use the information about the WIP in each station in the line. This leads to a very simple principle of optimal worker control in the AAP CONWIP line: First check Station 1 (the automated station); load it if there is a job and the machine is not automatically processing. Then check Station 3 which directly feeds Station 1 and work there if a job is available; otherwise, process a job, if there is any, at Station 2. Idle only when none of the above conditions are satisfied.

[1] Non-idling means that the worker is never intentionally idle.
[2] Being starved means that there is no job available.

3.3. Heuristic Control Policies

In larger lines, in which several stations are automated and operation times may not be exponentially distributed, the optimal worker control policy would be too difficult to obtain and too complex to implement in practice (i.e., because actions depend not only on queue length but also on elapsed processing times). In practice, two simple, commonly used heuristic policies, namely the cyclic and fixed-priority policies, are used.

Fixed-Priority Policy: Similar to the policy introduced in Theorem 1, the fixed-priority policy is a static priority policy. Under a fixed-priority policy, the worker attends stations in a fixed order. The order specifies stations from the highest priority to the lowest. When the worker arrives at a station, she works at that station until she becomes idle, whereupon she switches to the next lower-priority station. If a job becomes available at a higher-priority station (e.g., an automated machine finishes processing and requires unloading), the worker interrupts her work at the current station and immediately switches to the higher-priority station.

Cyclic Policy: Under a cyclic policy, the worker attends workstations in a cyclic fashion. When the worker arrives at an automated station, she waits for the machine to finish processing the preceding job (if it is not finished), then unloads the processed job, loads the new job on the machine, switches the machine on, and then carries the completed job to the next station. If the worker arrives at a manual station, she processes the job at that station. If the worker arrives at a station with no jobs, she immediately moves on to the next station in her cycle.

One of the characteristics of the cyclic policy is that there is an upper bound on the total WIP in the line, beyond which any increase in the WIP has no impact on throughput of the line. The following theorem presents this property of cyclic policies. Note that the following theorem is general and does not depend on the distribution of loading, processing, or unloading times. The proof is presented in Supporting Information Appendix II.

THEOREM 2: In CONWIP lines with k automated stations, m manual stations, and one agile worker, the maximum throughput under a cyclic policy in which the worker visits stations in the direction of the material flow is obtained if and only if the line operates under a WIP level of $k + 1$.

Now that we have some insights into the characteristics of the optimal and heuristic control policies, we turn to design issues of AAP CONWIP lines. Specifically, we address the automation position and automation concentration under the optimal and heuristic control policies.

4. DESIGN OF AGILE AUTOMATED CONWIP LINES

We consider two design questions that we presented in the introduction: (i) Where in the line is automation most effective? (ii) Should automation be concentrated or distributed?

4.1. Impact of Automation Position

The decision of which station (or stations) to automate in a line can have a significant impact on the performance of the line. This decision depends on several technological as well as financial constraints and is affected by the trade-off between the cost of automating a station and the benefit resulting from the improvement in line performance. In this subsection, we investigate the impact of automation position on performance (i.e., throughput) of AAP CONWIP lines. We consider automation concentration in the next subsection.

To address the question of automation position in a AAP CONWIP line, we begin with the following theorem:

THEOREM 3: In CONWIP production lines with only one automated station, if job processing times for all stations follow the same probability distribution, the position of automation does not matter.

The proof is simple and follows from the fact that CONWIP lines are closed, cyclic queueing systems, so that the throughput of the line can be measured at any station in the line. Moreover, because the line is balanced, there is no capacity advantage to automating any given station.

However, when the line is unbalanced, the question regarding the position of automation is not simple. Below, we investigate the impact of automation position under both the cyclic and fixed-priority policies.

4.1.1. Impact of Automation Position: Cyclic Policy

To study the position of automation in CONWIP lines operating under a cyclic policy, we consider a manual two-station line with one worker in which t_i, the operation time at Station i, follows probability density function $f_i(t)$, $i = 1, 2$. Suppose that we have a certain amount of automation time, α, to allocate to either of the two manual stations. (For example, if we choose to place the automation at Station i, then the automatic processing time at Station i is α while the total manual operation time (i.e., loading and unloading time) is $t_i - \alpha$, where $\alpha < t_i$, for $i = 1, 2$).

An intuitive choice for automation is the bottleneck station, because it has the longest total operation time and might well seem to be the resource constraint in the line. However, we show that this is not always true. On the contrary, it turns out that automating the bottleneck station may be the worst choice. We can show this rigorously for the two-station case as follows. In a two station line operating under a cyclic policy, when we choose to place α time units of automation at Station i, then we will have the following theorem:

THEOREM 4: If job processing time at Station 1 (Station 2) is stochastically larger than that at Station 2 (Station 1), then automating Station 2 (Station 1) results in a larger increase in throughput.

Note that the station with the stochastically larger operation time has a larger average job processing time and is therefore the bottleneck. Hence, Theorem 4 implies that automating the bottleneck station is less effective than automating the non-bottleneck station in a two-station line using a cyclic policy.

To investigate whether the insight of Theorem 4 for systems with two stations also holds in larger systems, we performed an extensive simulation study of 180 different scenarios. Our simulation model assumes that the operations at automated stations include loading, automatic processing, and unloading. The automatic processing times are deterministic, whereas the manual operation times (i.e., loading and unloading an automated station, and the entire operation in a manual station) follow Gamma distributions. The Gamma distribution is very flexible in shape and therefore covers a wide range of variability scenarios. We consider three scenarios for variability of operations: low variability (i.e., coefficient of variation CV = 0.5), moderate variability (CV = 1), and high variability (CV = 2). Supporting Information Appendix I describes how we generated different three- and five-station AAP CONWIP lines with different scenarios for variability in manual processing, loading and unloading operations.

We simulated three-station and five-station unbalanced CONWIP lines to investigate the magnitude of the impact of automation position. In each case, we suppose that one station is a bottleneck and set its total operation time to be 1.2, 1.5, or 1.8 units, whereas the rest of the stations have operation times of 1 unit. Without loss of generality, we let Station 1 be the bottleneck. We then compare the effects of automating each station under scenarios with various automation and utilization levels. For example, in our five-station experiments, we consider different cases where Station 1, 2, 3, 4, or 5 is automated. For each case, we varied the automatic processing time in the automated station to be 0.2, 0.5, or 0.8, which presents, low, medium, and high level of automation. For example, when we automated 0.8 units of processing time of a station with a total of 1 unit, the remaining operation time (i.e., $1 - 0.8 = 0.2$) was divided equally between loading and unloading operations, each having a mean of 0.1 units and following a Gamma distribution.

We first ran simulations for AAP CONWIP lines under a cyclic policy. Note that there are $(3-1)! = 2$ ($(5-1)! = 24$) possible cyclic polices that can be used in a three-station (five-station) line. For any given WIP level, when Station i is automated, we searched exhaustively for the cyclic policy that results in the highest throughput. We call the resulting policy the optimal cyclic policy, and the corresponding throughput the optimal throughput. We compared the optimal throughput for different automation position scenarios (e.g., automating bottleneck and non-bottleneck stations).

Our simulation was written in C++. Runs ended after 20,000 jobs exited the line, in addition to a warm-up period of 20,000 jobs. Each run was replicated 20 times. For variance reduction purposes, common random number streams were used. Different random number streams were used for loading and unloading times at different stations to ensure independence. The standard error for each throughput result is no more than 0.0003. The throughput obtained by the simulation were compared at the 95% confidence level to check whether there is a statistically significant difference among them.

We observed in our 180 scenarios that the difference in throughput of the same line with different automation position was either not statistically significant, or if it was statistically significant, the difference was very small (i.e., the maximum percent difference for three-station lines was 1.04%, while that for five-station lines was 0.19%). We also observed that, similar to our findings for two-station lines, automating the bottleneck station results in a lower throughput than that achieved by automating a non-bottleneck station.

The intuition behind this phenomenon is as follows. When none of the stations are automated, the maximum throughput of the line is $1/\sum_{\forall i} t_i$, where t_i is the total average processing time in Station i. When Station j is automated, then the maximum possible throughput of the line is $1/[\sum_{\forall i} t_i - a_j]$, where a_j ($a_j < t_j$) is the automatic processing time at Station j. This maximum throughput is only achieved if the worker never idles at the automated station (waiting for the automated machine to finish its processing). The probability that the worker arrives at the automated station and sees that the machine is still working is higher as the time between the worker's departure from Station j and its next arrival at Station j becomes smaller. Note that this time is equal to $\sum_{\forall i \neq j} t_i$. On the other hand, if the bottleneck station is automated, then $\sum_{\forall i \neq j} t_i$ is smaller than when a non-bottleneck station is automated. This implies that, when the bottleneck station is automated, the worker idle time at Station j is longer, and thus throughput is smaller.

These results lead us to the conclusion that, when a cyclic policy is used in an AAP CONWIP line, the position of automation does not have a significant impact on throughput.

4.1.2. Impact of Automation Position: Fixed Priority Policy

To investigate the impact of automation position in AAP CONWIP lines operating under a fixed-priority policy, we conducted simulation experiments for three-station and five-station lines similar to those conducted for the cyclic policy. We found that the position of automation has very little effect on throughput (less than 0.2% for three-station lines and less than 0.21% for five-station lines). We also observed that automating the non-bottleneck station is more effective than automating the bottleneck station, but the difference is not very large. The intuition is similar to that for the cyclic policy.

Overall, our analysis suggests that in an AAP CONWIP line using either a fixed-priority or a cyclic policy, managers can concentrate on issues like financial and technological constraints without worrying too much about the impact of automation position on operational performance. This is in sharp contrast with the results for AAP Push lines, which we will discuss in Section 5.

4.2. Impact of Automation Concentration

The question we consider in this section is: Should automation be concentrated or distributed? In more practical terms, under a fixed budget, should we install one highly automated machine or several partially automated machines?

4.2.1. Automated Concentration: Cyclic Policy

In this section, we investigate the impact of automation concentration if the line operates under a cyclic policy. We first compare the throughput of the following two-station AAP CONWIP balanced lines. In both lines the total average processing times in each station is $t = 1/\tau$. The automated station consists of loading and automatic processing (we consider insignificant (zero) unloading time).

CON-CYC Line: In this line, automation is concentrated in Station 1. Specifically, Station 1 is automated, with $1/\mu$ units of automation time. The average loading time in this station is $1/l_c$ time units, and $(1/\mu) + (1/l_c) = 1/\tau$.

DST-CYC Line: In this line, automation is evenly distributed. Specifically, both stations have $1/2\mu$ units of automation time, and $1/l_d$ units of loading time, where $(1/2\mu) + (1/l_d) = 1/\tau$.

By Theorem 2 we know that, when two stations are automated, the maximum throughput can be obtained under a cyclic policy with a WIP level of 3. Assuming all manual operation times and automatic processing times are exponentially distributed, we can establish a Markov chain and obtain the throughput of the line (see Supporting Information

Table 1. Experiments on the effect of automation concentration (three-station, CV = 2).

	Cyclic policy				Fixed-priority policy			
Automation distribution scenarios	(1, 1, 1)	(2, 1, 1)	(1, 2, 1)	(1, 1, 2)	(1, 1, 1)	(2, 1, 1)	(1, 2, 1)	(1, 1, 2)
(0–0.9–0)	0.415	0.297	0.293	0.297	**0.472**	**0.321**	**0.321**	**0.321**
(0.1–0.7–0.1)	0.450	0.313	0.310	0.314	0.456	0.313	0.313	0.313
(0.2–0.5–0.2)	0.458	0.316	0.314	0.316	0.440	0.305	0.305	0.306
(0.3–0.3–0.3)	**0.462**	**0.317**	**0.317**	**0.317**	0.425	0.298	0.298	0.299
(0.4–0.1–0.4)	0.459	0.315	0.316	0.316	0.455	0.313	0.313	0.313
(0.45–0–0.45)	0.447	0.310	0.311	0.311	0.468	0.319	0.321	0.319
max difference (%)	10.2	6.3	7.6	6.3	10.0	7.2	7.2	6.9

The bolded value indicates the maximum throughput in each column.

Appendix II for details of the Markov model). The throughput of the line DST-CYC line can be obtained as follows:

$$\text{TH(DST-CYC)} = \frac{\mu l_d(3l_d + 4\mu)}{2[(l_d)^2 + 3\mu l_d + 4\mu^2]}. \quad (3)$$

For the CON-CYC line which has only one automated station, we know from Theorem 2 that the maximum throughput can be obtained under a cyclic policy with a WIP level of 2. Similarly, by developing a continuous time Markov chain for the CON-CYC line, we can calculate the throughput of the line as:

$$\text{TH(CON-CYC)} = \frac{\tau(\tau + \mu)}{2\mu + \tau}.$$

By substituting for l_d, the difference between the throughput values can be reduced to:

$$\text{TH(DST-CYC)} - \text{TH(CON-CYC)} = \frac{(2\mu - \tau)\tau^3}{(2\mu + \tau)(\tau^2 - 2\mu\tau + 8\mu^2)},$$

which is always positive, because $\mu > \tau$. This implies that distributed automation is more effective than concentrated automation when a cyclic policy is implemented in our two-station balanced AAP CONWIP line.

We used simulation to investigate the validity of this observation in our general models with three-station and five-station lines, and to evaluate the magnitude of the impact of automation distribution. We first considered a three-station balanced line where the mean total operation time on each station is 1 unit, which we denote by (1,1,1). We also studied unbalanced lines (2,1,1), (1,2,1), and (1,1,2). We assume that the total amount of automation (i.e., the sum of the automatic processing times in all stations) is 0.9 units. For each of the balanced and unbalanced lines, we started with the case in which Station 2 is highly automated (having all 0.9 min of automation), whereas the other two stations are manual, denoted as automation distribution scenario 0–0.9–0. Then we distributed automation gradually and evenly to stations 1 and 3. Note that 0–0.9–0 is the most concentrated case, whereas 0.3–0.3–0.3 is the most distributed case.[3] The resulting throughput values for the various automation distributions, under the cyclic policy, are given in Table 1.

Note that Table 1 presents the results for a three-station line under one of our five variability scenarios in which $CV = 2$ for all loading, unloading and manual processing times. The throughput of each AAP CONWIP line in the table is the maximum throughput of the line obtained by setting a high WIP level, which we call WIP_{max}.[4] Thus, the difference in line performance is not due to a difference in WIP levels, but instead is due to automation position or the worker control policy (i.e., cyclic or fixed-priority). Also, as we mentioned earlier, the throughput in the table corresponds to the optimal throughput under the optimal cyclic policy.

We observe from Table 1 that, for both balanced and unbalanced lines, the best line performance under the cyclic policy corresponds to the scenarios with the most distributed automation (i.e., (0.3–0.3–0.3)), shown in bold font in the table, and the worst line performance corresponds to the scenarios with the most concentrated automation (i.e., (0–0.9–0)). The "max difference" at the bottom of Table 1 thus corresponds to the relative difference between these two scenarios under the cyclic policy.[5] Therefore, as Table 1

[3] Because of the closed-loop property of a CONWIP line, we can significantly reduce the case space. For instance, automation distribution scenario 0.9–0–0 for unbalanced line (2,1,1) and automation distribution scenario 0–0–0.9 for unbalanced line (1,1,2) are both in effect the same as automation distribution scenario 0–0.9–0 for unbalanced line (1,2,1).

[4] We obtained the maximum throughput of the line by setting the WIP level high enough (i.e., to WIP_{max}), so that any increase in WIP level does not have a significant impact on line performance.

[5] The "max difference" is the difference between the throughput of the automation scenario with the maximum throughput and the throughout of the automation scenario with the minimum throughput, divided by the throughput of the automation scenario with the maximum throughput. It represents the percent increase in line throughput if the automation configuration is changed from the worst to the best (in each column).

shows, under a cyclic policy, concentrated automation is not as effective as distributed automation. For example, in the balanced line in Table 1, the maximum relative difference between the (0.3–0.3–0.3) automation scenario and the least throughput scenario is 10.2%, while this number of about 6.3% to 7.6% for the unbalanced lines. We repeated this experiment for three-station and five-station lines under different variability and WIP levels, and we observed the same phenomenon: If the line is operated under a cyclic policy, concentrated automation is not as effective as distributed automation.

The intuition behind the fact that distributed automation is more effective than concentrated automation when line operates under a cyclic policy is as follows. As Theorem 1 implies, it is optimal to give high priority to automated stations. When only one station is automated, that station gets the highest priority under the optimal control policy. However, when all stations are automated, as it is the case in of automation distribution, giving the highest priority to only one of the automated stations is not an effective policy. In fact, it is more effective to allocate the worker's time evenly among stations. The cyclic policy achieves this by visiting all stations in a cyclic manner, under which all automated stations get the same level of worker attention. Conversely, when automation is concentrated, the cyclic policy does not perform well, because it does not give the highest priority to the automated station(s).

Although in general, under the cyclic policy, distributed automation is better than concentrated automation, we also observed interesting cases in our simulation study in which automation concentration was slightly better than automation distribution. This occurred in cases corresponding to the variability scenario "load/unload CV = 2" and "manual CV = 0.5." The reason for this is as follows. When variability in loading and unloading operations is high and variability in manual operation is low, under automation concentration, the line has only one station with a very high variability (i.e., the automated station with high variability in loading and unloading), while the rest of stations (i.e., manual stations) have low variability (CV = 0.5). When automation is distributed, however, all of the stations have high variability in their operations (i.e., all stations have loading and unloading operations with high variability CV = 2). Thus, although the cyclic policy is more effective in cases with distributed automation, in this case the benefit is not high enough to outweigh the negative impact of high variability. Therefore, in lines using a cyclic policy, automation concentration is more effective than automation distribution (due to the high level of variability that results from automation distribution.) This suggests that, if the automation technology involves highly variable loading and unloading operations, then under the cyclic policy, having a single highly automated station (i.e., automation concentration) may be more beneficial than having several stations with low levels of automation, particularly if the manual operations in the line have low variability.

We would like to emphasize that, our numerical study includes 60 cases in which "load/unload CV = 2" and "manual CV = 0.5." We observed in only 24 out of 60 cases that under the cyclic policy, automation concentration was better than automation distribution. For those cases, the differences between the two configurations were very small, with an average difference of 2.1% and a maximum difference of 3.0%.

4.2.2. *Automated Concentration: Fixed-Priority Policy*

In this section, we investigate the impact of automation concentration if the line is run under a fixed-priority policy. To develop insight into this issue, we compare the throughput of a pair of two-station lines controlled by fixed priority policies:

CON-FXD Line: In this line the automation is concentrated in Station 1. Specifically, Station 1 is automated, with $1/\mu$ units of automation time. The average loading time at this station is $1/l_c$ time units, where $(1/\mu)+(1/l_c) = (1/\tau)$. The line operates under a fixed-priority policy that gives high priority to Station 1.

DST-FXD Line: In this line automation is evenly distributed. Specifically, both stations have $(1/2\mu)$ units of automation time, and $(1/l_d)$ units of loading time, where $(1/2\mu) + (1/l_d) = (1/\tau)$. Similar to the above, we assume that the line operates under a fixed-priority policy, and without loss of generality, we assume the worker gives high priority to Station 1.

Similar to Section 4.2.1, we can calculate the throughput for both lines by developing a continuous time Markov chain. Details are presented in Supporting Information Appendix II.

When WIP levels in both lines are set to WIP = 3, the throughput of the lines are found to be:

$$\mathrm{TH(DST\text{-}FXD)} = \frac{\mu l_d\left(32\mu^3 + 32\mu^2 l_d + 14\mu l_d^2 + 3l_d^3\right)}{64\mu^4 + 80\mu^3 l_d + 44\mu^2 l_d^2 + 13\mu l_d^3 + 2l_d^4} \tag{4}$$

$$\mathrm{TH(CON\text{-}FXD)} = \frac{\tau(\mu^2 + \mu\tau + \tau^2)}{2\mu^2 + \mu\tau + \tau^2}$$

By substituting for l_d, the difference between these two throughput levels can be reduced to:

$$\mathrm{TH(CON\text{-}FXD)} - \mathrm{TH(DST\text{-}FXD)} = \frac{\tau^2(32\mu^5 + 16\mu^4\tau - 42\mu^3\tau^2 + 23\mu^2\tau^3 - 5\mu\tau^4 + \tau^5)}{(128\mu^4 - 96\mu^3\tau + 40\mu^2\tau^2 - 6\mu\tau^3 + \tau^4)(\tau^2 + \mu\tau + 2\mu^2)}$$

which is positive, because $\mu > \tau$. This implies that concentrated automation is more effective than distributed automation when a fixed-priority policy is used in a balanced two-station AAP CONWIP line with WIP = 3.[6]

To confirm the analytical results obtained for the two-station line, we performed a simulation study (similar to that for cyclic policy) for three- and five-station lines. The right part of Table 1 shows our results for the three-station line under the optimal fixed-priority policy that results in the optimal throughput for each automation distribution scenario. As Table 1 shows, under a fixed-priority policy, concentrated automation is more effective than distributed automation. Table 1 also shows that the maximum relative difference between the throughput of the best automation distribution scenario, which is the most concentrated case (i.e., (0–0.9–0)), and that of the worst automation distribution scenario, which is the most distributed case (i.e., (0.3–0.3–0.3)), is 10.0% for a balanced line and around 7% for the unbalanced lines.

Our study of three-station lines under other variability and WIP level scenarios, as well as our analysis of five-station lines, also confirmed that automation concentration is more effective that automation distribution if the line operates under a fixed-priority policy.

The intuition behind this observation is as follows. As Theorem 1 implies, when only one station is automated (i.e., automation concentration), static priority policies, including fixed-priority policies, which give the highest priority to the automated station are optimal. Thus, fixed-priority policies are superior when automation is concentrated. When automation is distributed, as we explained in Section 4.2.1, policies that do not give the highest priority to one station and do not prioritize one automated machine over another (e.g., the cyclic policy) are more effective.

4.2.3. *Automation Concentration: Cyclic Policy Versus Fixed-Priority Policy*

In Section 4.2.1, we showed that the maximum throughput of the line under a cyclic policy is obtained when automation is evenly distributed. For example, for the balanced line case in Table 1, the maximum throughput under the cyclic policy is 0.462 (shown in bold font) and occurs under the automation distribution scenario (0.3–0.3–0.3). In contrast, in Section 4.2.2 we showed that the maximum throughput of the line under a fixed-priority policy is obtained when automation is concentrated. For example, Table 1 shows that the maximum throughput under the fixed-priority policy for the balanced line is 0.472 and occurs under automation distribution scenario (0–0.9–0). As this example depicts, concentrated automation combined with a fixed-priority policy results in 2.2% ($= (0.472 - 0.462)/0.462$) more throughput than distributed automation with the cyclic policy. This raises the questions: (i) Is concentrated automation with a fixed-priority policy is always more effective than distributed automation with a cyclic policy? (ii) What is the magnitude of the difference?

To show that concentrated automation with a fixed-priority policy is more effective than distributed automation with a cyclic policy, we compare the throughput of our CON-FXD line with that of DST-CYC line. Using (3) and (4), after some algebra, we get:

$$\text{TH(CON-FXD)} - \text{TH(DST-CYC)} = \frac{t^3(2\mu^2 - 2\mu t + t^2)}{(2\mu^2 + \mu t + t^2)(t^2 - 2\mu t + 8\mu^2)},$$

which is positive because $\mu > t$. Note that both TH(CON-FXD) and TH(DST-CYC) are for WIP level of three jobs. On the other hand, TH(DST-CYC) does not change for WIP levels above WIP $= 3$ (see Theorem 2), whereas TH(CON-FXD) is nondecreasing in WIP level. Thus, we still have TH(CON-FXD) $>$ TH(DST-CYC) for WIP > 3, which implies that in the balanced two-station line, a fixed-priority policy with concentrated automation results in higher throughput than a cyclic policy with distributed automation for all WIP ≥ 3.[7]

We further investigate the above insight in our general three- and five-station lines. Specifically, we compare the throughput of each AAP CONWIP line under the fixed-priority policy and under the cyclic policy, in order to answer the following questions:

1. If automation is concentrated, how much does the throughput increase if one switches from a cyclic policy to a fixed-priority policy?
2. If automation is distributed, how much does the throughput increase if one switches from a fixed-priority policy to a cyclic policy?
3. Is concentrated automation with a fixed-priority policy always more effective than distributed automation with a cyclic policy? What is the magnitude of the difference?

To study the first question, we compared the throughput of the lines with automation concentrated in one station operating under the fixed-priority policy and under the cyclic policy for all 300 cases. We observed that the fixed-priority policy

[6] One can develop a similar Markov chain for the cases with WIP $= 4$, or 5, which result in more complex expressions for the TH, but eventually lead to the same conclusion that TH(CON-FXD) $>$ TH(DST-FXD).

[7] It is easy to show that this result also holds for WIP = 2.

can result in up to a 13.8% increase in throughput (with an average increase of 4.5%) in three-station lines. In five-station lines, this difference was more subtle, with a maximum difference of 2.3% and an average difference of 0.7%. The larger differences were observed in cases where the line was balanced and the CV of the manual process times was high. Note that the smaller differences observed in five-station lines is due to the fact that operation time in one out of five stations in a five-station line constitutes a smaller portion of the total operation time in the line than does the time of one station out of three stations. Thus, if one station is automated in a five-station line, failure to use the best control policy has less impact on line performance than in a three-station line.

To investigate the second question, we compared the throughput of the lines with evenly distributed automation under the cyclic policy and under the fixed-priority policy for all 300 cases. We observed that the cyclic policy performs better than the fixed-priority policy, with a maximum difference of 9.6% and an average difference of 6.5% in three-station lines. In five-station lines, similar to the above, this difference was smaller, with a maximum difference of 3.3% and an average difference of 2.3%.

To find the answer to the third question, we compared the throughput of the lines with automation concentration and a fixed-priority policy with that of lines with evenly distributed automation and a cyclic policy for all 300 cases. We observed that, in general, the concentrated automation line with a fixed-priority policy achieves a higher throughput than does the distributed automation line with a cyclic policy. This is consistent with our analytical results for the two-station balanced line, which we presented in the beginning of this section. In three-station lines, the difference is most significant when the load/unload process times exhibit high CV. In such cases, a fixed-priority policy in a concentrated automation line achieves a throughput that is as much as 3.2% higher than that of the cyclic policy in an analogous line with distributed automation. In other CV scenarios for three-station lines, the difference between the two types of line is not significant. In the case of five-station lines, the difference is occasionally statistically significant, but the fixed-priority line exhibits a maximum of only 0.2% higher throughput, so it is not economically significant.

We observed an interesting behavior in the cases in which the manual operations have high variability (CV = 2) and loading and unloading operations have low variability (CV = 0.5). As opposed to what we observed in all other cases of our simulation study, some of the lines with distributed automation and a cyclic policy performed better than corresponding lines with concentrated automation and a fixed-priority policy. The explanation for this behavior is as follows. Because in these cases manual stations have very high variability, lines with concentrated automation have more variability in their operations than lines in which the same amount of automation is distributed across all stations (i.e., a line with no manual stations). The negative impact of high variability in the concentrated distribution lines is so high that it outweighs the effectiveness of the fixed-priority policy. However, we note that for those cases, the average and maximum difference between lines with distributed automation and a cyclic policy and lines with concentrated automation and a fixed-priority policy are very small, namely, 0.4% and 0.8%, respectively. So we view this result as interesting, but not of great practical importance.

The managerial implication of the above results is that in designing a CONWIP line with one cross-trained worker, one should consider either a concentrated automation configuration with a fixed priority policy, or a distributed automation configuration with a cyclic policy. However, the performance difference between the two is typically small if the load/unload CV is not high. Because automation concentration has a larger impact on performance in push lines (Hopp et al. [11]), this observation further substantiates our conclusion that design flexibility is a benefit of CONWIP.

5. AAP CONWIP LINES VERSUS AAP PUSH LINES

A key part of the pull literature has focused on the operational benefits of pull relative to push. For example, Spearman and Zazanis [22] provided theoretical evidence for the superior performance of pull. They showed that pull systems have less congestion (i.e., are more efficient), while pull systems are inherently easier to control (i.e., are more robust). By comparing CONWIP with kanban, they argued that the effectiveness of pull does not result from pulling per se, but rather from limiting WIP and WIP variability. Roderick et al. [21] compared CONWIP to other production systems using simulation and found that CONWIP has superior performance with respect to due date and cycle times. Buzacott and Shanthikumar [6] showed that if the value added at each station in the line is negligible, then CONWIP exhibits superior performance with respect to maximizing customer service subject to a WIP constraint. These articles are all for conventional production lines (i.e., with neither automated machinery nor workforce agility). A natural question to ask, therefore, is how the relative performance of CONWIP and Push is affected by the introduction of automated equipment and workforce agility.

To address this question, we compare agile automated CONWIP serial lines with agile automated push lines with respect to (i) control, (ii) design, and (iii) efficiency and robustness.

5.1. Control

For one-worker serial production lines with one automated machine operating under a push protocol, Hopp et al. [11]

showed that the optimal worker allocation policy that minimizes the average WIP in a two-station push line has a threshold structure. For example, when the automated machine in nonempty Station 1 is not processing a job (i.e., Case (ii) in Theorem 1), the agile worker must take into account the number of jobs at each station in order to decide whether to load the automated machine or process a job at Station 2. Using MDP models we can show that under a push protocol the optimal policy for three-station lines is even more complicated. In contrast, in the CONWIP environment the optimal action is simple; for example, as Case (ii) of Theorem 1 shows, when the automated machine in nonempty Station 1 is not processing a job, the worker should load the automated station regardless of the number of jobs at different stations. This confirms that, from the perspective of optimal control, single-worker AAP CONWIP lines are inherently easier to control than their AAP Push counterparts.

5.2. Design

In this section, we compare the design issues of automation position and automated concentration in AAP CONWIP and AAP push lines.

Automation Position: Hopp et al. [11] showed that if an AAP line is run as a push line, automating downstream stations is generally more effective than automating upstream stations. However, as we have shown in Section 4.1, in an AAP CONWIP line, the position of automation does not have a significant impact on line performance. This allows managers to concentrate on issues such as financial and technological constraints without worrying too much about the impact of automation position when they design an AAP CONWIP line. Consequently, design flexibility with respect to automation position is one of the benefits of CONWIP over push in AAP environments.

Automation Concentration: As we have shown, in designing AAP CONWIP lines with one cross-trained worker, one should consider either a concentrated automation configuration with a fixed priority policy, or a distributed automation configuration with a cyclic policy. However, the performance difference between the two is typically small. For AAP push lines, however, Hopp et al. [11] showed that automation concentration has a large impact on performance in push lines. Hence, we conclude that design flexibility with respect to automation concentration is a benefit of AAP CONWIP lines over AAP Push lines.

5.3. Efficiency and Robustness

Our analysis of efficiency and robustness of AAP CONWIP and AAP push lines parallels that of Spearman and Zazanis [22] for conventional CONWIP and push lines (i.e., lines with no automated machinery or workforce agility). In particular, we consider a simple static optimization model that balances the cost of lost production with the cost of added WIP, via a simple profit function of the form:

$$\text{Profit} = p\text{TH} - h\text{WIP},$$

where p is the marginal profit per job, TH is the throughput of the line, and h is the cost for each unit of WIP (including costs for increased cycle time, decreased quality, etc.). We consider several cases in which $h/p = 0.01$, 0.05, 0.1, and 0.25.

For a CONWIP line, throughput is a function of WIP; so, we seek a value of WIP that maximizes profit. In contrast, for a push line the average WIP is a function of the release rate; so, we find the value of TH (i.e., the job release rate) that maximizes profit. The question of efficiency is concerned with the relative performance (i.e., profit) of the two release protocols when their controls are optimized, whereas robustness is concerned with how fast profit degrades when WIP or TH are set at suboptimal levels.

We first compare the efficiency and robustness of CONWIP and push lines when they are run under their corresponding optimal worker control policies. Then, we extend our study to our general lines under cyclic and fixed priority policies.

5.3.1. Efficiency and Robustness Under the Optimal Control Policy

Because the MDP for three-station push lines becomes impractically large (Hopp et al. [11]), we focus on two-station lines to compare the efficiency and robustness of the CONWIP and push systems under the optimal worker control policies. Figure 1 (left) compares the profit curves of a two-station CONWIP line and a push line in which the second station is automated with an automation level of 80% (i.e., the mean automatic processing time accounts for 80% of the mean total operation time in the second station). The mean manual operation time at the first station, and the mean total operation time at the second station (i.e., loading and automatic processing time) are set to one unit to represent a balanced line. Fig. 1 (left) is for a case where $h/p = 0.01$, the same h/p ratio considered in Spearman and Zazanis [22]. We use our MDP models to obtain the throughput in CONWIP line and the MDP model from Hopp et al. [11] to compute the average WIP in the push line. Because WIP and throughput are measured in different units, we measure suboptimality in terms of percent error. For example, 70% on the horizontal axis indicates that the current control parameter (i.e., WIP in

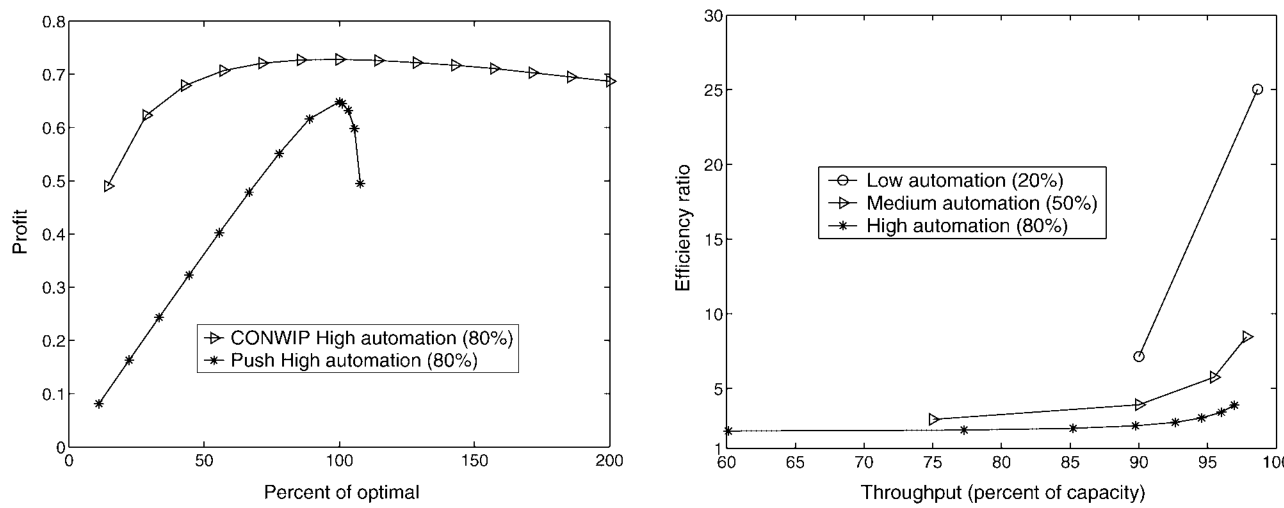

Figure 1. Comparisons for a balanced line operating under optimal Push and Pull settings: (left) efficiency and robustness and (right) impact of automation on robustness.

the CONWIP system or TH in the push system) is set at 70% of the optimal level.

Figure 1 (left) shows that CONWIP is more efficient than push, because it generates higher profit at the optimum. Furthermore, the profit function of the CONWIP line is very flat near the optimal WIP level, while in contrast, the profit function of the push line declines sharply when the release rate is set above or below the optimum level. Hence, Fig. 1 (left) also indicates that CONWIP is more robust than push with respect to the error in setting the optimal control parameter. We observed the same phenomenon in cases where $h/p = 0.05, 0.1, 0.25$.

To characterize the relative efficiency of push and pull in greater detail, we define the efficiency ratio between the CONWIP and the push lines for a given throughput rate as:

$$\text{efficiency ratio} = \frac{\text{average WIP of the push line}}{\text{WIP of the equivalent CONWIP line}}.$$

By "WIP of the equivalent CONWIP line" we mean the WIP in the line under CONWIP protocol that results in the same throughput as that achieved by the line when it is run under a push protocol. Therefore, an efficiency ratio greater than 1 indicates that the line under CONWIP carries less WIP than the same line under push, given the same throughput.

To investigate how the magnitude of automation in the line affects the relative performance of CONWIP and push strategies (i.e., the efficiency ratio), we considered three scenarios: (i) *high automation*, where the automated machine requires 0.2 units of mean manual loading time and 0.8 units of mean automatic processing time, which represents an 80% automation level; (ii) *medium automation*, where the automation level is 50%; and (iii) *low automation*, where the automation level is 20%. In our experiments, we successively increased the WIP level in the CONWIP line and recorded each average throughput rate, until the line saturated and throughput reached its maximum. Then we computed the average WIP level in the equivalent push line for each throughput rate achieved in the CONWIP line.

The results are shown in Fig. 1 (right). Note that lines with different automation levels have different capacities. This would make the comparison between scenarios unfair if we were to directly compare the efficiency ratios with regard to throughput rates. To avoid this, we plot the efficiency ratio vs. throughput as percent of capacity (which can be considered as line utilization). Note that there are fewer data points for the lower automation-level scenarios, because the line saturates at lower WIP levels. For example, in the 20% automation level scenario, throughput is essentially constant for all WIP levels of two or higher; as a result, the CONWIP line produces only two values of throughput to feed into the push line.

As Fig. 1 (right) shows, the efficiency ratios are always greater than one. This confirms that CONWIP is more efficient than push. Moreover, this efficiency advantage increases in the line utilization. However, it decreases in the level of automation. The reason for this behavior can be understood by looking at the two extreme cases of 100% and 0% automation. When the automation level is 100%, then the one-worker, two-station AAP CONWIP and push lines are both equivalent to a one-worker, one-station manual production line (because jobs flow through the automated station automatically without any help from the worker), and so the efficiency ratio is one. However, when the automation level of the automated machine is 0% (i.e., both the

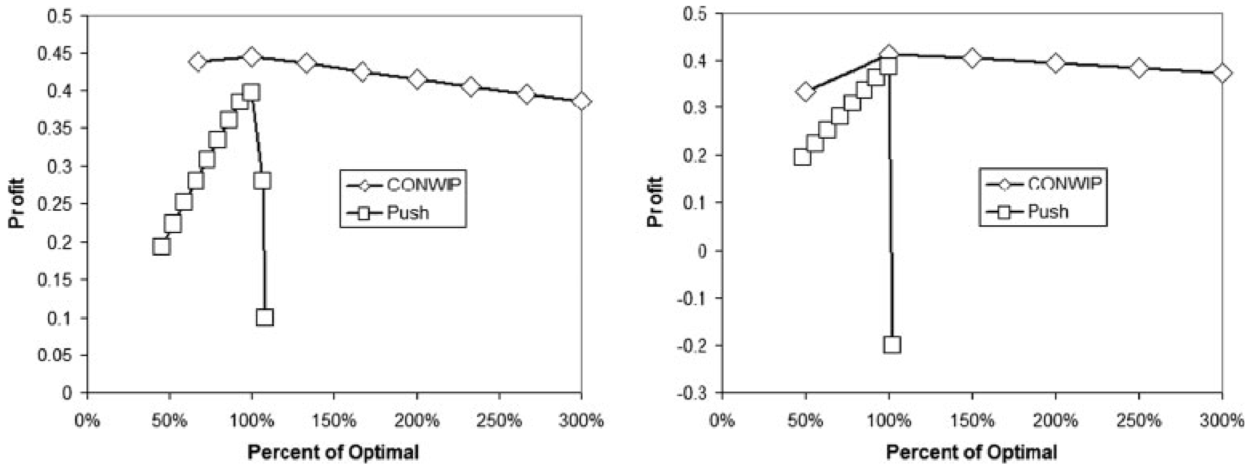

Figure 2. Robustness comparison in the three-station balanced line for "all CV = 1" scenario, $h/p = 0.01$, and fixed-priority policy: (left) one automated station and (right) three automated stations.

CONWIP and push lines are manual), then it is easy to see that a WIP level of 1 is sufficient for the CONWIP line to reach its capacity, while it takes infinite WIP for the push line to reach its capacity; hence, the efficiency ratio is infinite.

When the first station is automated, we observed that the above results still hold. We also conducted experiments for various line unbalance scenarios, by fixing the operation time in one station to be 1 unit and adjusting that in the other station to be 0.1, 0.2,…, 1. For each of these scenarios, we placed high (80%), medium (50%), and low (20%) automation at one station and used our MDP model to compare the performance of CONWIP and push, in a manner similar to our approach for balanced lines. We observed behavior similar to that shown in Fig. 1.

5.3.2. *Efficiency and Robustness Under Heuristic Controls*

We also examined the efficiency and robustness of AAP CONWIP line in our three-station and five-station lines with different levels of variability and under cyclic and fixed-priority policies.

We started with three-station balanced lines and assumed that the mean total operation time in each station is 1 unit, denoted as (1,1,1). We first considered cases in which only one station is automated. We set the automation level in the automated station to be 90% (i.e., 0.9 units of automation time in the balanced line), ran simulations for CONWIP and push lines, and computed the efficiency ratios at various WIP and variability levels. We then considered cases where all three stations are automated. We set the automation level on

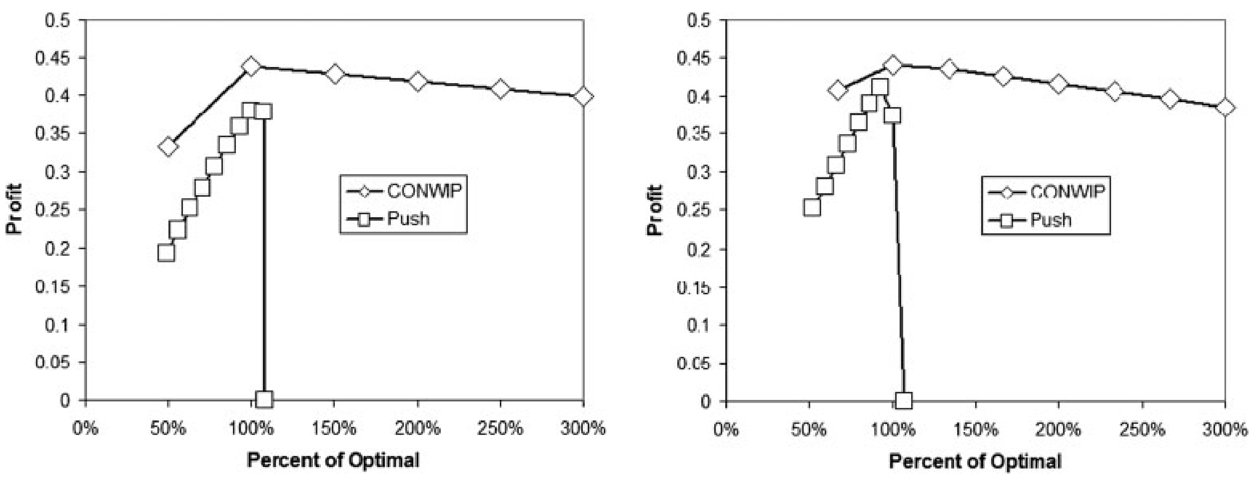

Figure 3. Robustness comparison in the three-station balanced line for "all CV = 1" scenario, $h/p = 0.01$, and cyclic policy: (left) one automated station and (right) three automated stations.

these stations to be 30% (i.e., 0.3 units of automation in each station of the balanced line), computed the efficiency ratios at various variability and WIP levels. We also conducted similar experiments for unbalanced lines where one station is the bottleneck with total operation time of 2 units, denoted as (2,1,1), (1,2,1), and (1,1,2). See Supporting Information Appendix I for the details of our design of experiments for unbalanced lines.

For each case, we studied the performance of CONWIP and push lines under the optimal fixed-priority policy and the optimal cyclic policy.[8] We observe behavior similar to that shown in Fig. 1 (left), that is, AAP CONWIP lines are more robust to control errors than AAP push lines. Fig. 2 demonstrates the robust performance of CONWIP relative to push in one of our three-station balanced lines where the best fixed-priority policy is implemented. Figure 2 (left) shows the case where only one station (i.e., Station 2) is automated with an automation level of 90%; Fig. 2 (right) shows the case where all three stations are automated and each has an automation level of 30%. Figure 3 shows the robustness of CONWIP relative to push in a similar setting except that the best cyclic policy is implemented.

Similar to what we observed previously under the optimal worker control policies, we find that the efficiency ratio between CONWIP vs. push is always greater than 1. Hence, AAP CONWIP lines always carry less WIP than their push counterparts to achieve the same throughput.

We repeated this experiment for our five-station balanced and unbalanced lines with different variability and WIP levels under both fixed-priority and cyclic policies and observed the same phenomenon. Therefore, based on our results for lines operating under both optimal and heuristic control policies, we conclude that CONWIP outperforms push with regard to efficiency and robustness in agile automated environments.

6. CONCLUSIONS

This article investigates the control, design, efficiency and robustness issues of manufacturing cells with automated equipment staffed with a single cross-trained (agile) worker. We show that the optimal worker control policy in a three-station line with one station automated is a static priority type policy, which gives the highest priority to the automated station and the second highest priority to the station that feeds the automated station. Through an extensive simulation study of lines with three and five stations under both fixed-priority and cyclic control policies, we show that for the same units of automation time, automating the non-bottleneck stations is more effective than automating the bottleneck station, although the difference is not very large. Furthermore, in general, under the fixed priority policy, automation concentration is better than distributed automation, whereas under the cyclic policy, automation distribution is more effective. Thus, in designing CONWIP AAP lines with one worker, one should consider either a concentrated automation configuration with fixed-priority or a distributed automation configuration with a cyclic policy. We also show that the primary benefits of pull (CONWIP in particular), namely, observability, efficiency, and control, extend to systems with automated equipment and cross-trained (agile) workforce. Finally, we identify line design flexibility as a fourth advantage of CONWIP protocol in the agile automated production environments. That is, although push systems are sensitive to the placement and concentration of automation, CONWIP systems are not, which means that automation decisions can be based on other issues, such as cost and quality.

To see whether our insights are indeed robust in more general systems, further research is needed into systems with multiple-product types, multiple workers, and other features, such as machine failures, rework, and partial cross-training. Given the growing importance of AAP systems in industry, such research would be of great practical significance.

[8] As we mentioned before, the *optimal* fixed-priority policy (cyclic policy) is the policy among all possible fixed-priority policies (cyclic policies) that results in the maximum profit.

ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation under grant DMI-0114598. The third author would also like to thank the research grant support from City University of Hong Kong.

REFERENCES

[1] H. Ahn, I. Duenyas, and R.Q. Zhang, Optimal stochastic scheduling of a two-stage tandem queue with parallel servers, Adv Appl Prob 31 (1999), 1095–1117.
[2] S. Andradottir, H. Ayhan, and D.G. Down, Server assignment policies for maximizing the steady-state throughput of finite queueing systems, Management Sci 47 (2001), 1421–1439.
[3] R.G. Askin and S. Strada, "Investigation of cellular manufacturing practices," Handbook of cellular manufacturing systems, S. Irani (Editor), John Wiley, New York, 1999, pp. 25–34.
[4] J.J. Bartholdi III and D.D. Eisenstein, A production line that balances itself, Oper Res 44 (1996), 21–34.
[5] J.J. Bartholdi III, D.D. Eisenstein, and R. Foley, Performance of bucket brigades when work is stochastic, Oper Res 49 (2001), 710–719.
[6] J.A. Buzacott and J.G. Shanthikumar, "The PAC system: a general approach for coordinating multiple cell manufacturing systems," Conference on Multi-Echelon Inventory systems, Berkeley, CA, 1991.
[7] I. Duenyas, D. Gupta, and T.M. Lennon, Control of a single server tandem queueing system with setups, Oper Res 46 (1998), 218–230.

[8] P. Desruelle and H.J. Steudel, A queueing network model of a single operator manufacturing workcell with machine/operator interference, Management Sci 42 (1996), 576–590.
[9] T.M. Farrar, Optimal use of an extra server in a two station tandem queueing network, IEEE Trans Automatic Control AC-38 (1993), 1296–1299.
[10] E.S. Gel, W.J. Hopp, and M.P. Van Oyen, Factors affecting opportunity of worksharing as a dynamic line balancing mechanism, IIE Trans 34 (2002), 847–863.
[11] W.J. Hopp, S.M.R. Iravani, and B. Shou, Serial agile productions systems with automation, Oper Res 53 (2005), 852–866.
[12] W.J. Hopp and M.L. Spearman, Factory physics: Foundations of manufacturing management, Irwin/McGraw-Hill, Burr Ridge, IL, 2000.
[13] W.J. Hopp and M.P. Van Oyen, Agile workforce evaluation: A framework for cross-training and coordination, IIE Trans 36 (2004), 919–940.
[14] S.M.R. Iravani, M.J.M. Posner, and J.A. Buzacott, A two-stage tandem queue attended by a moving server with holding and switching costs, Queueing Syst 26 (1997), 203–228.
[15] J.O. McClain, K.L. Schultz, and L.J. Thomas, Management of worksharing systems, Manuf Serv Oper Management 2 (2000), 49–67.
[16] J.O. McClain, L.J. Thomas, and C. Sox, On-the-fly line balancing with very little WIP, Int J Prod Econ 27 (1992), 283–289.
[17] K. Nakade, K. Ohno, and J.G. Shanthikumar, Bounds and approximations for cycle times of a U-shaped production line, Oper Res Lett 21 (1997), 191–200.
[18] J. Ostolaza, J. McClain, and J. Thomas, The use of dynamic (state-dependent) assembly-line balancing to improve throughput, J Manuf Oper Management 3 (1990), 105–133.
[19] K. Ohno and K. Nakade, Analysis and optimization of a U-shaped production line, J Oper Res Soc Jpn 40 (1997), 90–104.
[20] M.L. Puterman, Markov decision processes, Wiley, New York, 1994.
[21] L.M. Roderick, D.T. Phillips, and G.L. Hogg, Comparison of order release strategies in production control systems, Int J Prod Res 30 (1992), 611–626.
[22] M.L. Spearman and M.A. Zazanis, Push and pull production systems: issues and comparisons, Oper Res 40 (1992), 521–532.
[23] L.I. Sennott, M.P. Van Oyen, and S.M.R. Iravani, Optimal dynamic assignment of a flexible worker on an open production line with specialists, Eur J Oper Res 170 (2006), 541–566.
[24] R. Uzsoy and L.A. Martin-Vega, Modelling Kanban-based demand-pull systems: A survey and critique, Proc Manuf Inter '90, Atlanta, GA, March, 1990.
[25] M.P. Van Oyen, E.S. Gel, and W.J. Hopp, Performance opportunity of workforce agility in collaborative and noncollaborative work systems, IIE Trans 33 (2001), 761–777.
[26] E. Zavadlav, J.O. McClain, and L.J. Thomas, Self-buffering, self-balancing, self-flushing production lines, Management Sci 42 (1996), 1151–1164.