CLARIFYING CONWIP VERSUS PUSH SYSTEM BEHAVIOR USING SIMULATION

S. T. Enns
Paul Rogers

Dept. of Mechanical and Manufacturing Engineering
University of Calgary
Calgary, AB., T2N-1N4, CANADA

ABSTRACT

This research examines the performance of CONWIP versus "push" workload control in a simple, balanced manufacturing flowline. Analytical models and simulation experiments are used to evaluate the tradeoffs between throughput and inventory performance. Tradeoff curves based on inflating the inventory level for the CONWIP system and the arrival rate for the "push" system are generated. As well, the variability of interarrival and processing times are considered as experimental factors. Results show that, contrary to what some previous studies have indicated, CONWIP efficiency is not inherently superior to "push" system efficiency. Instead, the release mechanism used for the "push" system has a significant impact on which system will perform better. Utilization levels and processing time variability also affect the relative performances.

1 INTRODUCTION

The constant-work-in-process, or CONWIP, system of workload control has received significant attention since being introduced by Spearman et al. (1990). While the popularity of CONWIP has increased since this study, the basic ideas behind maintaining constant workloads have been around for many decades. This is pointed out in an excellent review by Framinan et al. (2003). As well, similar ideas have long been implemented through the exercise of input control in manufacturing practice. Good references for research on input control include Wight (1970) and Bergamaschi et al. (1997).

The CONWIP system has some appealing characteristics but also some significant limitations with respect to implementation. Among CONWIP's strengths are its simplicity and its robustness in dealing with processing time uncertainty or floating bottlenecks. Limitations include the difficulty of coordinating parts going into assemblies and the difficulty of treating transfer and processing lot sizes differently from stage to stage. Most often CONWIP systems are compared to Kanban systems, since both are "pull" systems. It would appear that Kanban systems, especially the two-card type, are more suitable for complicated environments.

One of the problems in comparing the performance of a CONWIP system with any other system is the lack of a unified framework for comparison (Framinan et al. 2003). This leads to results that are difficult to generalize and may even be contradictory. For example, several studies have concluded that a CONWIP system will outperform a Kanban system in a flowline structure. However, Gstettner and Kuhn (1996) reached the opposite conclusion.

This research addresses the comparison of CONWIP with another type of control system. In particular, Hopp and Spearman (2000) concluded that CONWIP will dominate a "push" workload control system in a simple flowline. More specifically, it was concluded that a CONWIP system will yield more throughput for a given average inventory level. However, their analysis was based on queuing relationships that may not be suitable.

In this research CONWIP and "push" systems are compared experimentally. The purpose is not only to verify whether performance using CONWIP is superior but also to gain insights into the behavior of these two control systems over a range of conditions. Therefore, a flexible simulation model was constructed to compare these systems. The modeling considerations used were similar to those discussed by Marek et al. (2001). Analysis was then performed considering both analytical relationships and experimental results.

The next section describes the flowline model used in this research. Following this, an initial comparison is made based on Hopp and Spearman's example. A larger experimental design is then presented, followed by analysis of more comprehensive results.

2 THE FLOWLINE MODEL

This research is based on a flowline model constructed using ARENA® discrete-event simulation software (Kelton et al. 2004). The flowline consists of five machines ($N=5$). All jobs visit all five machines sequentially so flow is uni-

directional. The jobs are independent and any setup time is assumed to be included in the processing time. When jobs are completed they leave the system. In other words, demand is assumed infinite and there is no storage of finished goods. These assumptions are consistent with those used in other studies, such as Hopp and Spearman (2000) or Gstettner and Kuhn (1996). Figure 1 illustrates the configuration assumed.
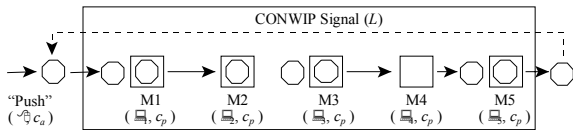

Figure 1: Flowline Configuration

The processing times are stochastic and specified by the machine processing rate, μ, and the processing time coefficient of variation, c_p, which is defined to be the standard deviation of the processing times divided by the mean processing time, $1/\mu$. The distribution used to generate appropriate processing times is the Gamma distribution, which is defined by parameters α and β. The mean of the distribution is equal to $\alpha\beta$ and the variance is equal to $\alpha\beta^2$. It can easily be shown that the value of α is equal to $1/(c_p^2)$ and the value of β is equal to c_p^2/μ. When c_p values are set equal to 1.0, the processing times are negative exponential.

In this research the processing rates, μ, are set equal to 1.0 at each machine. This means the line is balanced. Although no bottlenecks are assumed, the simulation model itself has the flexibility to incorporate bottlenecks.

Performance evaluation is based on throughput rates and inventory levels. Throughput rates, λ, are equal to the arrival rate of jobs into the flowline while inventory levels are based on the number of jobs in the system, L. It is desirable to have high throughput rates with low inventories but there is obviously a tradeoff between these two measures. Therefore, tradeoff curves are generated for the purpose of determining dominance of one system over another.

2.1 CONWIP Control

CONWIP performance is determined by the number of jobs allowed in the system, L. It is assumed there is never a shortage of jobs (orders) available for release into the shop. Therefore, the number of jobs in the system is always constant. When a job is completed and exits the flowline, a signal is sent to release a new job into the flowline. This signal could be sent through the use of cards. However, in this research the signal transmission is assumed to be instantaneous. Therefore the distribution of job interdeparture times from the flowline is the same as the distribution of job interarrival times into the flowline.

In the case of CONWIP systems the throughput level, λ, for a given inventory level, L, is the main performance measure.

2.2 "Push" Control

For the "push" system there is assumed to be no limit on the inventory within the flowline. Throughput is determined by specifying the arrival rate of jobs to the flowline, λ. As well, the variability of job interarrival times can be controlled by specifying the interarrival time coefficient of variation, c_a.

The distribution used to generate appropriate interarrival times is again the Gamma distribution. This distribution generates interarrival times which are independent. Procedures for setting the α and β parameters are similar to those described for setting the processing time distribution parameters. When the c_a value is specified as 1.0, the interarrival times are negative exponential. In other words, job arrivals into the system are then based on a Poisson process.

In the case of "push" systems the time-average inventory level, L, for a given throughput, λ, is the main performance measure.

3 AN INITIAL EXAMPLE

Assume that a flowline consists of 5 machines (N=5) and that processing times are negative exponential with a mean processing time, $1/\mu$, of 1.0 at each machine. This would be considered a balanced line. Furthermore, assume that the CONWIP system is being operated with a work-in-process level, L, of 6.

3.1 An Analytical Analysis

According to Hopp and Spearman (2000) the performance of this system can be estimated by viewing the behavior along the flowline as if it is memoryless. A new job arrival will expect to see L-1 jobs spread evenly across the N machines in the line. Therefore the expected flowtime along the line would be as follows.

$$W = N\left[\left(\frac{L-1}{N}\right)\frac{1}{\mu} + \frac{1}{\mu}\right] \qquad (1)$$

The first term in the square brackets is the expected processing time remaining for jobs already at any machine when a new job arrives. This plus the expected processing time for the new job, $1/\mu$, yields the expected delay at any machine along the flowline. Multiplying this by the number of machines then gives the expected job flowtime, W. The throughput can then be determined by using Little's Law, $\lambda = L/W$. Substituting Equation (1) for W and simplifying leads to the following.

$$\lambda = \frac{L\mu}{L-1+N} \qquad (2)$$

This equation is meant to apply under only memoryless conditions

Since μ is assumed to be 1.0, the throughput, λ, for the example is then given by the following.

$$\lambda = \frac{L}{L-1+N} = \frac{6}{6-1+5} = 0.6 \qquad (3)$$

Therefore, the analytically estimated throughput rate, using a constant work-in-process level of 6, will the 0.6. Using Little's Law, the expected flowtime, W, is 10. Finally, the utilization level at each of the machines, ρ, is also equal to 0.6. This can be confirmed by noting that $\rho = \lambda / \mu$ for any system.

For comparison a flowline with unconstrained inventory may also be considered. In this case, it is assumed there is some external release mechanism that is controlling the job arrival stream. Interarrival times are independent of the number of jobs currently within the flowline. Hopp and Spearman (2000) refer to this as a "push" system.

If the arrivals to this "push" system are described by a Poisson process (negative exponential interarrival times) and if the processing times at each of the N machines are again negative exponential, this system can be analytically modeled using *M/M/1* queuing relationships. In *M/M/1* queues the departure stream is also described by a Poisson process. Therefore the behavior of each station in a series can be analyzed independently and the results simply summed to determine the average flowtime, W, or time-average inventory, L, for the flowline.

The time-average average inventory, L, expected for a given throughput rate, λ, when N machines are linked in series can be determined as follows.

$$L = N\left(\frac{\lambda}{\mu - \lambda}\right) = N\left(\frac{\rho}{1-\rho}\right) \qquad (4)$$

If the throughput rate is set equal to that in the previous CONWIP case, the expected average inventory level for the "push" system is determined as follows.

$$L = 5\left(\frac{0.6}{1-0.6}\right) = 7.5 \qquad (5)$$

The expected flowtime through the system, W, would be the following.

$$W = 5\left(\frac{1}{\mu - \lambda}\right) = 5\left(\frac{1}{1-0.6}\right) = 5(2.5) = 12.5 \qquad (6)$$

Since an average inventory level of 7.5 is required to get a throughput rate of 0.6 for the "push" system, versus an inventory level of 6.0 for the CONWIP system, Hopp and Spearman conclude the CONWIP system is more efficient.

3.2 An Experimental Analysis

A set of simulation experiments was initially run to illustrate the problem of comparing CONWIP and "push" systems using analytical relationships. The first experiment was conducted assuming a "push" system with $\lambda = 0.6$ and a Poisson arrival process. Five replications were run. Each involved an initialization period of 1000 time units and data collection over 100,000 time units. The flowline was preloaded with one job at each machine to facilitate reaching steady-state conditions faster. The observed time-average jobs in the system, L, was 7.465, with a 95% confidence interval of (7.331, 7.598). The average flowtime, W, was 12.384, with a 95% confidence interval of (12.234, 12.534). The theoretical values previously calculated fall within these confidence intervals. The time-average standard deviation of the jobs in the system was 4.305.

A second experiment was conducted assuming a CONWIP system with $L = 6$. Procedures for data collection were identical to previous. As well, common random numbers were used. In this case the observed throughput rate, λ, was 0.605 and the average flowtime, W, was 9.925, with a 95% confidence interval of (9.888, 9.962). The average inventory, L, was confirmed to be 6.00 with a standard deviation of 0.0. The mean interarrival time coefficient of variation for new jobs, c_a, was measured to be 0.937. The fact that c_a and W are lower than the analytical values while λ is higher implies the assumption of the system being memoryless may not be entirely accurate. In other words, if the system were memoryless the expected c_a value should be 1.0, as for *M/M/1* queuing relationships.

The slightly higher throughput rate along with lower inventory illustrates that the performance of the CONWIP system is superior for this example. These results are consistent with those of Hopp and Spearman. However, these comparisons are somewhat problematic with respect to the job input stream. In the case of the CONWIP model it is implicitly assumed that there is always a new job available to enter the system. This means there must be some sort of job (or order) pool but this is not being taken into account in the analysis. However, for the "push" system it is assumed jobs directly enter the flowline based on some external release process.

If in fact there are always jobs (orders) available and the objective is to minimize inventory at a specified throughput level it would not be reasonable to release jobs into the flowline using a Poisson process. A more reasonable alternative would be to release jobs at a constant rate. Therefore, assuming a "push " system with constant interarrival times for comparison with a CONWIP system is more fair.

A simulation with constant interarrival times of 1.667 ($1/\lambda = 1/0.6$) resulted in observed time-average jobs in the system, L, of 5.675, with a 95% confidence interval of (5.644, 5.707). The average flowtime, W, was 9.459, with a

95% confidence interval of (9.407, 9.512). The time-average standard deviation of the jobs in the system was 2.057. These results indicate that this "push" system with constant interarrival times requires less work-in-process inventory, L, while yielding the same output as the CONWIP system. It is therefore clear that CONWIP systems do not always dominate in terms of throughput versus inventory performance efficiency.

In order to obtain more insight into the behavior of these systems, performance tradeoff curves can be developed using structured simulation experiments. The following section describes experiments designed to support this type of analysis.

4 EXPERIMENTAL DESIGN

The flowline, processing time, measurement and data collection assumptions used in these experiments are the same as previous unless otherwise specifically stated. For the CONWIP system, tradeoff curves of throughput versus inventory performance were made by incrementing the inventory level, L. The range used was from 4 to 21 jobs in the system. In other words, 18 levels were used to obtain points along each tradeoff curve. The only other factor in the CONWIP experiments was the coefficient of variation of the machine processing times, c_p. These were set at two levels; namely 1.0 and 0.5. Each combination of CONWIP setting was run for three replications. Therefore, the total number of runs using CONWIP was 108 (18*2*3).

For the "push" system, tradeoff curves were made by incrementing the throughput rate, λ. The range used was from 0.45 to 0.85 jobs per time unit, in steps of 0.05. In other words, 9 levels were used to obtain points along each tradeoff curve. Two other factors were also considered. The first was again the coefficient of variation of the machine processing times, c_p. These were set at two levels; namely 1.0 and 0.5.

The second additional factor was the coefficient of variation of job interarrival times, c_a. The c_a values could be due to choice if a pool of available jobs (orders) is assumed, or it could be due to some external order release mechanism. For example, orders could be generated from an MRP system. In practice these order releases would not likely be Poisson but would rather have some lower level of variability. Rapid modeling software, such as MPX, assumes a default c_a value of 0.30 (MPX, 1991). The factor levels of c_a chosen for the experimental design were 1.0, 0.5 and 0.0. Each combination of "push" settings was also run for three replications. Therefore, the total number of runs using "push" workload control was 162 (9*2*3*3).

5 EXPERIMENTAL RESULTS AND ANALYSIS

This section first compares CONWIP and "push" system performance. Further analysis of CONWIP system behavior is then presented to gain insights into the circumstances under which the CONWIP system is likely to outperform the "push" system.

5.1 CONWIP and "Push" Comparisons

The "push" system tradeoff curves for throughput, λ, versus inventory, L, with c_p values of 1.0 are shown in Figure 2 as dashed lines. Separate curves are shown for c_a values of 1.0, 0.5 and 0.0. The CONWIP tradeoff curve for c_p values of 1.0 is shown as a solid line.
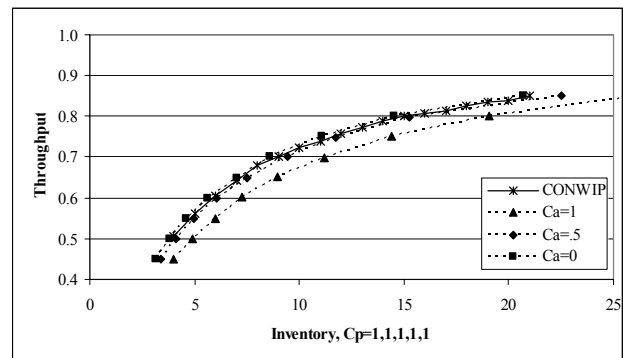

Figure 2: Comparison with $c_p = 1.0$

Similarly, Figure 3 illustrates the results when a c_p value of 0.5 is used at each machine. These results are consistent with those in Figure 2. The main difference is that the complete set of curves is shifted up and to the left. As expected, this indicates improved performance with lower processing time variability.
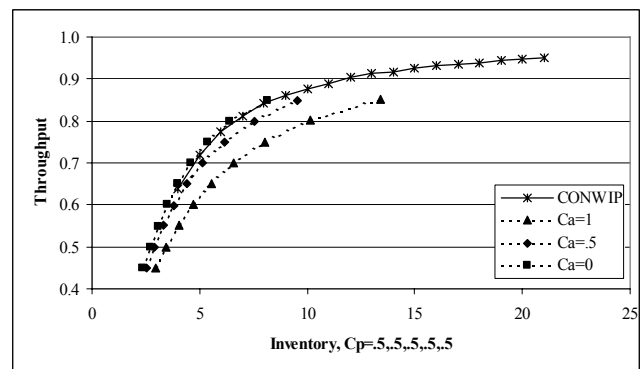

Figure 3: Comparison with $c_p = 0.5$

The "push" system results in Figures 2 and 3 show that for any level of throughput, inventory increases with increasing interarrival time variability. As well, for any given level of throughput, inventory increases with increasing processing time variability. The CONWIP results show that the throughput, λ, increases with increasing inventory levels, L. As well, throughput decreases with increasing

processing time variability. These results are all consistent with expectations based on queuing relationships.

These results show that a CONWIP system results in better performance than a "push" system when interarrival times to the "push" system are highly variable. In other words, when the new job arrivals are Poisson (c_a=1.0) a "push" system performs relatively poorly. However a "push" system, which has no inventory constraints, can perform slightly better than a CONWIP system if jobs are released at constant intervals.

In summary, these comparisons indicate that a CONWIP system is not inherently superior to a "push" system and that the relative performance is conditional on the interarrival time variability assumed for the "push" system. As well, it is dependent on the interarrival time variability observed in the CONWIP system. This in turn is dependent on the utilization levels and processing time variability, especially at the last machine in the flowline.

5.2   CONWIP Interarrival Time Analysis

It is obvious that for the "push" system the job interarival times are independent of the job processing time characteristics. This is not true for the CONWIP system. This subsection takes a closer look at the dependence of the interarrival time variability on processing time variability in CONWIP systems. In other words, the dependence of c_a on c_p is examined. Figure 4 shows the throughput performance as a function of the inventory levels, for both c_p values of 1.0 and 0.5, as solid lines. The dashed lines, along with the right-hand axis, show the observed interarrival time coefficients of variation, c_a, for new jobs entering the system.
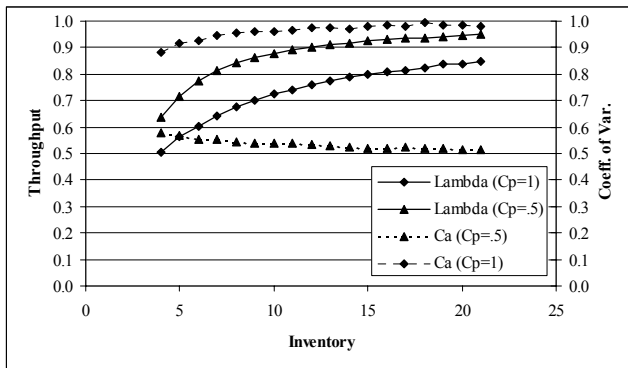

Figure 4: CONWIP results

When CONWIP inventory levels are kept low, the c_a values are well below 1.0 even when processing times are negative exponential. This illustrates why it is problematic to use the assumption of memoryless properties for the analysis, even if the processing time coefficients of varia-

tion, c_p, are set equal to 1.0. The assumption of negative exponential interarrival times ($c_a = 1.0$) is being violated. However, if c_p values are 1.0 and inventory levels are allowed to increase, the interarrival time coefficient of variation, c_a, will approach 1.0 and analysis based on memoryless properties will provide a better approximation.

Analytically, this behavior can be explained by considering Marshall's formula for interdeparture time coefficients of variation, c_d. This formula is exact for independent interarrival times. (Whitt, 1983).

$$c_d^2 = \rho^2 c_p^2 + \left(1 - \rho^2\right)c_a^2 \qquad (7)$$

where ρ is the utilization level, equal to λ / μ. As the machine utilizations increase, the variability of interdeparture times will be increasingly dictated by the processing time variability. Therefore, as the CONWIP inventory levels increase, ρ will also increase and the interdeparture time distribution will increasingly approach the processing time distribution.

Since the departures from the last machine are equivalent to the arrivals at the first machine under CONWIP control, it is easy to see why the c_a value will approach the c_p value for the last machine as utilization levels increase. If the processing times at the last stage are highly variable, the interdeparture time coefficient of variation, c_d, will be high. This in turn will cause new job interarrival times to be highly variable, negatively affecting upstream flowtimes. Performance for the flowline will deteriorate as a result.

One additional set of experiments was run in which the processing times at the last stage were set to be more variable than those at other upstream stages. Specifically, the processing time coefficient of variation, c_p, for the last machine was set to 1.0 while all other machines were set to 0.5. Under these conditions a relative improvement in the "push" system performance compared to the CONWIP system would be expected.

Figure 5 shows the results for the CONWIP system, as well as the "push" system at three levels of interarrival time variability. The behavior anticipated is confirmed. The "push" system performance improves relative to the CONWIP system performance when the c_a values are low. In other words, the interarrival time variability at upstream stages is relatively low and this helps reduce flowtimes as well as inventory. With a c_a value of 0.0 it is obvious that the "push" system performance is better than that of the CONWIP system.
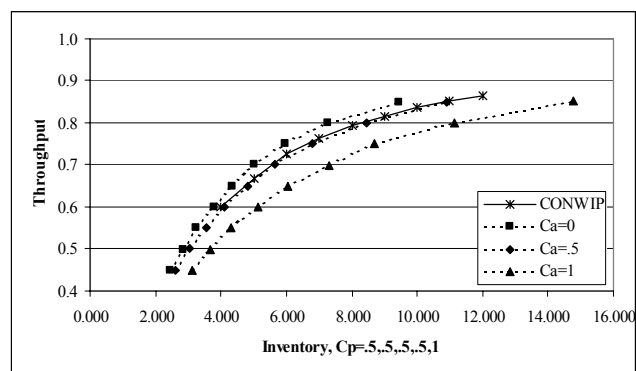
Figure 5: Comparison with c_p high for last stage

6   CONCLUSIONS

This paper contributes an examination of the underlying behavior of CONWIP versus "push" workload control in a simple flowline. Preliminary results indicate that comparison is not straightforward and that it may not be possible to conclude that either system dominates, as has been suggested in some previous research. Further study is necessary to fully understand the relative behavior of these two types of workload control systems.

ACKNOWLEDGMENTS

REFERENCES

Bergamaschi, D., R. Cigolini, M. Perona, and A. Portoli. 1997. Order review and release strategies in a job shop environment: a review and a classification. *International Journal of Production Research* 35:399-420.

Framinan, J. M., P. L. González, and R. Ruiz-Usano. 2003. The CONWIP production control system: review and research issues. *Production Planning and Control* 14: 255-265.

Gstettner, S., and H. Kuhn. 1996. Analysis of production control systems kanban and CONWIP. *International Journal of Production Research* 34:3253-3273.

Hopp, W. J., and M. L. Spearman. 2000. *Factory Physics: Foundations of Manufacturing Management*. 2nd ed. New York, NY: McGraw-Hill, Inc.

Kelton, W.D., R.P. Sadowski, and D.T. Sturrock. 2004. *Simulation with Arena*, 3rd ed. New York, NY: McGraw-Hill.

Marek, R. P., D. A.Elkins, and D. R. Smith. 2001. Understanding the fundamentals of kanban and CONWIP pull systems using simulation. In *Proceeding of the 1994 Winter Simulation Conference*, ed. B. A. Peters, J. S. Smith, D. J Medeiros, and M. W. Rohrer, 921–929. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

MPX User Manual. 1991. Burlington, MA: Network Dynamics.

Spearman, M. L., D. L. Woodruff, and W. J. Hopp. 1990. CONWIP: a pull alternative to kanban. *Production Planning and Control*, 28: 879-894.

Whitt, W. 1983. The queuing network analyzer. *The Bell Systems Technical Journal* 62:2779-2815.

Wight, O. 1970. Input/output control a real handle on lead time. *Production and Inventory Management*, 11:9-31.

AUTHOR BIOGRAPHY

SILVANUS T. ENNS is an Associate Professor at the University of Calgary, Canada. His research interests lie in job shop, batch production and supply chain modeling and analysis.

PAUL ROGERS is an Associate Professor in the Department of Mechanical and Manufacturing Engineering at the University of Calgary. His research and teaching interests include production planning and control systems, manufacturing strategy, distributed and dynamic scheduling, and models for the design and analysis of healthcare systems. He is a Professional Engineer registered with the Association of Professional Engineers, Geologists, and Geophysicists of Alberta (APEGGA), a member of IIE, APICS, and INFORMS, and serves on the Editorial Board of the International Journal of Computer Integrated Manufacturing. He holds Ph.D. and M.Eng. degrees from Cambridge University in England. He can be contacted by email at <rogers@ucalgary.ca>.