Designing Productive Manufacturing Systems without Buffers

T. Freiheit*, M. Shpitalni** (1), S. J. Hu*, Y. Koren* (1)
*Department of Mechanical Engineering, The University of Michigan, U.S.A.
**Faculty of Mechanical Engineering, Technion, Haifa, Israel

Abstract
Modern industrial practice is to minimize work-in-process in order to eliminate inventory-carrying costs and quickly detect quality problems. Reduced work-in-process results from eliminating in-process buffers between operations in serial lines, but is accompanied by decreased system efficiency. Inventories are created before system expansion in order to offset production lost during construction. Furthermore, serial line expansion implies doubling line output. In reconfigurable manufacturing systems, new configurations that have not yet been fully explored by industry can be used to compensate for loss of buffered system isolation failure, creation of inventories, and step-size production expansion. Numerical models are applied to predict productivity and explicitly show the equivalency of alternative configurations to buffered serial transfer lines. Parallel-serial configurations as well as the newly proposed reserve capacity configurations are examined.

Keywords:
Productivity, Manufacturing System Performance, Reconfigurable

1 INTRODUCTION

Pure serial configuration, as schematically illustrated in Figure 1(a), is traditionally used in automated transfer lines. Its advantages include relatively low work-in-process and throughput time, as compared to process layout type configurations. High productivity can be achieved when buffers are utilized to isolate system failure from individual machine failure. However, buffered serial lines are not necessarily the best production strategy for addressing the current manufacturing environment. The unpredictable market changes of recent years have led to the need for an environment that can react rapidly and cost effectively [1].

Modern industrial practice is to minimize the amount of work-in-process in order to eliminate inventory-carrying costs and facilitate detection and reduction of quality-associated rework costs. Although the elimination of buffers achieves this goal, it also greatly reduces system productivity, since a single machine failure causes system failure. In addition, undesirable inventories of finished goods are typically held to buffer the customer from system failure.

Globalization and international competition have made it essential for manufacturing operations to produce efficiently in response to market demand. Historically, manufacturing systems have been designed with excess capacity in the anticipation that demand will eventually reach capacity. However, market projections are notoriously inaccurate, resulting in a significant risk that the eventual demand will not equal the planned capacity. When capacity exceeds demand, capital has been inefficiently invested. When capacity is under demand, it is normal that capacity be expanded.

Typically, inventories are created before a manufacturing system is expanded to compensate for disruptions during construction. In a growing market, it may be difficult to accumulate these inventories. Moreover, expansion of a serial production line implies duplicating the line, thus doubling its output, which again can yield a capacity higher than the ultimate demand.

Manufacturing system performance is influenced by system configuration [2]. In reconfigurable manufacturing systems, non-traditional configurations other than serial lines can be exploited to provide low levels of work-in-process, reduce inventories, and provide incremental production expansion step sizes. Productivity improvement can be achieved by parallelism rather than through the use of buffers in a serial line. Moreover, with parallelism and flexible production equipment, the throughput of a production system can be incrementally expanded. The parallel-serial and the newly proposed reserve capacity classes of production system configurations are examples of such configurations.

This paper applies numerical models to predict the productivity of non-traditional manufacturing system configurations and to demonstrate their equivalency to buffered serial transfer lines. The paper is organized as follows: Sections 2 describes the productivity variance reduction of parallel-serial lines and their productivity equivalency to buffered serial transfer lines. Section 3 shows how production lines may be scaled up incrementally by using reserve capacity. Section 4 discusses the configurations and concludes the paper.

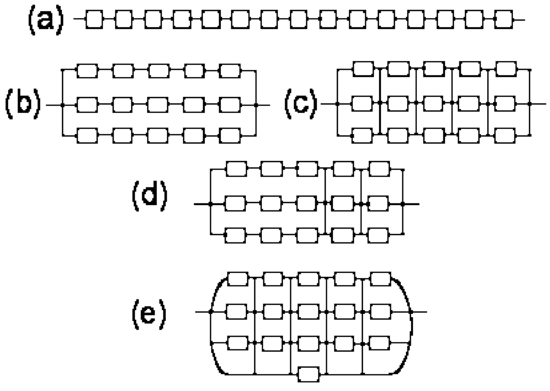

Figure 1: Schematics of Production System Configurations without Buffers.

2 CONFIGURATION AFFECT ON PRODUCTIVITY VARIANCE AND BUFFER EQUIVALENCY

The parallel-serial class of production system configurations is defined as a set of *m* serial machining lines of *n* stages, configured in parallel to each other [3]. The serial lines can either be completely independent of each other or can have crossover between every operation. Crossover refers to a situation in which product flow may be transferred to another line between stages, e.g. when a line is blocked due to a failure upstream. Parallel-serial systems may also be hybrid, where crossover points do not occur following every stage. A hybrid configuration may be desirable when quality considerations dictate consistency between operations. In this model, machines and stages are assumed to be independent and have paced production rates. Figures 1(b) and 1(c) show example schematics of parallel-serial configurations of *m*=3 and *n*=5 without and with crossover, respectively. Figure 1(d) shows a schematic of a parallel-serial hybrid configuration.

Unlike bufferless pure serial configurations, these non-traditional system configurations can still produce when one or more machines fail. Thus, in order to predict their productivity, it is necessary to model all of their productive system states. The productivity of a configuration is defined as the normalized expectation of the production rate of all system states. Productivity is also called system availability or effectiveness. Mathematically, productivity *P* can be expressed as:

$$P = \frac{1}{\mu_{\max}} \sum_{i=1}^{n} \mu_i \Pr(ith\ state) \tag{1}$$

where *n* is the number of system states, μ_i is either zero for a non-productive state or the production rate associated with the *i*th state, μ_{max} is the highest production rate of the states, and Pr*(ith state)* is the probability that the *i*th state occurs. From basic statistics, the productivity variance is:

$$\sigma^2 = \frac{1}{\mu_{\max}^2} \left(\sum_{i=1}^{n} \mu_i^2 \Pr(ith\ state) - \left(\sum_{i=1}^{n} \mu_i \Pr(ith\ state) \right)^2 \right) \tag{2}$$

The productivity of a parallel-serial configuration with no crossover and with all machines having the same availability is determined by summing the probabilities of all permutations of operational lines, where ($1-R^n$) is the probability of any line being down, and *R* is the probability that a single machine is functional. This productivity is:

$$\begin{aligned} P_{\parallel-serial\ no\ x/o} &= \sum_{i=1}^{m} \binom{m}{i} (1-R^n)^{m-i} R^{ni} \frac{i}{m} \\ &= R^n (1-R^n)^m \left(1 + \frac{R^n}{1-R^n}\right)^m \\ &= R^n \end{aligned} \tag{3}$$

Therefore, Equation (3) shows that the productivity of a parallel-serial production system with no crossover is independent of the number of lines and, like the pure serial line, entirely dependent on the number of stages (machines). That is, a system with *m* parallel-serial lines of *n* machines per line and no crossover having a theoretical production capacity of *P* (*P*/*m* per line) has an equivalent production rate to a pure serial line (*m*=1) of *n* machines with production capacity *P*. Moreover, redistributing the *n* machines of a pure serial line into a parallel-serial configuration of *m* lines without crossover shortens each line, giving a productivity improvement of 100($R^{n/m-n}$-1)%, where *n* is evenly divisible by *m*. Furthermore, the additional lines reduce the variance of productivity over that of a pure serial line.

The productivity of a parallel-serial configuration with crossover is determined from the number of permutations of functioning machines in each stage. When all stages have the same availability and production rate, the productivity is given by:

$$P_{\parallel-serial\ with\ x/o} =$$

$$\frac{1}{m}(1-R)^{nm} \sum_{b=1}^{m} \sum_{a_1=b}^{m} \cdots \sum_{a_n=b}^{m} \binom{m}{a_1} \cdots \binom{m}{a_n} \left(\frac{R}{1-R}\right)^{q}$$

$$\text{with} \quad q = \sum_{j=1}^{n} a_j \tag{4}$$

The summation with index *b* is included to account for the bottlenecks that occur due to one or more machine failures in a stage. A comparison of the productivity advantage of crossover over no crossover is given in Freiheit et. al [3].

As noted above, one advantage of the parallel-serial type configuration is that it has reduced productivity variance. Pure serial lines without buffers are limited to a single productive system state, giving a variance based on the probabilities either of no machine failures or of one or more machine failures. Parallel-serial lines have multiple productive states, where productivity is scaled by the equivalent number of lines that are functioning. For example, in a parallel-serial line with crossover, the equivalent number of lines is the number of functioning machines in the bottleneck stage, i.e. the minimum of the a_i's of Equation (4).

Figure 2 plots the productivity coefficient of variation (CV) of parallel-serial lines both with and without crossover. This plot shows that more lines in parallel yield a lower CV. Additionally, more line in parallel result in a lower increase in CV as the lines get longer. Finally, crossover configurations have inherently less variance than non-crossover configurations. Since consistency in production levels is desirable in a manufacturing system, a parallel-serial production line design should take advantage of these trends and develop the process plans and resource requirements for wider lines.

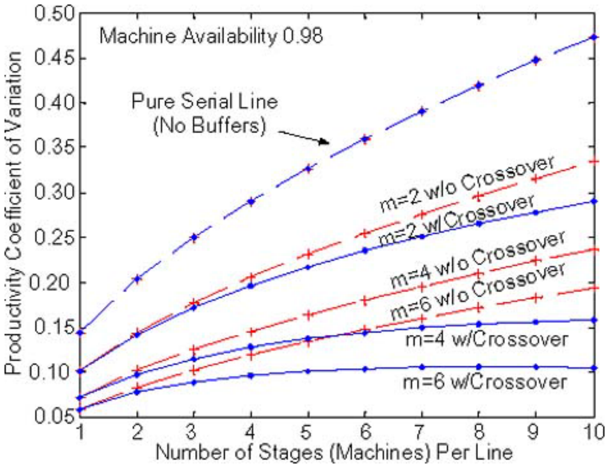

Figure 2: Productivity Coefficient of Variation Versus Line Length.

Configuration		Productivity	Pure Serial w/Buffers		Productivity	Pure Serial w/Buffers	
Width (m)	Length (n)	No cross-over	Total Buffers	Total Avg. WIP	With Cross-over	Total Buffers	Total Avg. WIP
24	1	–	–	–	0.950	Inf	–
8	3	0.855	16	8	0.891	595	297
6	4	0.813	18	9	0.862	367	183
4	6	0.733	20	10	0.807	191	95
3	8	0.663	21	10	0.756	118	59
2	12	0.548	22	11	0.665	56	28
1	24	0.346	0	12	–	–	–

Table 1: Productivity Equivalence of Parallelism to Optimal Buffers in a Pure Serial Line of 24 Machines.

The improvement in productivity of parallel-serial type configurations over pure serial lines can be equated to the productivity improvements achieved in pure serial lines by the addition of buffers. The buffer size necessary in a pure serial line for the same productivity given by a parallel-serial configuration can be calculated using an aggregation/decomposition algorithm per Gershwin [4] and Dallery et al. [5]. This algorithm, incorporated by Yang [6] into a prototype software program, calculates an optimum buffer allocation to achieve a given level of productivity in a pure serial line and its average work-in-process. Required inputs are production rate, mean-time-to-failure and mean-time-to-repair of each machine in the line. The program outputs the total size and location of buffers to meet the desired productivity, as well as the average work-in-process in every buffer.

In a system of 24 fully flexible reconfigurable machines, it is possible to create twelve parallel-serial configurations, as detailed in Table 1. It is assumed that it is possible to reallocate the machine tasks such that the effective system production rate is the same in all configurations. Table 1 presents the productivity for these configurations and compares it to the productivity of a pure serial line with an optimal allocation of buffers to provide the same level of productivity. All machines have the same availability of 0.95. As can be seen, a pure serial line requires more buffer space to achieve equivalent productivity levels when more lines are placed in parallel, and the average work-in-process likewise increases.

Note also that crossover in a parallel-serial line is approximately equivalent to having a buffer of size one between every stage of a non-crossover line. For example, in a line of width eight and length three, a buffer of size one accounts for the productivity difference of 0.036. Since there are eight lines in parallel, crossover in this case is equivalent to a total system buffer size requirement of 16, with an average WIP of 8. In addition, note that the reserve capacity configuration of the next section also provides productivity improvements that are equivalent to a buffer.

3 PRODUCTIVITY AND LINE CAPACITY

Ideally, a production system is initially designed with changes in capacity in mind. One typical method in a serial line is to leave open stages close to anticipated bottleneck points to provide space for production expansion. However, depending on the provisions made, the line must be shut down during construction, implying a build-up of inventory prior to construction. Furthermore, large-scale capacity changes in serial lines imply duplicating a line, thus doubling system output. The open stages method is less applicable for large-scale changes because bottleneck points may not be adequately anticipated. Moreover, leaving stages open for expansion normally does nothing for reducing capacity.

The proposed reserve capacity configuration provides a means for scaling production up or down, using adjustments in productivity as well as modifications to capacity to address current throughput requirements. The capacity change step size is typically small enough to provide a better match of system throughput to product demand. Additionally, in the reconfigurable manufacturing paradigm, production modules can be reused, so production assets do not need to be wasted with capacity changes. The idea is to add or remove productive standby machines in parallel to a main production line by providing one or more flexible standby machines capable of performing any operation in parallel to a serial or parallel-serial transfer line, e.g. Figure 1(e).

The standby machines can either contribute independently to production when they are not being called upon to substitute for a failed machine in the transfer line, or they can remain idle until a failure occurs in the main transfer line. Note that a parallel-serial line with reserve capacity has crossover by necessity. Consideration of reserve capacity is facilitated by the increasing use of identical, highly flexible CNC machines in production system design.

For a paced parallel-serial system of identical machine availability and with k standby machines, reserve capacity productivity is derived by determining the equivalent number of parallel lines of n operational stages and noting that every excess functional standby machines contribute $1/n$ to productivity:

$$P_{rc} = \sum_{i=0}^{k} \sum_{a_1=0}^{m} \cdots \sum_{a_n=0}^{m} \left[\frac{\begin{array}{l} \binom{m}{a_1} \cdots \binom{m}{a_n} (1-R)^{nm-q} R^{q} \times \\ \left(\binom{k}{i} R_s^{i} (1-R_s)^{k-i} \right) \times \\ \left(u_s - (1-\alpha)\gamma + \dfrac{\beta(i-\gamma)}{n} \right) \end{array}}{m + k/n} \right] \tag{5}$$

$$q = \sum_{j=1}^{n} a_j$$

$$u_s = \min_{t=0}^{i}\left(a_1, \ldots, a_{\beta(t)} + 1, \ldots, a_n\right)$$

$$\gamma = n u_s - \sum_{i=1}^{n} \min\left(a_i, u_s\right)$$

where:

- n is the number of stages in the main line.
- a_i is the number of functioning machines in the ith stage.
- q is the total number of functional machines in the main line.
- i is the number of functional standby machines.
- R is the availability of main line machines.
- R_S is the availability of the standby machines.

- $u_{\blacksquare}$ is the equivalent number of parallel lines.
- γ is the number of standby machines being used as main line machine replacements.
- α is a productivity-discounting factor associated with transfer efficiency to the standby machines.
- β is an efficiency-discounting factor for the multi-functional standby machine in full production mode.

Alpha and beta equal one when there is complete efficiency, e.g. when transfer can be completed within the main line cycle time. As appropriate, α can be interpreted as a scaling factor for slowing the pace of production. Note that in the calculation of $u_{\blacksquare}$, $b^{(j)}$ is the index of the jth minimization operation and $b^{(0)}=\min(a_1,\ldots,a_n)$. Models have also been developed for pure serial and non-productive standby machine reserve capacity configurations.

The expansion of a line by the addition of productive standby machines to a parallel-serial main line is almost equivalent to placing all the standby machines in a pure parallel configuration in parallel to the line. However, the standby machines do not truly constitute a pure parallel configuration because their operation is dependent on the main line, as they can be called upon to substitute for a failed main line machine. More productivity is gained from the standby machines in the reserve capacity configuration than in a pure parallel configuration because of synergies derived from bottleneck mediation.

Figure 3 plots an adapted version of Equation (5), giving the relative throughput as a line is expanded by adding standby machines. When the number of standby machines equals the length of the line, the standby machines are integrated into the main line, resulting in a parallel-serial configuration. This integration results in a small loss of throughput, as the standby machines no longer operate like a pure parallel configuration. However, that loss is reversed with the next addition of a standby machine. Note that when sequential stages fail in the main line, multiple standby machines can be operated serially. However, this operating scheme is not modeled in Equation (5), and operating a pure parallel configuration is generally more productive.

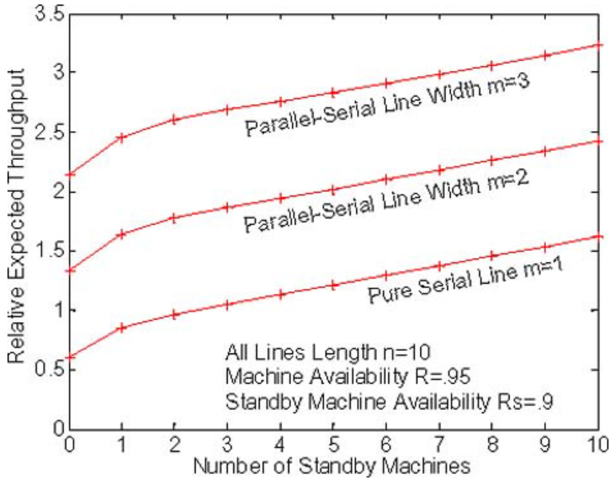

Figure 3: Production Expansion Using Reserve Capacity. Alpha =0.7 Beta =0.9.

4 DISCUSSION AND CONCLUSIONS

The design of productive manufacturing systems without buffers can be achieved by judicious use of non-traditional configurations such as the parallel-serial and reserve capacity types. A few additional issues in the use of configurations should be noted. Reliable material handling is necessary to maintain the synergistic effect of crossover in both configuration types. The availability of the material handling system should exceed the availability of the machines. If the material-handling system is not sufficiently reliable, a buffered system may be preferable. Machine availability has a large impact on the synergistic gains of crossover. When machine availability is high, crossover may not be used frequently enough to justify the additional expense of its material-handling system. Finally, if main line machine failures are frequent and have low repair time, it is possible that no productivity gains can be had from a reserve capacity configuration, as a main line machine may be repaired by the time a standby machine has been configured to perform the operation that it is substituting. In this case, a buffered system is preferable. On the other hand, if the repair time is long, regardless of how frequently the failure occurs, a reserve capacity system is preferable as optimal buffer capacity is large. In every case, it is necessary to model the productivity and compare system costs to their performance benefits and operational costs.

The intent of this work is not to replace buffers as a suitable method of productivity improvement in production system design, but rather to provide other options for consideration. Using either configuration or buffers with serial lines alone or blending buffers with non-traditional configurations may provide the optimal solution for a given production circumstance. Future work will deal with the trade-off between failure frequency and repair duration to optimal implementation of reserve capacity, and the required level of material handling availability for effective crossover system design.

5 ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support of the NSF Engineering Research Center for Reconfigurable Manufacturing Systems (NSF Grant EEC95-92125) at The University of Michigan, the valuable input from the Center's industrial partners and the support of the Minerva Schlesinger Laboratory at the Technion.

6 REFERENCES

[1] Koren, Y., Heisel, U., Jovane, F., Moriwaki, T., Pritschow, G., Ulsoy, G., Van Brussel, H., 1999, Reconfigurable Manufacturing Systems, *Annals of the CIRP*, 48/2:527-540.

[2] Koren, Y., Hu, S.J., Weber, T., 1998, Impact of Manufacturing System Configuration on Performance, *Annals of the CIRP*, 47/1:369-372.

[3] Freiheit, T., Shpitalni, M., Hu, S.J., 2002, Productivity of Parallel-Serial Manufacturing Lines with and without Crossover, *ASME International Mechanical Engineering Congress and Exposition*, Nov. 17-22, 2002, New Orleans, LA.

[4] Gershwin, S.B., 1987, An Efficient Decomposition Method for the Approximate Evaluation of Tandem Queues with Finite Storage Space and Blocking, *Operations Research*, 35/2:291-305.

[5] Dallery, Y., David, R., Xie, X-L., 1989, Approximate Analysis of Transfer Lines with Unreliable Machines and Finite Buffers, *IEEE Transactions on Automatic Control*, 34/9:943-953.

[6] Yang, S., Hu, S.J., Koren, Y., 1999, Approximate Analysis of Production Lines with Loop, Scrapping, and Serial/Parallel Blocks, *Second Aegean International Conference on "Analysis and Modeling of Manufacturing Systems,"* May 16-20, 1999, Tinos Island, Greece.