Local and integral control of workload†

RENÉ DE KOSTER‡ and JACOB WIJNGAARD‡

In most of the literature on performance analysis of production systems, buffer are assumed to be controlled locally. In automated production systems buffers are not always the result of local physical space restrictions, but may also be software-controlled. Software-controlled buffers allow a more efficient use of the production units and a better control of the work in progress. Some of these possible control principles are known from production control and goods flow theory. We present a model for production systems in which the goods flow is continuous. The effect of different control mechanisms on the performance of the system (that is throughput and throughput time) is studied. Examples of control systems are BSC (Base Stock Control), Workload Control and control by local buffers only. It appears that for our model many of these control systems yield equivalent performance measures.

1. Introduction

There exists a vast literature on performance analysis of production systems. Questions that are considered are for instance:

- (a) What will the throughput of a certain system be?
- (b) What will the distribution of the buffer contents be?
- (c) What will the throughput time of the products be?

And many others.

Different techniques are used to answer such questions. The main tools are simulation and heuristics based on queueing theory. Authors using simulation are, for instance, Freeman (1964) and Anderson and Moodie (1969). Analysis of 1-product flow lines, based on the results of queueing theory, can be found in Gershwin (1987), Perros and Altiok (1986) and Pollock *et al.* (1985). More complicated lay-outs are dealt with by Altiok and Perros (1987), and Perros and Snyder (1986). A description of a real-life implementation of such a tool is given by Durlinger (1985).

In flow lines it is also possible to use regeneration methods in models with (continuous) production rates instead of service times. See Wijngaard (1979) and De Koster (1987). In almost all these studies the buffers are local. A local buffer in a flow line is depicted in Figure 1.

Machine B has to stop (is starved) if there are no products available in the buffer. Machine A has to stop (is blocked) if the buffer is full. Local buffers are the consequence of physical space restrictions, for instance, a transportation chain with a limited number of spaces for product carriers.

Now more flexible automated production systems are becoming available, which are also more flexible with respect to buffer usage. Instead of being hardware-

Received May 1987.

[†]Research supported by The Netherlands Organization for the Advancement of Pure Research (NWO).

[‡]Eindhoven University of Technology, Department of Industrial Engineering and Management Science, P.O. Box 513, NL 5600 MB, Eindhoven, The Netherlands.

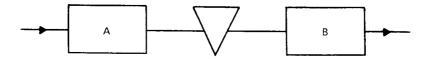


Figure 1. Two-stage flow line with buffer.

controlled, the buffers are software-controlled. It is also possible to implement non-local control limits. For instance, the sum of the contents of a number of buffers may not exceed some critical R. Non-local buffers caused by physical restrictions occur in flexible manufacturing systems. Buzacott and Shanthikumar (1980) discuss flexible manufacturing systems as sketched in Fig. 2, where the machines have a common storage, realized by a loop conveyor.

The possibility of the use of software-induced control limits leads to the question of how to set these control limits, in order to get optimum performance. As performance measures for flow lines we use in this paper: the throughput, v (average system output and throughput time, S (average time a product stays in the system after it has been released to the system). Instead of throughput time we can also consider the average number of products in the system, L, since by Little's formula we have L = vS.

In the determination of control limits it may be possible to use insights from production control and goods flow theory, because in that field of control hard space restrictions have never played an important role. Restricting the attention to lines there are three approaches in production control that may be useful here. These approaches are local buffer control, Base Stock Control (BSC) and Workload Control (WC). In a locally controlled system a shop order for an item is generated if the level of the inventory in the buffer succeeding the production unit gets below a certain value. An example of a locally controlled system is the Kanban system (Vollmann et al. (1984) or Schonberger (1982)), where the production order for one unit of product is released by the receipt of a production kanban (card). Hence the sizes of the buffers are determined by the number of kanbans in the system.

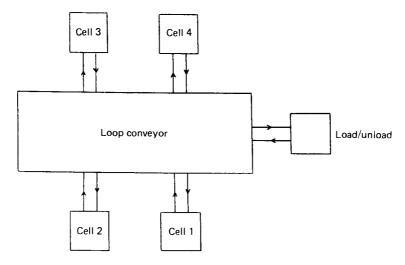


Figure 2. FMS with loop conveyor.

In the Base Stock System (BSS) (see Magee (1958) or Bertrand et al. (1985)), the buffers are not local. In Fig. 3 a production line controlled by the BSS is sketched. Production unit i (PU_i) (i=1,...,N) produces at its maximum possible rate if the echelon stock in right echelon i drops below the level K_i . It is blocked if the content in right echelon i increases to K_i . Here, according to Clark and Scarf (1960), right echelon i is defined as PU_i plus all PU's and stockpoints downstream of PU_i, and the echelon stock in the right echelon i is defined as the work-in-process in PU_i plus all stock and work-in-process downstream of PU_i. PU_N may in this context be interpreted as the demand by customers as experienced by stockpoint B_{N-1} . Base Stock Control reduces the delay in information processing which arises when orders for products are offset to preceding production units. In locally controlled systems this delay in information processing may cause the in-process stocks to get out of control. By controlling the work-in-process the throughput times are controlled as well.

Workload Control, as defined by Bertrand and Wortmann (1981) is also an integral control system. By applying WC the amount of work in the system that still has to be finished by the various work centres is kept under control. WC as applied on a flow line is indicated in Fig. 4.

 PU_1 in Fig. 4 produces at its maximum possible rate if the work-in-process in left echelon i is below K_i , for all i. PU_1 is blocked if the content in any left echelon i increases to K_i . Left echelon i is defined as B_i plus all PU's and stockpoints upstream of B_i . The aim of WC is to load the relevant capacities in an efficient way and to control the lead times. Note that if $K_i = K$, for all i, in both Figs. 3 and 4, then WC and BSC are identical.

It is also possible to allow additional, local storage restrictions in both the lines of Figs. 3 and 4. Furthermore, production units can be controlled by combinations of control, WC and BSC.

These three approaches of control are use now in a single product flow line and compared with each other. It might be expected that WC and BSC would perform better than control with local buffers, since WC uses the capacities more efficiently and with BSC there is more global tuning between the capacities. However, it appears that for the model used in this paper the three control systems are equivalent as far as the performance measures mentioned earlier are concerned.

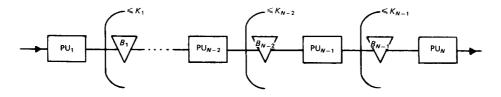


Figure 3. Production line, controlled by the BSS.

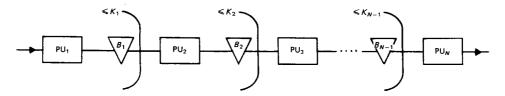


Figure 4. Production line controlled by workload control.

2. The model

In the rest of this paper N-stage flow lines as sketched in Figs. 3 and 4 are considered. The production is supposed to be continuous. The production rate of PU_i is denoted by v_i. The PU's are subject to machine failures; the time to the first failure for PU_i is exponentially distributed with parameter λ_i . Duration of repair is also exponentially distributed, with parameter μ_i . Since the echelons have a finite capacity, blocking of the PU's may occur. In the model of Fig. 3 PU_i may be slowed down when right echelon i is full (i = 1, ..., N - 1). If right echelon i is full, that is, the echelon stock is at level K_i , PU_i has to slow down to the rate at which PU_N is working (if this rate is smaller than PU_i 's rate). Hence, if PU_N is down and echelon i is full, then PU_i will stop completely and is then said to be blocked. Similarly in the model of Fig. 4, PU₁ may be slowed down to the production rate at which PU_i is working if left echelon (i-1) is full. In both the models of Fig. 3 and 4 PU_i may also have to slow down if the preceding buffer is empty. It then has to slow down to the rate of PU_{i-1} . If PU_{i-1} is down PU_i stops completely and is then said to be starved. We assume PU_1 is never slowed down by lack of sufficient input and PU_N is never slowed down by lack of sufficient storage capacity for finished products.

The major difference between the models of Figs. 3 and 4 is that in Fig. 3 all PU's are coordinated with PU_N (that is, the output process is controlled) and in Fig. 4 PU_1 is coordinated with all other PU's in order to control the input process.

We assume failures to be time-dependent, that is, they occur at the same rate when the station is forced not to run or to run at a lower speed. However, the assumption of operation-dependent failures can be solved similarly. The capacity of right echelon i in Fig. 3 is denoted by K_i . To avoid trivialities we suppose $K_1 \ge K_2 \ge ... \ge K_{N-1}$. For the line of Fig. 4 we suppose $K_1 \le K_2 \le ... \le K_{N-1}$. The throughput of a line is denoted by v(K), where $K = (K_1, ..., K_{N-1})$. In De Koster (1986) it has been shown that production lines as sketched in Fig. 3, controlled with the BSS, are comparable in performance with flow lines with local buffer control only, as sketched in Fig. 5. The buffer capacities in the locally controlled line are indicated by K_i . In order to be able to compare a base stock-controlled line with a locally controlled line, a method is needed to determine the local buffer capacities.

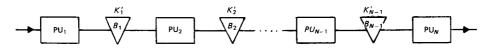


Figure 5. Flow line with local buffer control.

In De Koster (1986) a method is given to obtain the local buffer capacities K'_i as a function of the parameters of the production units and echelon capacities of the original line. Although in general the way of operating of PU_i as shown in Fig. 3 and in Fig. 5 is completely different, since the blocking process is different, the performance of the lines is approximately the same.

In the next section some examples of line equivalences will be given. It will be shown that production lines controlled with the BSS or WC have the same performance as production lines with local buffers only.

3. Numerical examples

In Table 1 an example is given of a three-stage line as sketched in Fig. 3, controlled with the BSS, and an approximating line controlled with local buffers only. L is the average content of the whole line. Similar quantities for the approximating line are denoted with index ap. For the approximating line we have $K'_2 = K_2$ and K_1 is determined as in De Koster (1986). The efficiency of PU_i, eff_i, is equal to $v_i \mu_i / (\lambda_i + \mu_i)$, which is the stand-alone average speed of PU_i. Both the line and the approximating line were simulated by 10 runs of 500,000 units of time per run, with the same random numbers, and the same initial state.

From Table 1 it appears that the differences in the performance measures for the two ways of controlling this line are small. In Table 2 a four-stage line controlled with the BSS as in Fig. 3 is approximated by a line with local buffers only. In the approximating line $K_3' = K_3$ and the other buffers are determined as in De Koster (1986). For both lines, 10 runs of 500,000 units of time per run were simulated again.

From Table 2 it again appears that the differences in the performance measures for the two ways of controlling the line are small. In order to investigate the influence of the run length in the simulations, on these results we simulated the line of Table 1 and the approximating line, for two different values for the K_i 's, 10 times with 2000 time units per run. The results are given in Table 3. σ_1^2 , denotes the variance in the throughput, σ_2^2 denotes the variance in the average line content.

		Origina	al line	Approximating line				
K_1	K_2	v(K)	L	K'_1	$v_{ap}(K')$	L_{ap}		
5	2	0.6771	4.16	4.71	0.6770	4.19		
10	4	0.7008	8.11	9.28	0.7010	8.19		
20	8	0.7295	16.13	18.29	0.7300	16.30		
30	12	0.7463	24.46	27.29	0.7469	24.75		
50	20	0.7646	41.99	45.49	0.7650	42.49		
20	20	0.7442	15.91	15.49	0.7466	16.18		
∞	∞	0.7875		_		_		

Table 1. $v = (1, 0.9, 1), \lambda = (0.01, 0.01, 0.01), \mu = (0.09, 0.07, 0.09), \text{ eff} = (0.9, 0.7875, 0.9).$

			Original line		Approximating line					
K_1	K_2	K 3	v(K)	L	K'_1	K_2'	$v_{ap}(K')$	L_{ap}		
5	3	2	0.6454	3.99	3.14	1.46	0.6420	3.75		
10	6	4	0.6756	7.86	6.31	3.15	0.6714	7.52		
20	12	8	0.7180	15.68	12.76	6.78	0.7136	15.27		
30	18	12	0.7467	23.59	19.37	10.49	0.7427	23.23		
50	30	20	0.7834	39.66	32.94	17.95	0.7801	39.56		
30	30	30	0.7624	23-38	13.83	12-24	0.7560	22.40		
∞	∞	∞	0.875							

Table 2. $v = (1.05, 1, 1.05, 1), \lambda = (0.01, 0.01, 0.01, 0.01), \mu = (0.07, 0.07, 0.07, 0.09), \text{ eff} = (0.919, 0.875, 0.919, 0.9).$

		Orig	ginal line			Approxi			
K_1	K_2	v(K)	$\sigma_1^2(\times 10^{-5})$	L	$\sigma_2^2(\times 10^{-2})$	$v_{ap}(\mathbf{K})'$	$\sigma_{1,ap}^2$	L_{ap}	$\sigma_{2,ap}^2$
5 30	2 12	0·6916 0·7571	7·5 8·1	4·15 23·42	0·40 42·1	0·6919 0·7585	7·6 8·6	4·17 24·00	0·70 55·3

Table 3. Comparison of the non-stationary behaviour of the lines of Table 1.

Also for the non-stationary behaviour there seems to be no significant difference in the performance of the line with respect to throughput and throughput time. The variances in these measures are greater for the locally controlled line, but in the simulations of Table 1 there was hardly any difference.

The next example we consider is a five-stage line which is controlled by WC as in Fig. 4. This line was again approximated by a locally controlled line. For both these lines, 10 runs of 500,000 units of time per run were simulated. The local buffers in the approximating line were obtained by a combined trial-and-error search method and a method similar to that described in De Koster (1986). The results are listed in Table 4.

In the approximating line K'_i had to be taken small, as can be seen in Table 4, since the first buffer tends to become nearly full, regardless of its capacity, since PU_i has the greatest efficiency.

				Origin	al line		Approxim				
K_1	K_2	K_3		$v(\mathbf{K})$	L	K_1'	K' ₂	K_3'	K_4'	$v_{ap}(\mathbf{K}')$	L_{ap}
4	4	4	4	0.5294	3.43	0.7	1.9	1.2	4.0	0.5276	3.34
2	3	5	7	0.5473	3.88	0.35	1.9	2.1	6.2	0.5438	3.81
4	6	10	14	0.5975	7.92	1.0	3.6	4.6	13.0	0.5931	7.77
10	20	30	40	0.6909	24.41	3.6	14.2	14-4	37.5	0.6849	23-84
15	30	40	50	0.7142	33.69	5.3	20.8	21.2	45.0	0.7100	33-16
∞	∞	∞	∞	0.7875	_				_	_	

Table 4. $v = (1.05, 1, 0.9, 0.9, 1.1), \lambda = (0.01, 0.01, 0.01, 0.01, 0.02), \mu = (0.07, 0.09, 0.07, 0.09, 0.09),$ eff = (0.919, 0.9, 0.7875, 0.81, 0.9).

4. Discrete product models

From the results presented in this paper it seems that, at least for the model under consideration, there is no significant difference in performance of control systems like Base Stock Control and Workload Control, on the one hand, and control by local buffers only on the other hand. Although this may seem strange at first sight, one should bear in mind that the advantage of BSC over local control is that the lag in information, which arises when orders for product are passed on to preceding production units, is reduced. This delay of information processing may cause the inprocess stocks to grow out of control in locally controlled production systems. However, in the model considered in this paper the goods flow is supposed to be continuous, so that little delay of information processing occurs. As soon as the stock in a certain stockpoint drops below the maximum level, the preceding PU starts producing (if possible), and the stock is instantaneously replenished.

It therefore remains questionable whether the results obtained here are due to the specific model used, or are also approximately valid for lines with discrete products, in which the PU's have (random) processing times. In Gershwin (1987) a discrete-product model is outlined in which the machines are unreliable and have equal processing times. With some minor adaptations this model can be seen as the discrete-product version of the continuous flow model used in this paper. Consider such a two-stage flow line with intermediate buffer of integer capacity K. We again assume that the machines are unreliable. While a machine is working (that is, operational and neither blocked nor starved), a fixed amount of time is required to process a part. This time is the same for both machines and is taken as the time unit. During a time unit, when PU_i is operational it has probability p_i of failing. When failed, PU_i has probability r_i of being repaired during a time unit. We suppose PU₁ starts working on an item only if the buffer is not full. This model is the same as the one of Gershwin (1987), except for the fact that in our model failures may occur when a machine is blocked or starved. Furthermore, we assume that a time unit is so small that no two events (failure or repair of a machine) occur in the same time unit. A more precise description of other assumptions can be found in Gershwin (1987). For this discrete-product model we can use the same type of regenerative method as Wijngaard (1979), in order to obtain throughput, throughput time and many other relevant line quantities. All equations are similar to the continuous goods flow case, except that λ_i has to be replaced by p_i and μ_i by r_i. The differential equations of Wijngaard (1979) correspond to difference equations. In Wijngaard's article the solution of the differential eqn. (8) is

$$w(x) = -\frac{\varepsilon + \eta}{r_f - r_g} + \left[\frac{-\varepsilon(K)}{r_f} + \frac{\varepsilon + \eta}{r_f - r_g} \right] \exp\left(-(r_f - r_g)(K - x)\right), \quad 0 \le x \le K$$
 (1)

In our model the solution of the corresponding difference equation is:

$$\tilde{w}(x) = -\frac{\tilde{\varepsilon} + -\tilde{\eta}}{\tilde{r}_f - \tilde{r}_g} + \left[\frac{-\tilde{\varepsilon}(K)}{\tilde{r}_f} + \frac{\tilde{\varepsilon} + \tilde{\eta}}{\tilde{r}_f - \tilde{r}_g} \right] \left(\frac{1 - \tilde{r}_g}{1 - \tilde{r}_f} \right)^{-(K - x)}, \quad x \in \{0, \dots, K\}$$
 (2)

where $\tilde{\epsilon}, \tilde{\eta}, \tilde{r}_f, \tilde{r}_g$ correspond to ϵ, η, r_f and r_g , respectively, with λ_i, μ_i replaced by p_i and r_i , respectively. By expanding the *e*-power in (1) and the corresponding power in (2) in a power series and dropping terms of $o(r_f)$, $o(r_f)$, $o(r_f)$, $o(r_f)$, $o(\tilde{r}_g)$, $o(\tilde{r}_g$

This correspondence also holds for N-stage lines, and hence it will be possible to approximate discrete product flow lines of the above type, controlled with WC or BSC and with echelon buffers of adequate size. If the echelon buffers are too small, the restriction that buffer sizes have to be integer-valued may prevent us from finding a good locally controlled equivalent line.

An objection to this model may be that the time unit has to be small, since it is assumed that no two events occur in the same time unit. This implies that the goods flow is *nearly* continuous. Therefore we present another discrete product model in which service times are random.

In this model of an (N+1)-stage flow line we assume PU_i $(0 \le i \le N)$ has a service time that is exponentially distributed with parameter μ_i . In front of PU_0 there is a buffer B which is always full. The service rate of PU_0 is λ . See also Fig. 6.

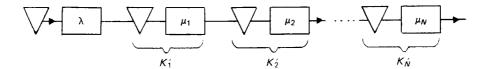


Figure 6. (N+1)-stage discrete product flow line.

When PU_i completes an operation on a product this product proceeds to the (i+1)st queue if there is space available. However, if the (i+1)st queue is full at that time, the product is forced to wait in front of PU_i until a departure occures from PU_{i+1} . The capacity of the *i*th queue, including the service position at PU_i is K_i . The line of Fig. 6 can also be controlled by BSC in a straightforward way. See Fig. 7.

In the sequel it will be shown that discrete product flow lines of this type controlled with BSC have also approximately the same performance as locally controlled lines. The K_i' in Fig. 6 are determined using a method as in De Koster (1986) combined with an approximation method for locally controlled flow lines of Perros and Snyder (1986) and trial and error. Some results are listed for a four-stage line in Table 5. The 'original' line is controlled with BSC, the 'approximating line' by local control. In the approximating line the buffer capacity of the last buffer is chosen equal to the echelon capacity of the last echelon. For both lines, 100,000 units of time were simulated, with the system empty initially. L stands for the mean number of products in PU_0 or downstream of PU_0 in the line. Note that the mean number of products in PU_0 is 1 at all times for both lines.

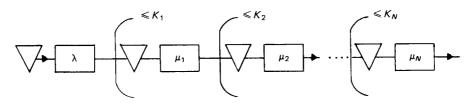


Figure 7. (N+1)-stage discrete product flow line controlled with BSC.

			Origina	ıl line		Appro	oximating li	ine
K_1	K_2	K_3	v(K)	L	K'_1	K_2'	$v_{ap}(\mathbf{K}')$	L_{ap}
7	5	3	0.7237	6.74	3	3	0.7194	6.56
10	6	4	0.7739	9.02	5	4	0.7779	9.10
15	10	5	0.8320	13.01	7	7	0.8306	12.92
20	15	10	0.8857	16.04	9	9	0.8838	16.19

Table 5. $\lambda = 1$, $\mu = (1.1, 1.05, 1)$.

From Table 5 we see that again both lines have approximately the same performance. However, for smaller-echelon buffers it is more difficult to find a locally controlled line, since we have the restriction that buffer capacities need to be integral.

This appears also from the first case listed in Table 6. The results listed in this table were obtained for both (four-stage) lines by one simulated run of 100,000 units of time, with the same initial state for both lines.

			Origina		Appro	oximating li	ne	
K_1	K_2	K_3	$v(\mathbf{K})$	L	K'_1	K_2'	$v_{ap}(\mathbf{K}')$	L_{ap}
3	3	2	0.5847	3.55	1	2	0.6109	3.50
5	5	3	0.6930	5.05	2	3	0.7028	4.93
6	6	4	0.7303	5.77	3	3	0.7345	5.95
9	7	5	0.7994	7.94	5	4	0.8040	8.14
19	16	11	0.9023	14.43	11	11	0.9038	14.41

Table 6. $\lambda = 1$, $\mu = (1, 1.2, 1.1)$.

5. Conclusions

For the different models studied in this paper there is no significant difference in performance of control systems like Base Stock Control and Workload Control on the one hand and control by local buffers only, on the other hand. The performance measures considered in the paper are throughput and mean work-in-process. With respect to the model with a continuous goods flow we can also state that this result holds for flow lines containing unreliable machines in parallel, or containing machines with more than two different speeds, where the duration of a certain speed-interval has an arbitrary distribution of phase type. It is not at all clear that BSC has any advantage in performance over local control for the models discussed. Maybe there is some gain in performance if many different products are to be controlled. The aim of WC is to load the capacities efficiently and to keep the throughput times constant. In our simulation runs for the continuous goods flow lines it was observed that the variance in mean work-in-process in the workload controlled line was always smaller than the variance in mean work-in-process of the approximating line. However, for the BSC examples this was not always the case. Although BSC, WC and local control yield approximately the same utilization factors and average throughput time, it still may be so that BSC and WC have a positive effect on the variance of the throughput time. Table 3 may indicate this.

References

ALTIOK, T., and Perros, H. G., 1987, Approximate analysis of arbitrary configurations of open queueing networks with blocking. *Annals of Operations Research*, **9**, 481–509.

Anderson, D. R., and Moodie, D. C., 1969, Optimal buffer storage capacity in production line systems. *International Journal of Production Research*, 7, 233–240.

Bertrand, J. W. M., Monhemius, W., and Timmer., 1985, Production and inventory control with the base stock system. Report EUT/BDK/12, Eindhoven University of Technology, The Netherlands.

Bertrand, J. W. M., and Wortmann, J. C., 1981, Production Control and Information Systems for Component Manufacturing Shops (Amsterdam: Elsevier).

BUZACOTT, J. A., and SHANTHIKUMAR, J. G., 1980, Models for understanding flexible manufacturing systems. AIIE Transactions 12, 339-349.

CLARK, A. J., and SCARF, H., 1960, Optimal policies for a multi-echelon inventory problem. Management Science 6, 475-490.

DE KOSTER, M. B. M., 1988, Estimation of line efficiency by aggregation. *International Journal of Production Research*, 25, 615-626.

- DE KOSTER, M. B. M., 1987, Approximations of flow lines with integrally controlled buffers. *IEE Transactions* (in the press).
- Durlinger, P. P. J., 1985, The trade-off-module; a real life implementation. Report BDK/KBS/85/03, Eindhoven University of Technology.
- Freeman, M. C., The effects of breakdowns and interstage storage on production line capacity. Journal of Industrial Engineering 15, 194-200.
- GERSHWIN, S. B., 1987, An efficient decomposition method for the approximate evaluation of tandem queues with finite storage and blocking. *Operations Research*, 35, 291–305.
- MAGEE, J. F., 1958, Production Planning and Inventory Control (McGraw-Hill, New York).
- Perros, H. G., and Altiok, T., 1986, Approximate analysis of open networks of queues with blocking: tandem configurations. *IEEE Transactions on Software Engineering* SE-12, 450-461
- Perros, H. G., and Synder, P. M., 1986, A computationally efficient approximation algorithm for analyzing open queueing networks with blocking. Research report, North Carolina State University, Raleigh.
- POLLOCK, S. M., BIRGE, J. R., and ALDEN, J. M., 1985, Approximation analysis for open tandem queues with blocking: exponential and general service distributions. Manuscript, I. E. and O. R. Dept. 83–16, Michigan University.
- SCHONBERGER, R. J., 1982, Japanese Manufacturing Techniques (New York: Free Press).
- VOLLMANN, T. E., BERRY, W. L., and WHYBARK, D. C., 1984, Manufacturing Planning and Control Systems (Homewood; Dow Jones-Irwin).
- WIJNGAARD, J., 1979, The effect of interstage buffer storage on the output of two unreliable production units in series, with different production rates. AIIE Transactions 11, 42–47.