SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK

TECHNICAL REPORT NO. 702

June 1986

BUFFER PLACEMENT IN SEQUENTIAL PRODUCTION LINES: FURTHER STUDIES

By

Yale Herer

This research was partially supported by National Science Foundation Grant No. DMC-8509131 and by a research grant from the IBM Center of Competency for Manufacturing Modelling.

This report is a continuation of the studies done previously by Joneja (1). Refer to The Model section of that report for clarification of the terminology used in this report.

Significance Study

The purpose of this section of the study is to determine whether certain earlier findings of Joneja (1) were significant. Three different setups were chosen for analysis. They were chosen for their representability, diversity, and the relative closeness of the numbers already obtained.

All the studies were on serial production lines with either 3 or 4 workcenters and a buffer between each workcenter. The issue is where to put these buffers to obtain the greatest utility.

The first study was on a four workcenter line with buffers between each workcenter (see Figure 1(a)). The production times at each workcenter had a uniform distribution [5,15]. There were three available buffer spaces with the following configurations analyzed.

	BUFFER b1	BUFFER b2	BUFFER b3	TOTAL
	1	1	1	3
	2	1	0	3
	1	2	0	3
*	0	1	2	3
*	0	2	1	3

(*) Denotes that it was not investigated directly, but rather the reversibility theorem [Muth (2)] was used to implicitly investigate these possibilities.

Results for all three setups are discussed jointly after each is introduced.

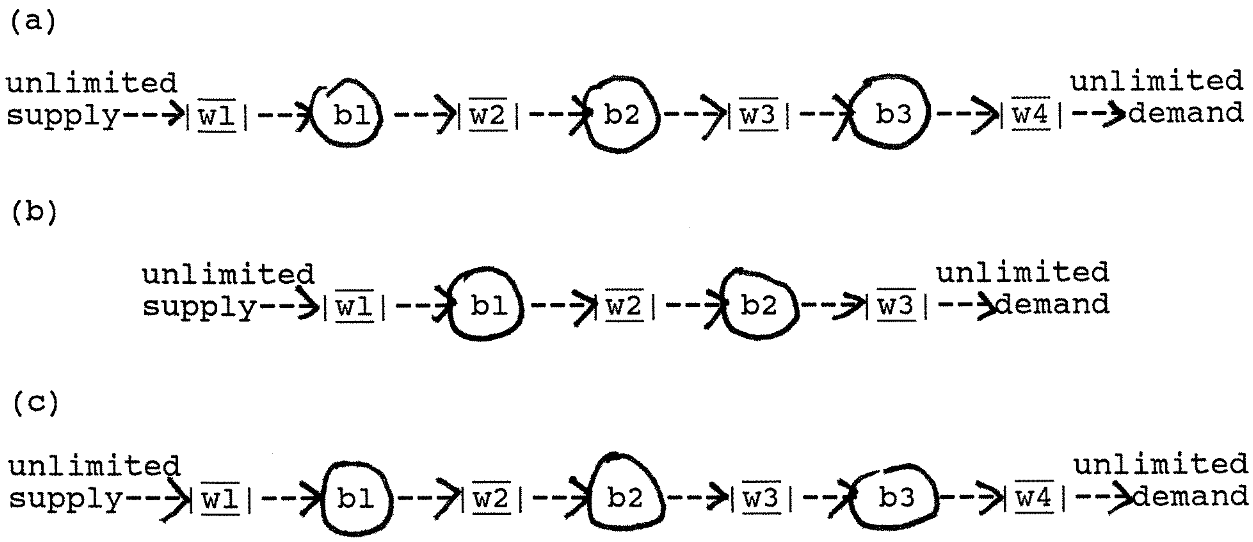


Figure 1.

The second study was on a 3 workcenter line with buffers between each workcenter (see Figure 1(b)). The production times at each workcenter were uniformly distributed with workcenters w1 and w3 having uniform [5,15] while workcenter w2 had uniform [6,16]; workcenter w2 was the bottleneck. There were two available buffer spaces with the following configurations analyzed.

	BUFFER b1	BUFFER b2	TOTAL
	1	1	2
	2	0	2
*	0	2	2

(*) Denotes same as above.

The third study was on a 4 workcenter line with buffers between each workcenter (see Figure 1(c)). The production times at each workcenter were uniform, with workcenters w1, w2, and w4 having uniform [5,15], while workcenter w3 had uniform [6,16], thus making workcenter w3 a bottleneck. There were two available buffer spaces with the following configurations analyzed.

BUFFER b1	BUFFER b2	BUFFER b3	TOTAL
0	1	1	2
0	2	0	2
1	1	0	2
1	0	1	2

The first configuration in each table was determined to be the "best" by the previous studies. To check whether this determination was statistically significant a Wilcoxan rank-sum test [Devore (3)] was performed. In all cases the hypothesized best configuration was found to be statistically better than all the other configurations at a significance level of over ninety-nine percent. Hence we can conclude that the earlier findings were significant, and thus the rules developed from these findings should be regarded as valid. That is; place buffers toward the middle of serial production lines, and near bottlenecks.

Rework Loop Study

The following configuration was studied:

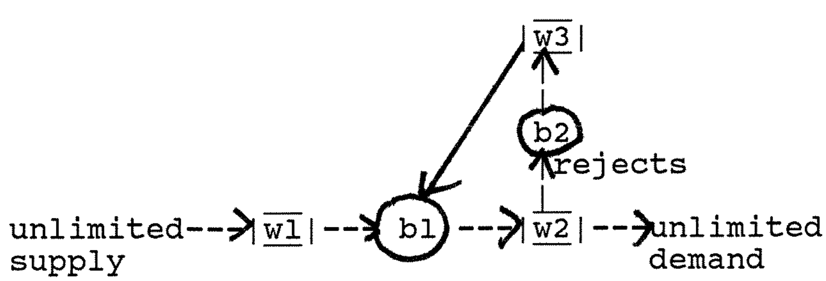


Figure 2.

The processing times were chosen so that they would balance the expected flow through the workcenters. The reject percentage was initially set at 5, 25, and 50 percent. The size of buffer b1 was set at 2, 10, and 20. The size of buffer b2 was set at 1, 10, and 20. This configuration was chosen because it represented the simplest non-trivial configuration.

NOTES TO XCELL USERS:

A value of zero for b1 is impossible because w2 is supplied by both w1 and w3. The only way to simulate this using XCell is to use a buffer.

To prevent a deadlock in the system a trigger was used at w1. For example if b1 had capacity 2 then w1 would be triggered low at 1. To simulate the triggering as actually being a block, one buffer space was added to b1. This was used to represent w1 working on the product and not <u>releasing</u> it until there was room in b1 as opposed to not starting work until there was room in b1.

Just as a capacity of 1 was used as the smallest value for b2, one over the minimum of zero. The smallest value used for the capacity of b1 was 2, one over the minimum.

The first observation was that no matter what the percentage of rejects were all the systems behaved similarly, with the only difference being that the smaller the reject percentage, the stronger was the observed trends in the system. For this reason, and because it was felt that a five percent reject rate was more representative of real problems, only a reject rate of five percent was used for further studies.

For the first part of the study all machine processing times were held constant (deterministic); thus the only random factor was due to rejections. For a reject percentage of five percent the following data was obtained. (Note: the two columns contain the same information, but are presented in a different order.)

BUFFER b1	BUFFER b2	THROUGHPUT	BUFFER b1	BUFFER b2	THROUGHPUT
2	1	4686	2	1	4686
2	10	5506	10	1	4705
2	20	5701	20	1	4706
10	1	4705	2	10	5506
10	10	5513	10	10	5513
10	20	5701	20	10	5513
20	1	4706	2	20	5701
20	10	5513	10	20	5701
20	20	5701	20	20	5701

Table 1.
The standard deviations of the throughputs ranged between 5 and 17.

The first trend that becomes apparent is increasing the size of buffer b2 has a notable impact on throughput, while increasing the size of buffer b1 has a very low (if any) effect on throughput.

The next step of the study was to change the processing time distribution at workcenter w3 to exponential, with the same mean as the previous constant processing time. No other changes were made.

BUFFER b1	BUFFER b2	THROUGHPUT	BUFFER b1	BUFFER b2	THROUGHPUT
2	1	4240	2	1	4240
2	10	5332	10	1	4238
2	20	5516	20	1	4237
10	1	4238	2	10	5332
10	10	5141	10	10	5141
10	20	5450	20	10	5190
20	1	4237	2	20	5516
20	10	5190	10	20	5450
20	20	5452	20	20	5452

Table 2.
The standard deviations of the throughputs ranged between 66 and 109.2

The trend observed above becomes more prominent here, in fact in some cases when the capacity of buffer b1 was increased the throughput actually went down. One should notice, however, that the high standard deviations of throughput imply that the differences, as shown here, fail to be statistically significant. But the important item to notice is that even if the throughput does not decrease it does not increase substantially. Thus the prudent course of action is to put excess (above a certain point) buffer capacity into buffer b2 of Figure 2, and not buffer b1. Thus one can get a larger impact on throughput from each buffer space added to the system. For example: Throughput is 326 units higher when a total of 22 buffer spaces is used with buffer b1 having capacity 2 and buffer b2 having capacity 20, over using 30 buffer spaces with buffer b1 having capacity 20 and buffer b2 having capacity 10. Thus it is possible to get six percent higher throughput, with twenty-seven percent fewer buffer spaces, if the buffer spaces are used judiciously. The other trend to notice, but not surprising, is that as more buffer space is allocated to buffer b2, the marginal benefit decreases.

One very definite trend in all the studies that does not appear in these results is the interaction of the buffers. When there are constant processing times at all three stations the interaction is most prominent. Buffer b2 always leads buffer b1. This means buffer b1 will not increase (though it will always oscillate; as product is alternately put in and taken out) until buffer b2 becomes full; and buffer b1 will not decrease until buffer b2 becomes empty. See Figure 3. This is due to the fact that w2 is faster than w1. W2 takes all the product w1 can

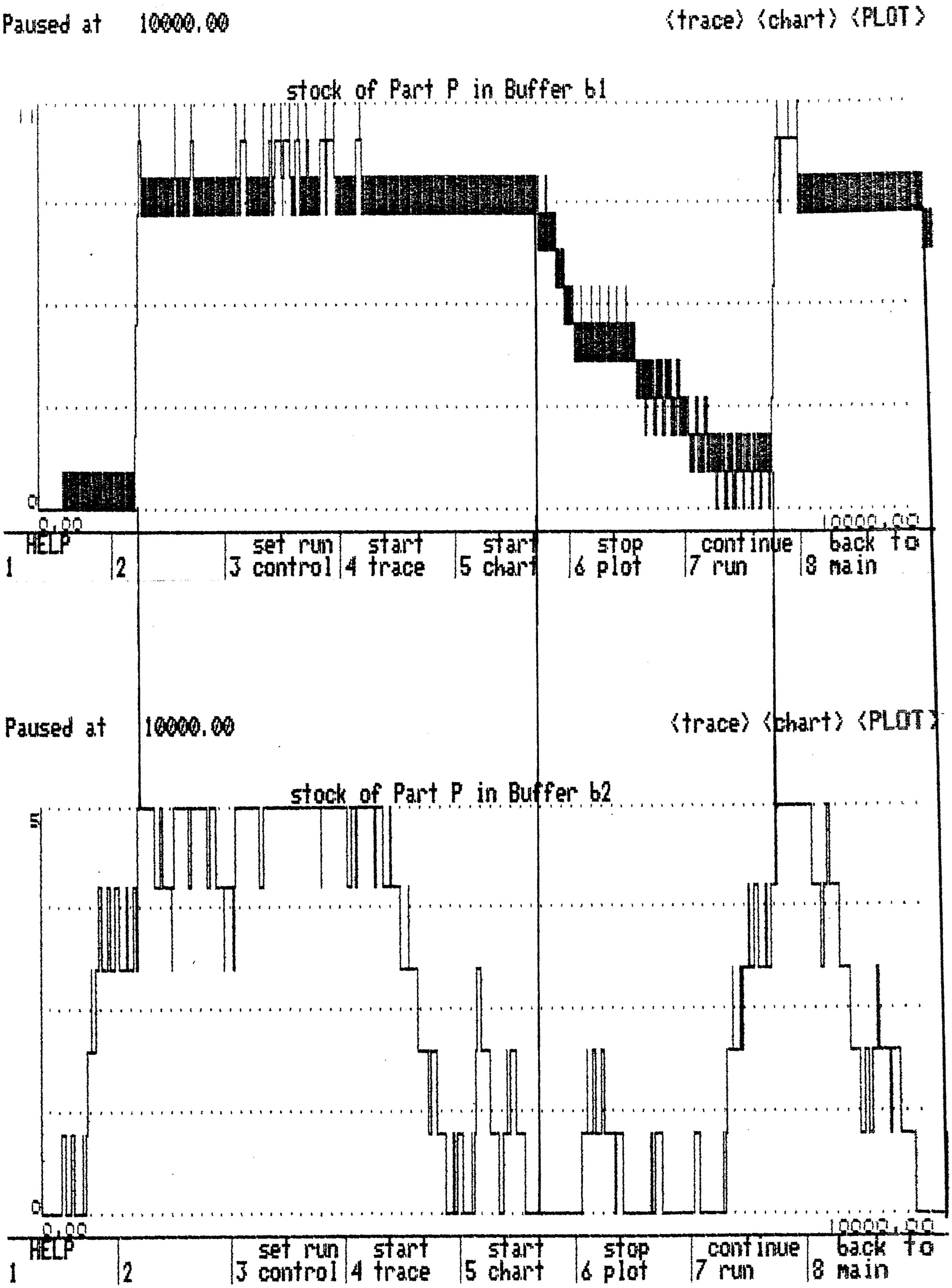

Paused at 10000.00
<trace> <chart> <PLOT>
stock of Part P in Buffer b1
0
0.00
10000.00
HELP
1
2
set run
3 control
start
4 trace
start
5 chart
stop
6 plot
continue
7 run
back to
8 main
Paused at 10000.00
<trace> <chart> <PLOT>
stock of Part P in Buffer b2
5
0
0.00
10000.00
HELP
1
2
set run
3 control
start
4 trace
start
5 chart
stop
6 plot
continue
7 run
back to
8 main

Figure 3.

produce out of b1. While if w3 dumps product into b1 the amount that w2 is faster than w1 will just clear the extra unit out of b1 before another unit can be dumped from w3. Hence the only way b1 can increase is if w2 is blocked (b2 full); and the only way b1 can decrease is if w3 does not put any units into b1 (b2 empty). This phenomenon was also strongly present when exponential processing times were introduced at workcenter w3, but the relationships were no longer absolute. This apparently explains why increasing the capacity of buffer b2 has a greater effect on throughput than increasing the capacity of buffer b1.

Response to Processing Time Changes

The next part of the study was to vary the processing times at workcenter w2. The configuration of Figure 2 was used again in this part of the study. With no changes in any values; except as noted above. Note: This created an unbalanced line. This was done to see what effect increasing the efficiency of one of the workcenters would have on total throughput. Buffer capacities were held at 10, workcenters w1 and w2 had constant processing times, while workcenter w3 had exponential processing times. The following data was collected:

Processing Times at Workcenter w2	Throughput
7	4070
6	4649
5	5141
4	5358
3	5258
2	5377
1	5461
0	5446

Table 3.

See Figure 4, for a graph of these numbers.

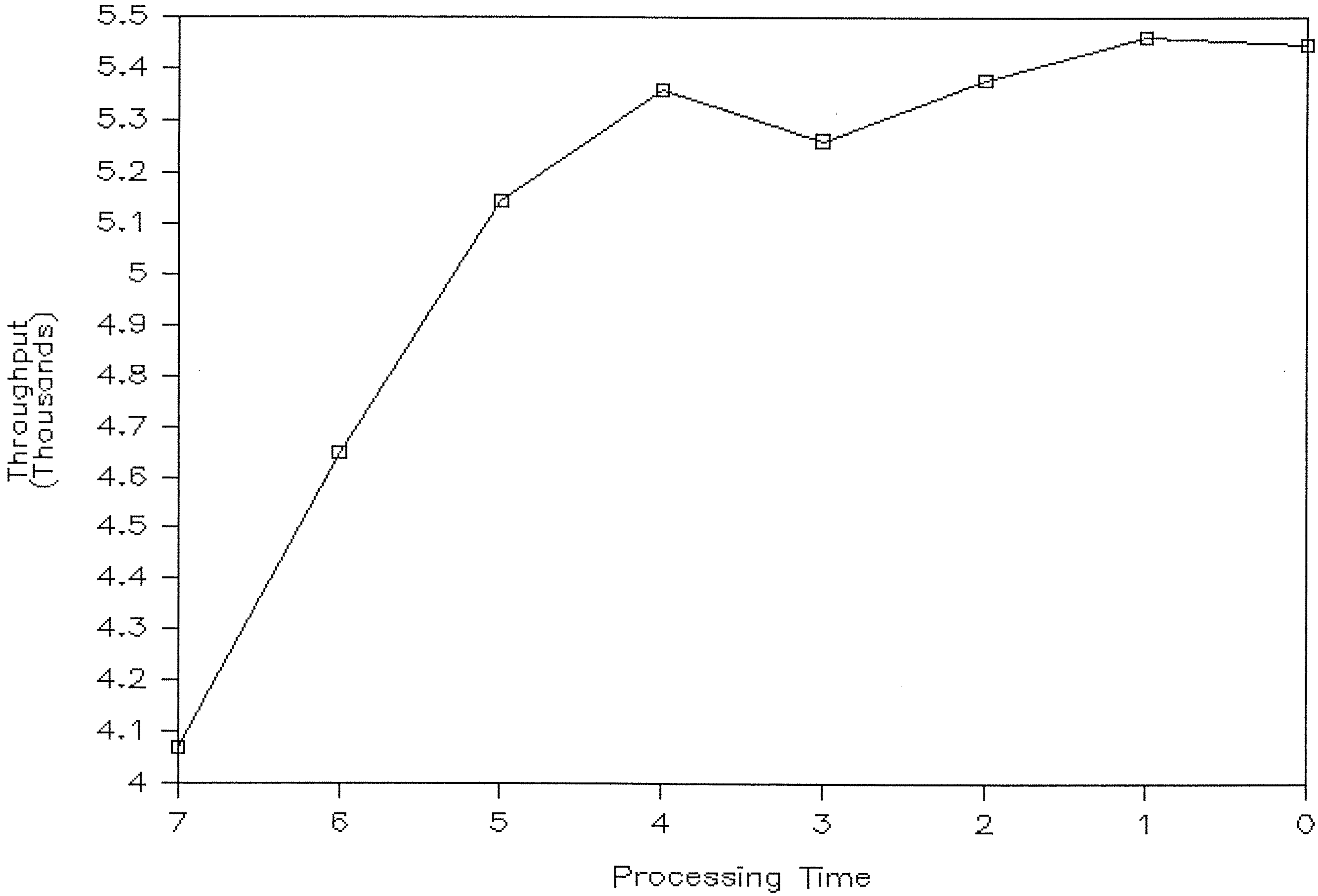
The interesting item in this table was the apparent decrease in the throughput as the processing time decreased from 4 to 3. To get a better handle on what was going on, replications were made and a Wilcoxan rank-sum test [Devore (3)] with a degree of significance of .95 was performed. The difference failed to be shown as significant, but the important item to notice is the throughput is relatively constant over a certain range of processing times, thus we can conclude that it is not a good idea to spend a lot of effort in reducing the processing time in workcenter w2 without reducing the processing times at workcenters w1 and w3.

Processing times above 7 were not considered because at a processing time of 7, workcenter two becomes an absolute bottleneck, meaning that the throughput of the system was equal to the throughput of a system that would contain only workcenter w2.

Maintenance Study

The configuration studied here is given as follows:

Figure 4


5.5
5.4
5.3
5.2
5.1
5
4.9
4.8
4.7
4.6
4.5
4.4
4.3
4.2
4.1
4
Throughput
(Thousands)
7
6
5
4
3
2
1
0
Processing Time

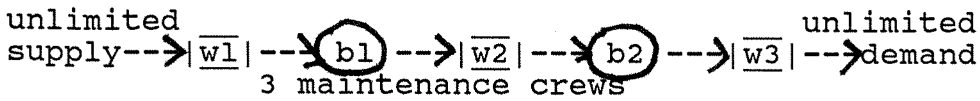


Figure 5.

Each workcenter has identical and constant processing times, and each workcenter is subject to random breakdowns. There are three maintenance crews, thus a workcenter never has to wait for another workcenter to be repaired before its repair can be started. An exponential distribution was used for the time between repair (uptime). Two distributions were used for repair time (downtime) 1) exponential and 2) discrete (set to the repair time values obtained through Wali Haider). The means of the uptime and downtime distribution were set such that the workcenters would be down for two minutes for every eight minutes working on a part.

The parameters were varied so that the expected down time was an integer multiple of the processing time. Integer multiples of 1, 2, 5, and 10 were used. Then the buffer sizes between the workcenters were varied, with the capacity of b1 always equal to the capacity of b2. A few runs were made where the processing time was varied. Two items were researched.

1) Does the probability mass of the amount of the amount of product in a buffer have higher mass at empty and full than the intervening states, as suggested by Malathronas et. al. (4).

2) What is the relationship between efficiency (actual throughput of the system divided by the theoretical throughput of the system with infinite buffering), downtime, processing time,

and buffer size.

Where to place buffers was not investigated because a workcenter that has maintenance is equivalent to a workcenter with a very strange and variable processing time distribution. In this case with exponential uptime and exponential downtime the processing time distribution is as follows:

p = Processing time (constant)
d = average Downtime
u = average Uptime

$$10 + \sum_{i=1}^{m} (t)$$

Where m has a poisson distribution with mean p / u and t has an exponential distribution with mean d. The mean of this strange distribution is 10 + ((p * d) / u) and its variance is (2 * d * d * p) / u. Thus the decision on where to place buffers can be obtained through findings in the earlier report given by Joneja (1).

1) Probability mass

There was found to be high probability masses at both empty and full. This can be seen in the printouts in appendix A, where the number of parts in the buffers are plotted on a time axis. The print outs are divided into sections where the ratio;

$$\frac{\text{processing time * buffer size}}{\text{downtime}}$$

remains constant. A buffer capacity of ten was used and the amount of material in both buffer b1 and b2 are plotted. The amount of material in the buffer definitely tends to stay at zero or ten, this is especially evident for low values of the ratio.

2) Relationship of efficiency

A suggested relationship is that efficiency is a function of:

$$\frac{\text{processing time} * \text{buffer size}}{\text{downtime}}$$

The reason this measure was suggested is that $\frac{\text{downtime}}{\text{processing time}}$ represents the amount of buffer space needed (on average) during one downtime to hold all items produced by the previous machine and conversely the amount of item the next machine (on average) can handle. Hence the suggested ratio represents the number of downtime cycles of product the buffer can hold. Thus the suggestion is that efficiency is primarily a function of the given ratio and not of buffer capacity alone. Buffer sizes used where integer multiples of $\frac{\text{downtime}}{\text{processing time}}$ including zero. We discovered efficiency to be primarily related to the ratio presented above for values of the ratio one and over. This can be seen in Figure 6 where, for exponential processing times, efficiency is plotted vs $\frac{\text{processing time} * \text{buffer size}}{\text{downtime}}$ on lines of constant $\frac{\text{downtime}}{\text{processing time}}$. Similar trends were observed for the discrete distribution. This means that efficiency, in this model, is primarily dependent on how many downtime cycles of product a buffer can hold. Also note from Figure 7 that the relationship appears to be a logarithmic one for values over one (zero could not be shown) and below "large" buffers. The right

Figure 6

exponential repair times

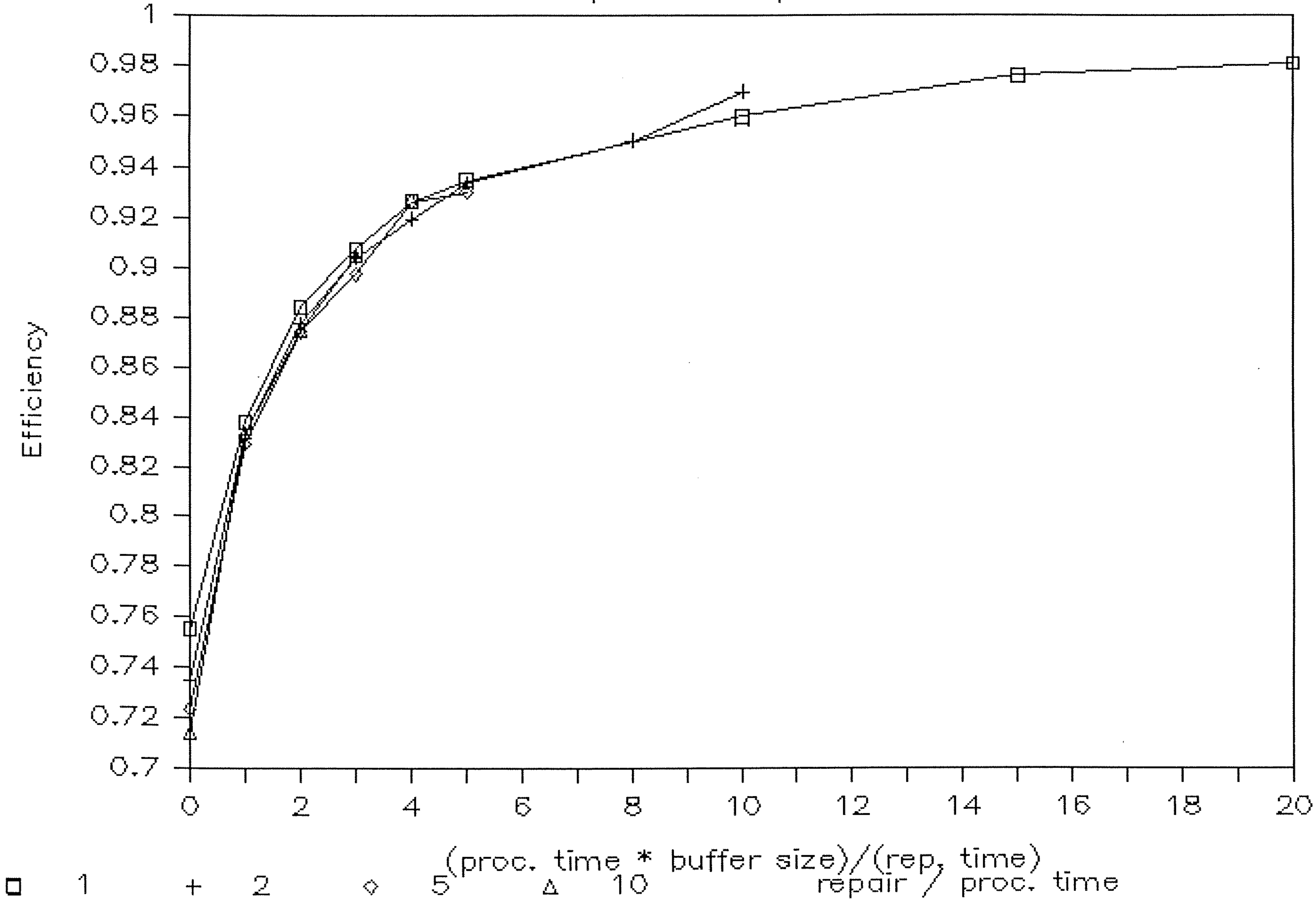
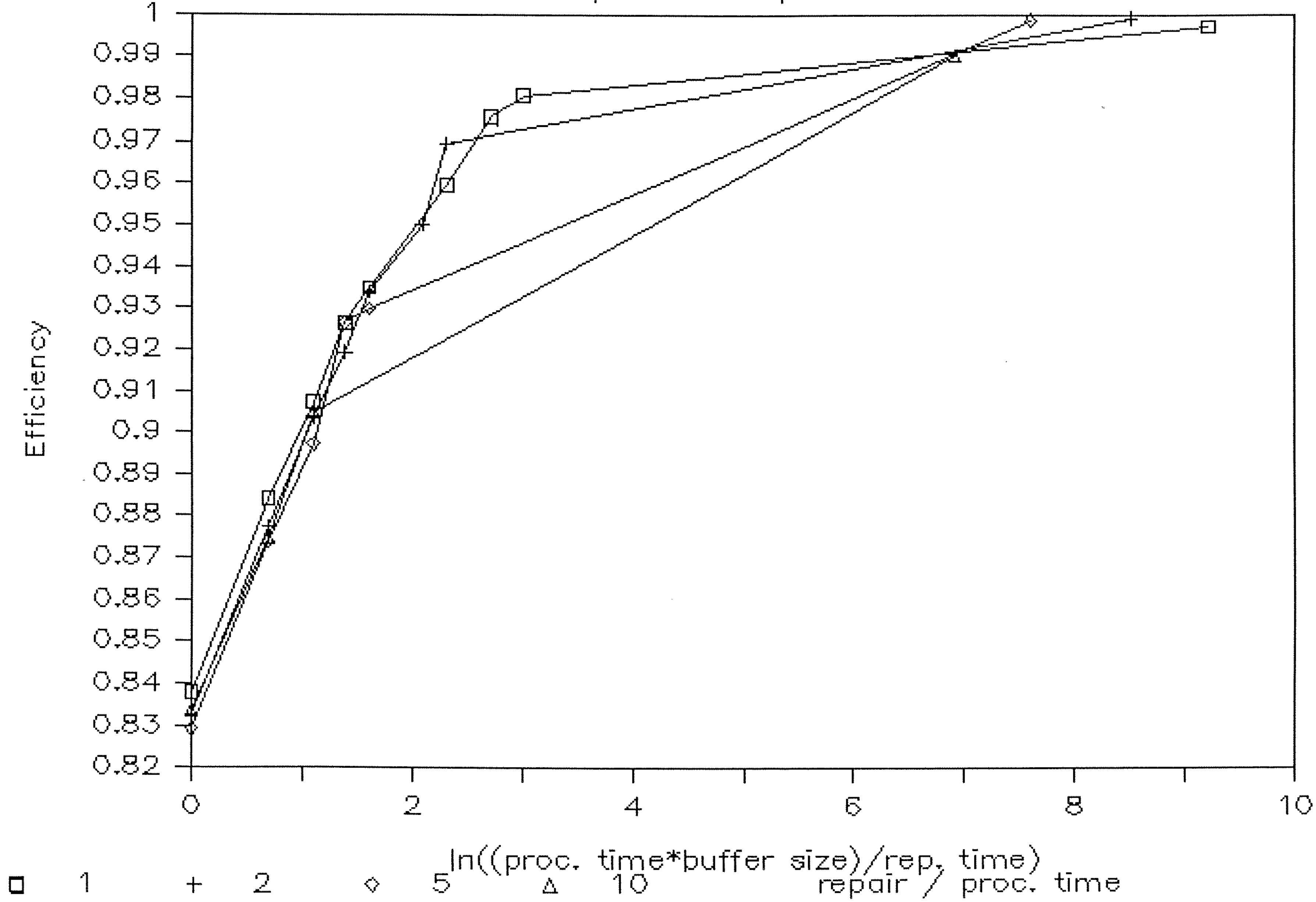


Figure 7

exponential repair times



most points represent an essentially infinite capacity.

Bibliography

(1) Joneja, D., W. L. Maxwell, "Buffer Placement in Sequential Production Lines: Considerations of Processing Time Variability," TR #700, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY.

(2) Muth, E. J., "The Reversibility Property of Production Lines," Management Science, Vol 25, No 2 pp. 152-158

(3) Devore, J. L., Probability & Statistics for Engineering and the Sciences, Brooks/Cole Publishing Company; Monterey, California 1982.

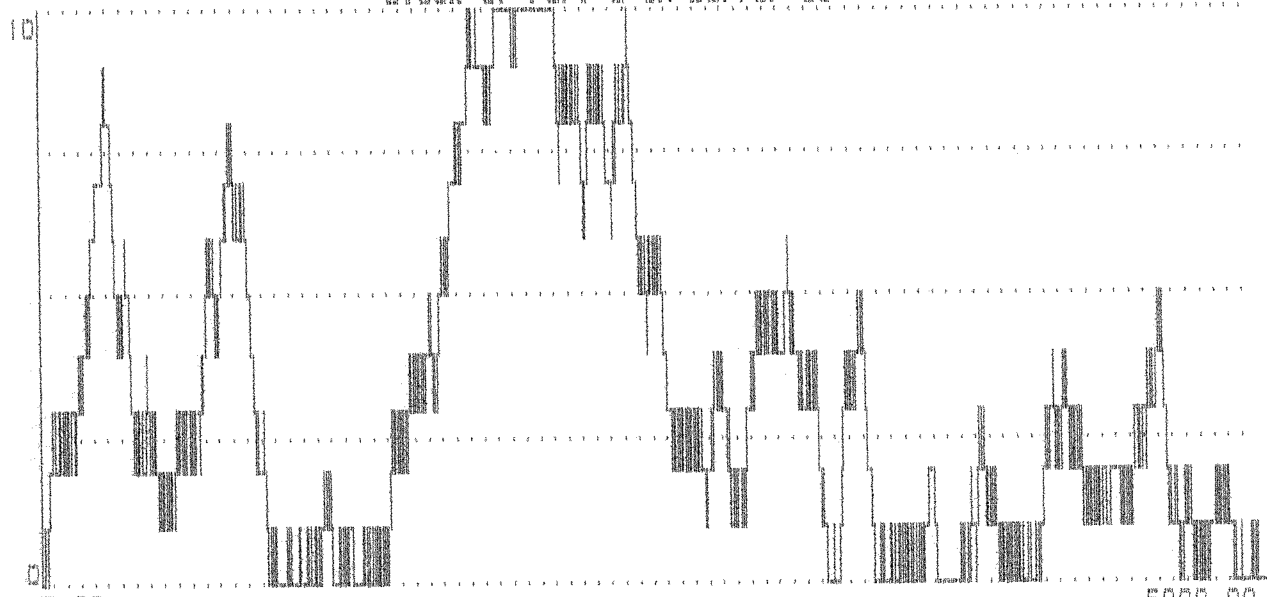
(4) Malathronas, J. P., J. D. Perkins, R. L. Smith; "The Availability of a System of Two Unreliable Machines Connected by an Intermediate Storage Tank," IIE Transactions, Vol 15, No 3 pp. 195-201.

APPENDIX A

$$\frac{\text{processing time} * \text{buffer size}}{\text{downtime}} = 10$$

Paused at 5000.00 <trace> <chart> <PLOT>

stock of Part a in Buffer b1

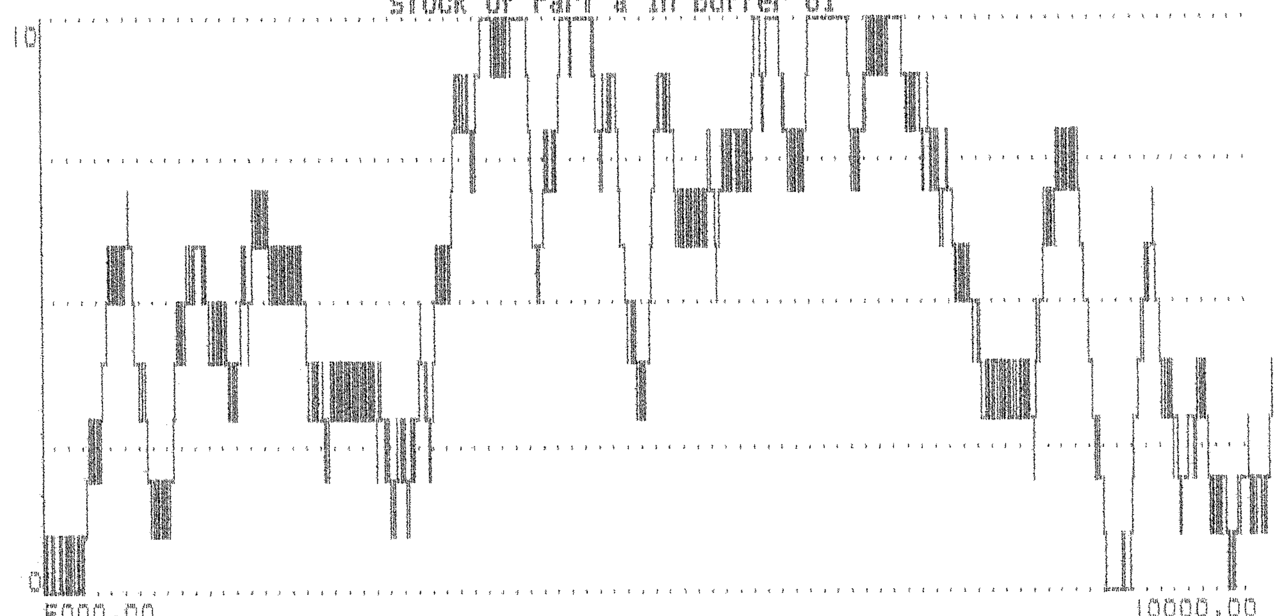


1 HELP	2	3 set run control	4 start trace	5 start chart	6 stop plot	7 continue run	8 back to main

Paused at 10000.00

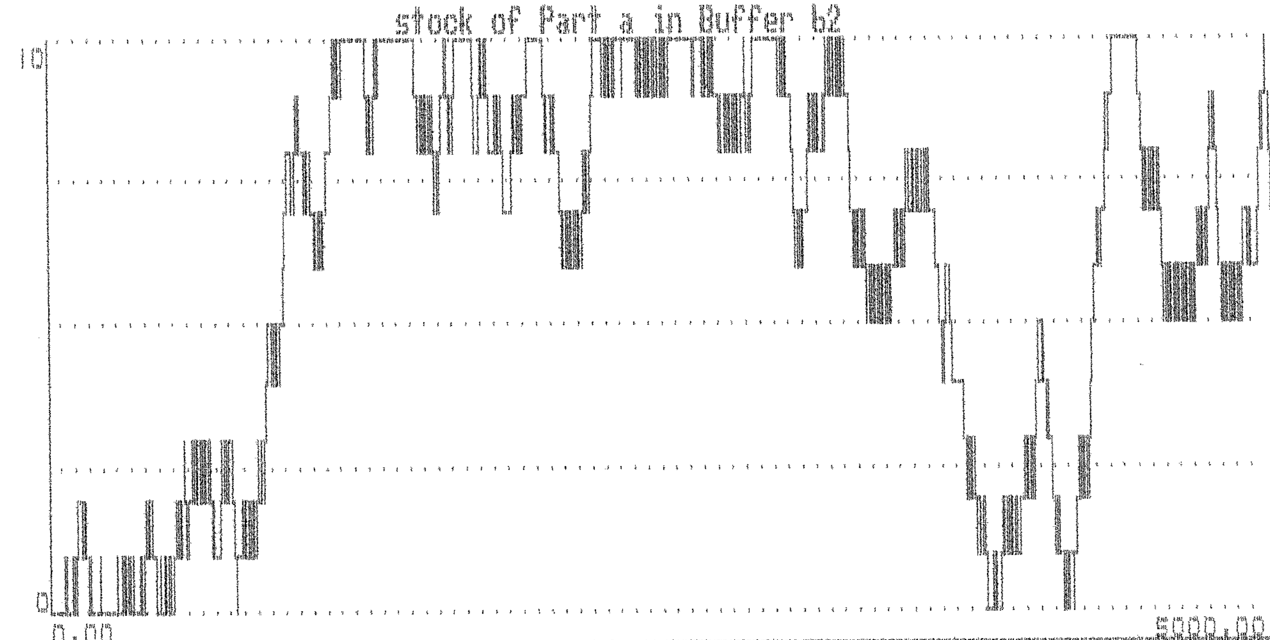
<trace> <chart> <PLOT>

stock of Part a in Buffer b1



1 HELP	2	3 set run control	4 start trace	5 start chart	6 stop plot	7 continue run	8 back to main

Paused at 5000.00 <trace> <chart> <PLOT>



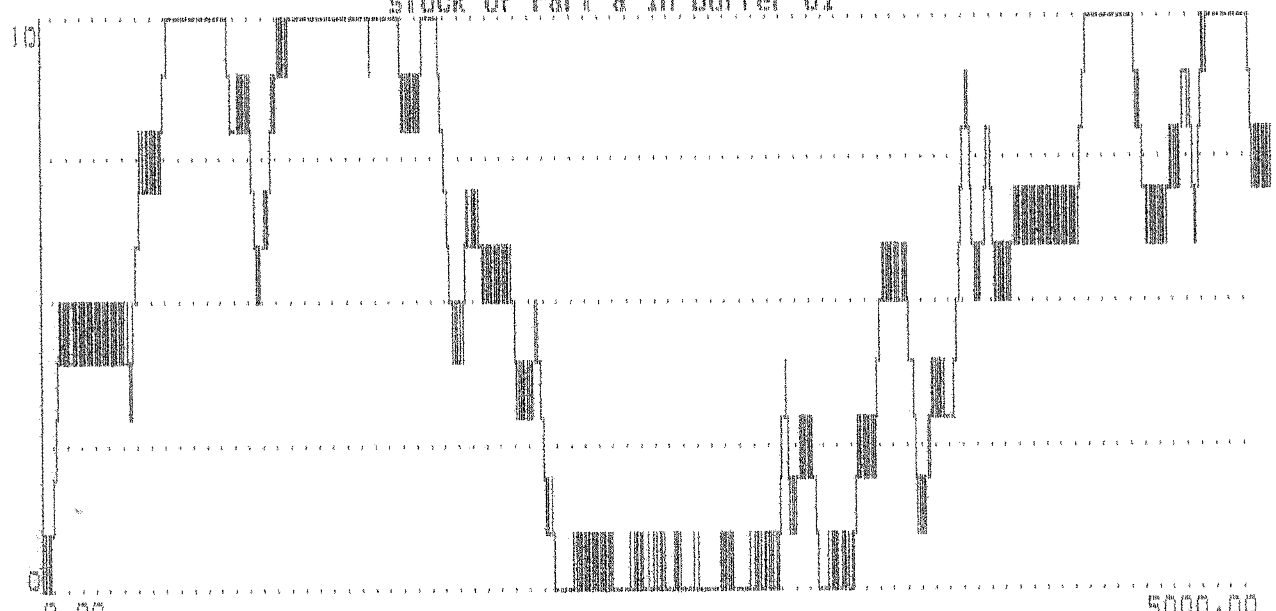
1 HELP	2	3 set run control	4 start trace	5 start chart	6 stop plot	7 continue run	8 back to main

Paused at 10000.00 <trace> <chart> <PLOT>

1 HELP	2	3 set run control	4 start trace	5 start chart	6 stop plot	7 continue run	8 back to main

$$\frac{\text{processing time} * \text{buffer size}}{\text{downtime}} = 5$$

Paused at 5000.00 <trace> <chart> <PLOT>

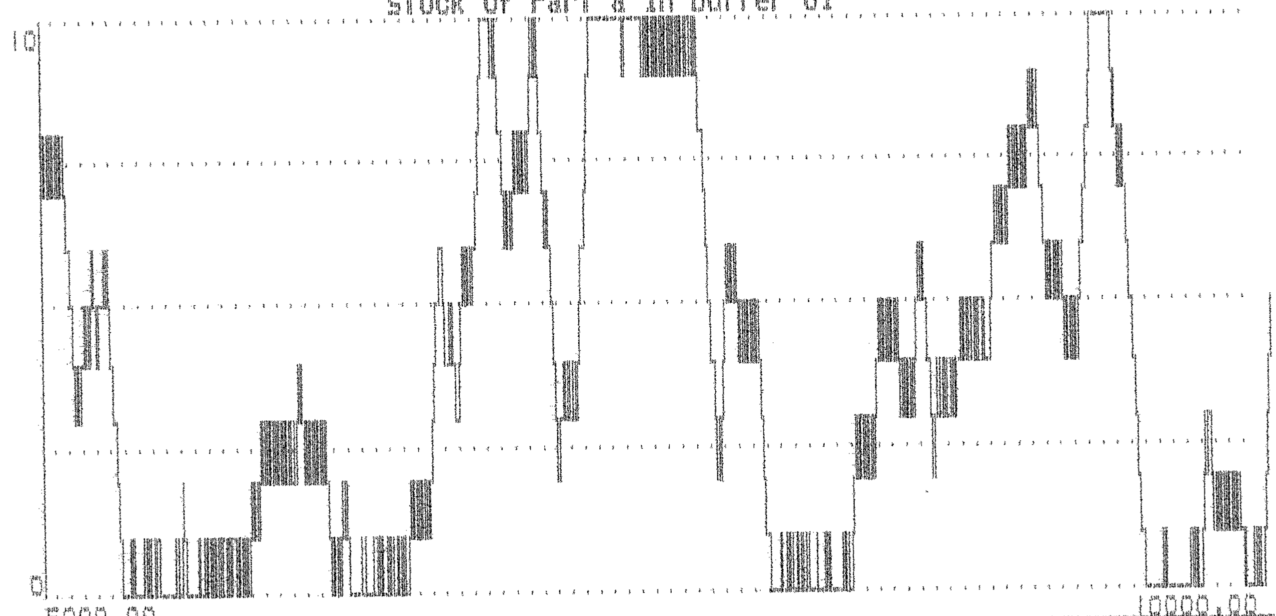


1 HELP	2	3 set run control	4 start trace	5 start chart	6 stop plot	7 continue run	8 back to main

Paused at 10000.00

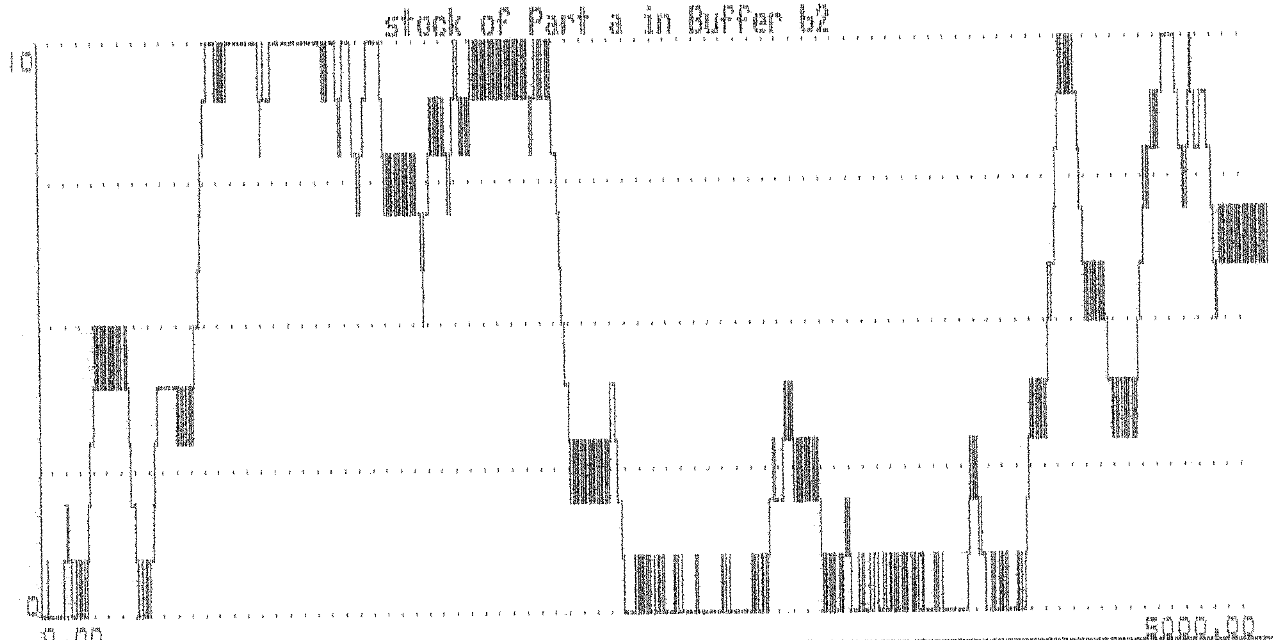
<trace> <chart> <PLOT>

stock of Part a in Buffer b1



1 HELP	2	3 set run control	4 start trace	5 start chart	6 stop plot	7 continue run	8 back to main

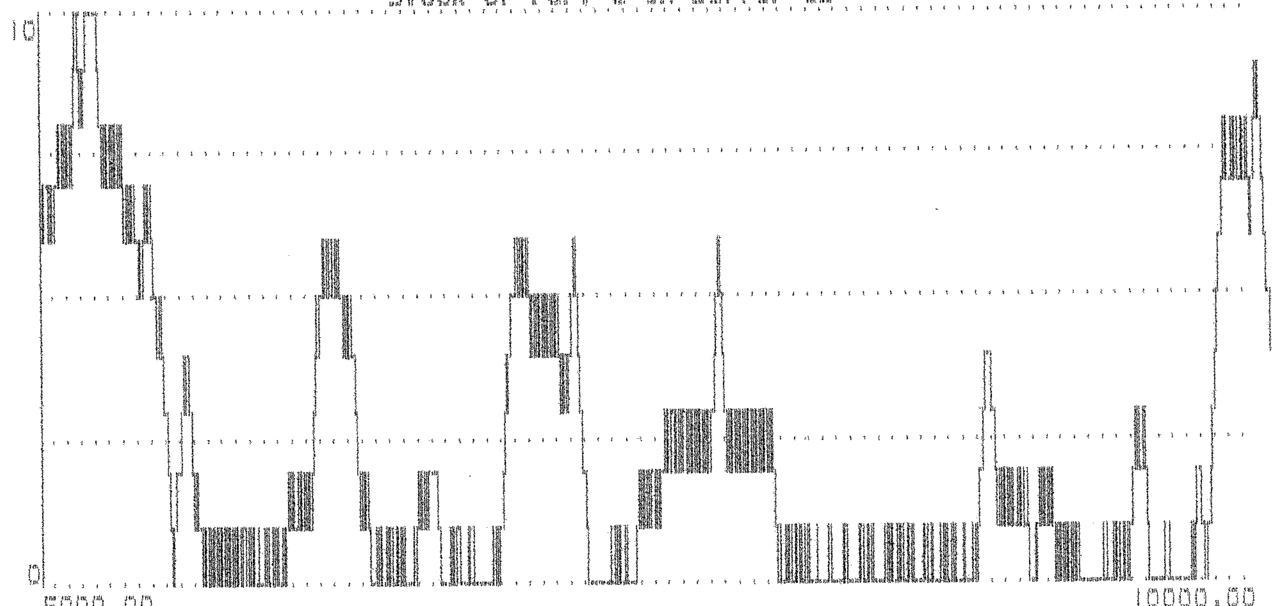
Paused at 5000.00 <trace> <chart> <PLOT>



1 HELP	2	3 set run control	4 start trace	5 start chart	6 stop plot	7 continue run	8 back to main

Paused at 10000.00 <trace> <chart> <PLOT>

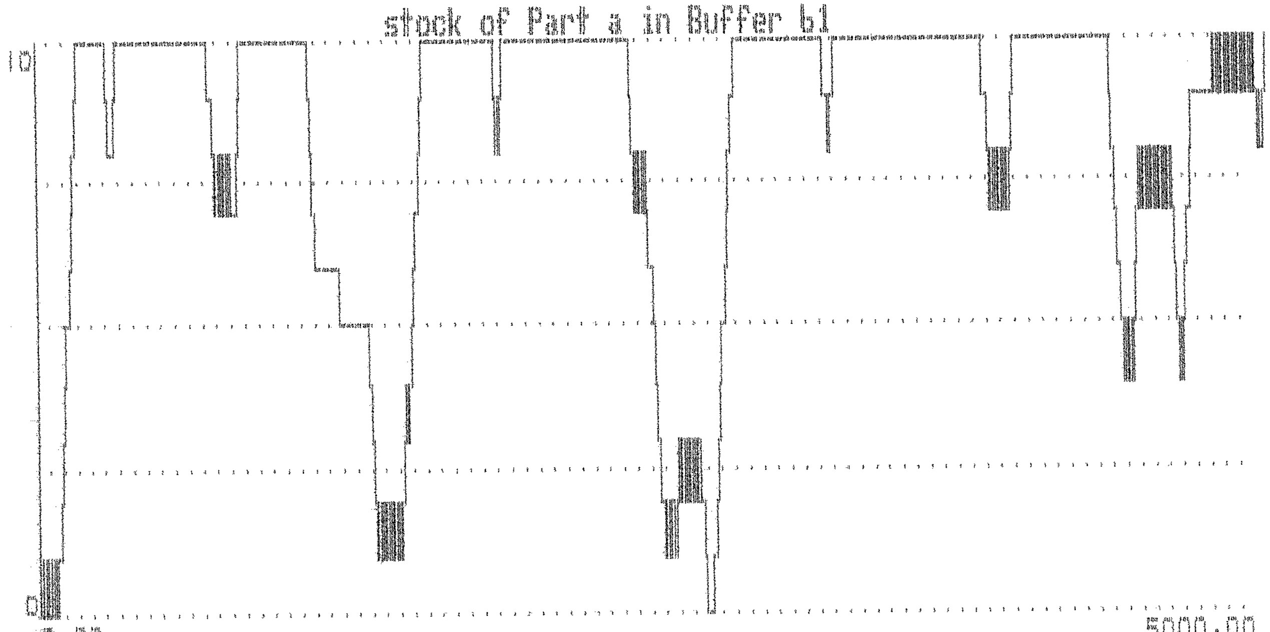
stock of Part a in Buffer b2



1 HELP	2	3 set run control	4 start trace	5 start chart	6 stop plot	7 continue run	8 back to main

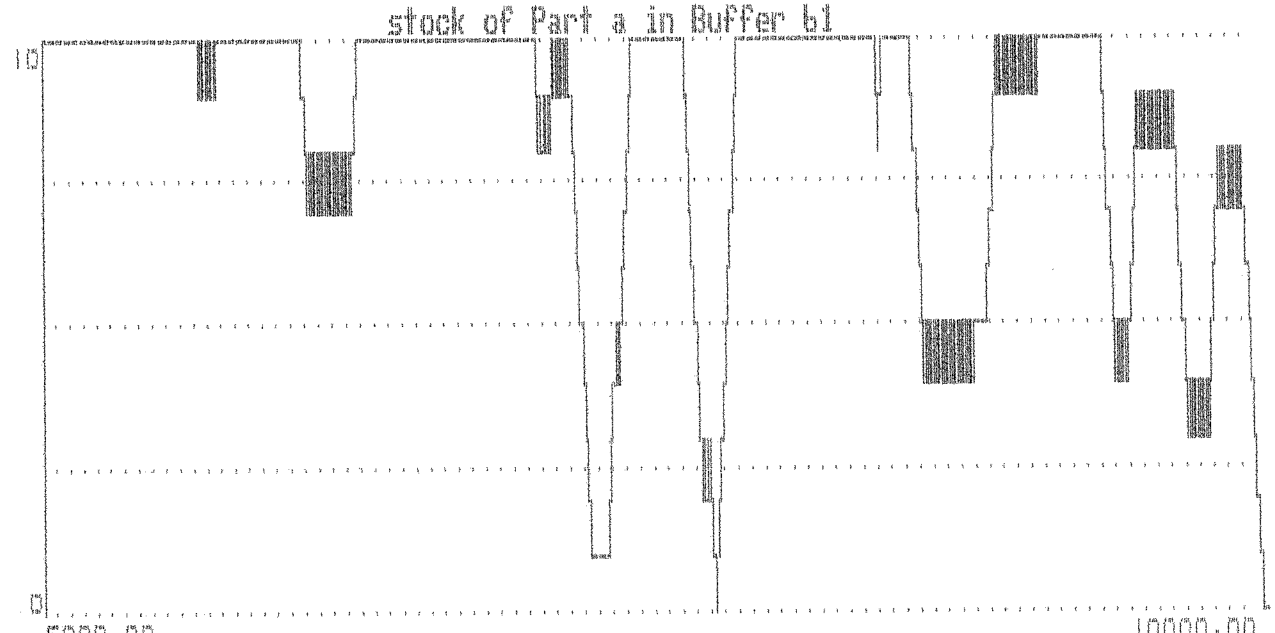
$$\frac{\text{processing time} * \text{buffer size}}{\text{downtime}} = 2$$

Paused at 5000.00 <trace> <chart> <PLOT>



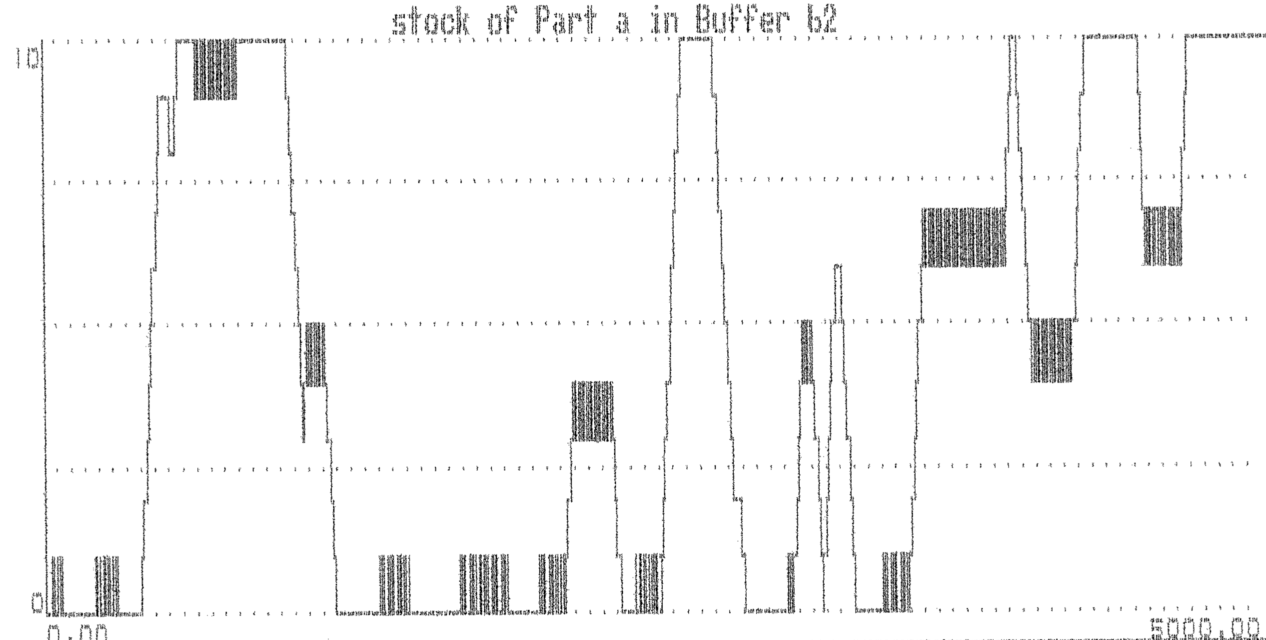
1 HELP	2	3 set run control	4 start trace	5 start chart	6 stop plot	7 continue run	8 back to main

Paused at 10000.00 <trace> <chart> <PLOT>



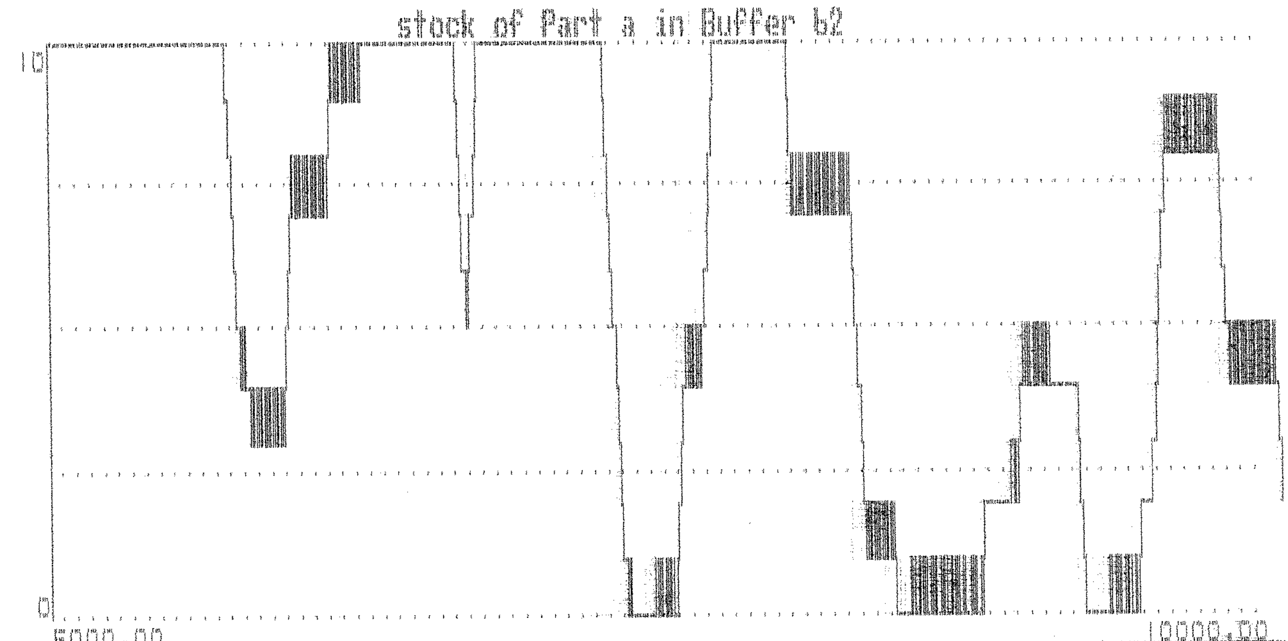
HELP		set run	start	start	stop	continue	back to
1	2	3 control	4 trace	5 chart	6 plot	7 run	8 main

Paused at 5000.00 <trace> <chart> <PLOT>



1 HELP	2	3 set run control	4 start trace	5 start chart	6 stop plot	7 continue run	8 back to main

Paused at 10000.00 <trace> <chart> <PLOT>

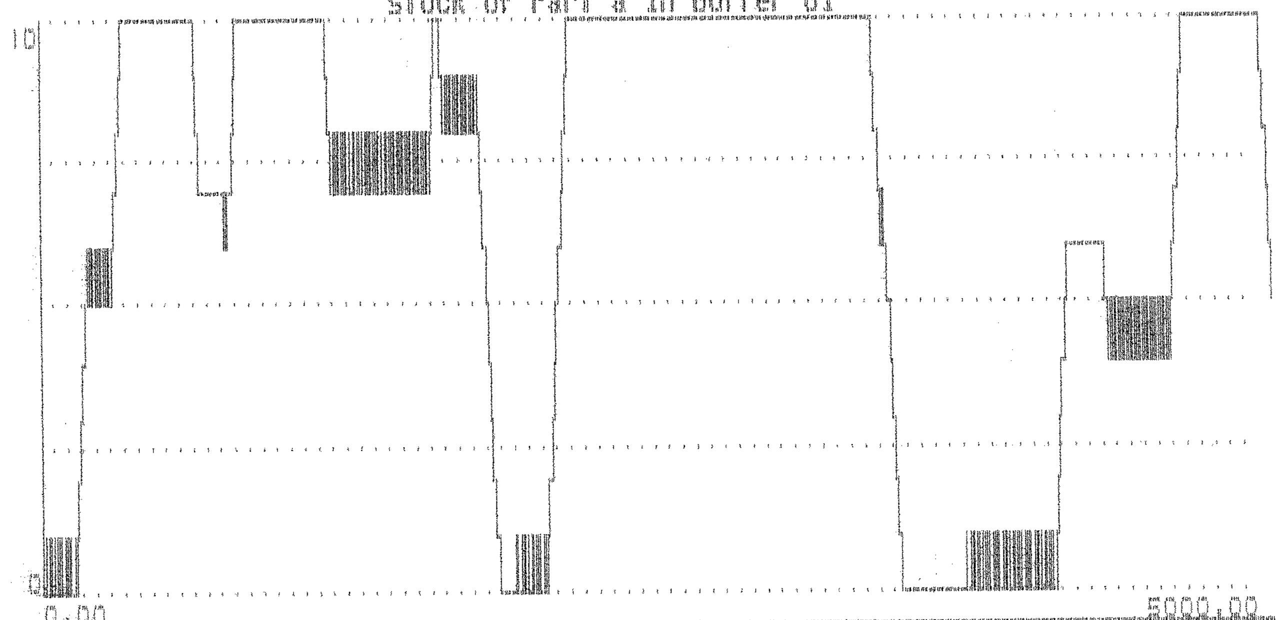


1 HELP	2	3 set run control	4 start trace	5 start chart	6 stop plot	7 continue run	8 back to main

$$\frac{\text{processing time} * \text{buffer size}}{\text{downtime}} = 1$$

Paused at 5000.00 <trace> <chart> <PLOT>

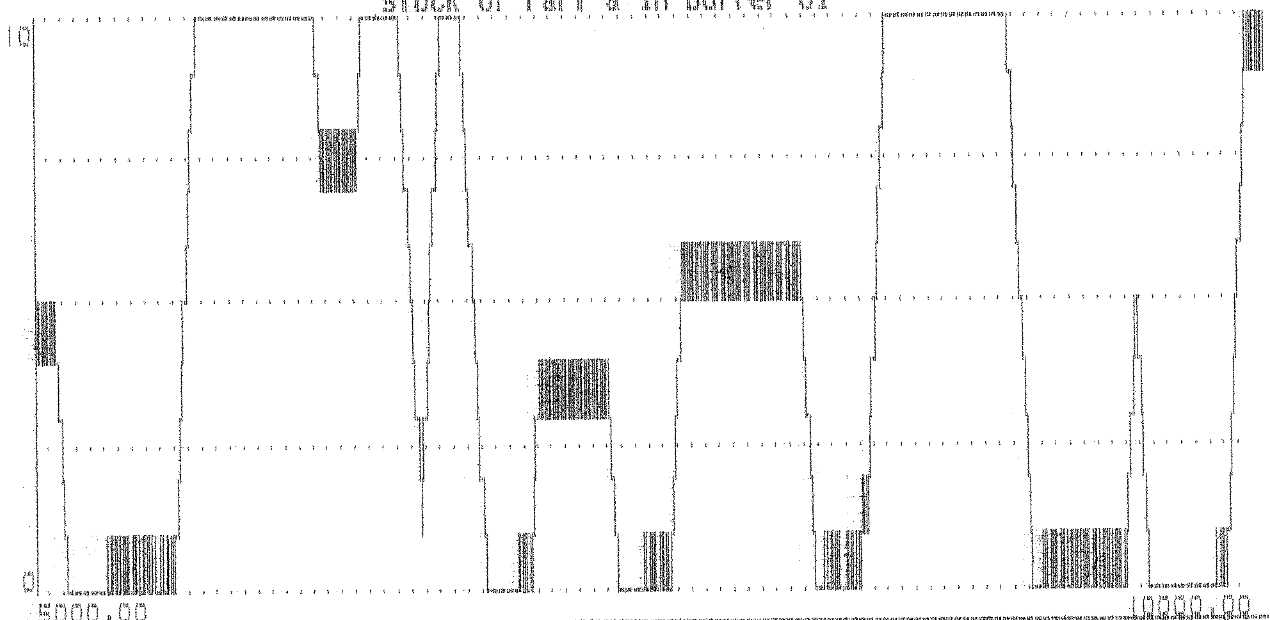
stock of Part a in Buffer b1



1 HELP	2	3 set run control	4 start trace	5 start chart	6 stop plot	7 continue run	8 back to main

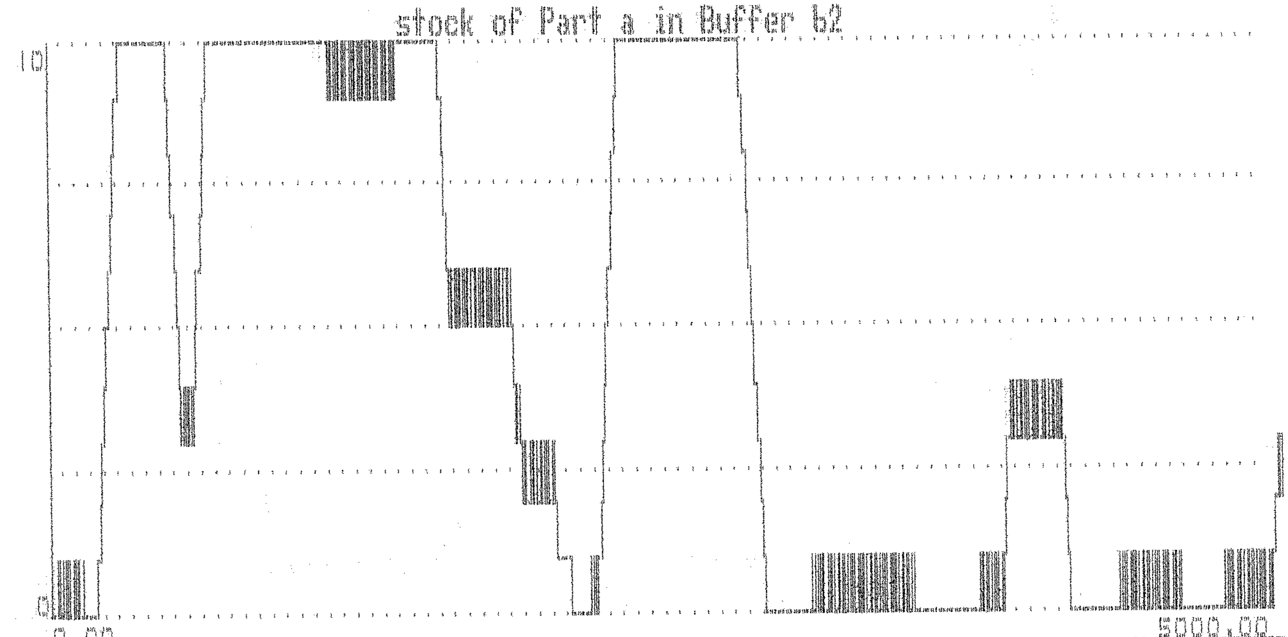
Paused at 10000.00 <trace> <chart> <PLOT>

stock of Part a in Buffer b1



HELP		set run	start	start	stop	continue	back to
1	2	3 control	4 trace	5 chart	6 plot	7 run	8 main

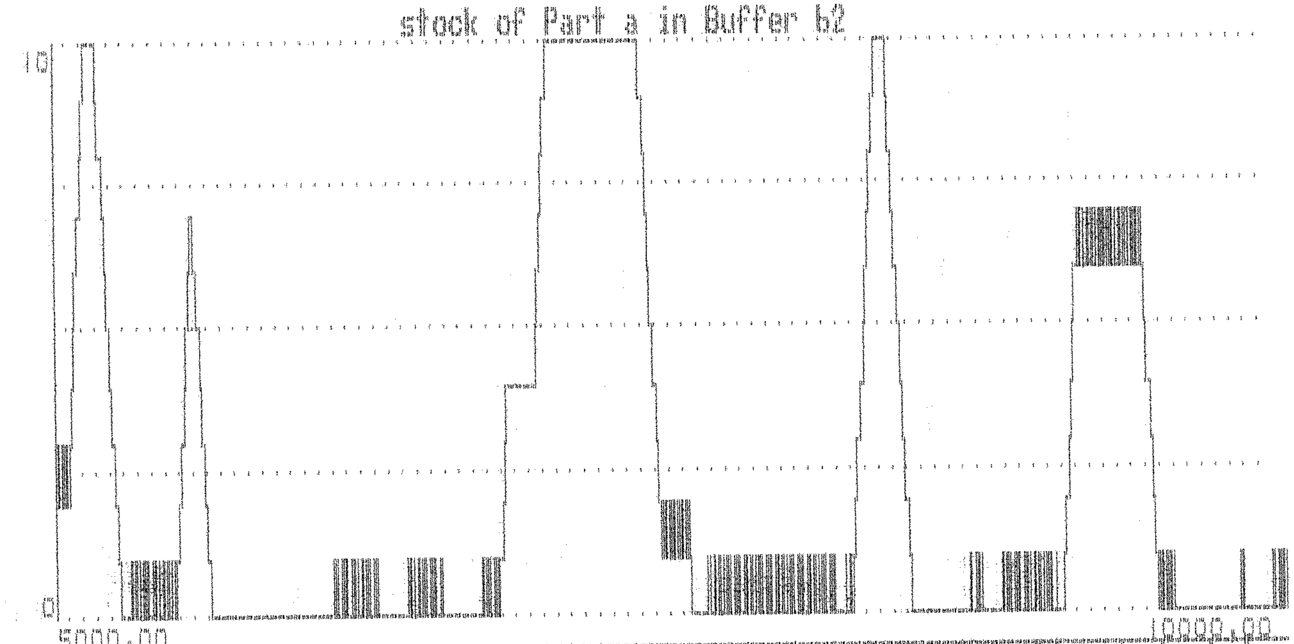
Paused at 5000.00 <trace> <chart> <PLOT>



1 HELP	2	3 set run control	4 start trace	5 start chart	6 stop plot	7 continue run	8 back to main

Paused at 10000.00

<trace> <chart> <PLOT>



1 HELP	2	3 set run control	4 start trace	5 start chart	6 stop plot	7 continue run	8 back to main

most points represent an essentially infinite capacity.

Bibliography

(1) Joneja, D., W. L. Maxwell, "Buffer Placement in Sequential Production Lines: Considerations of Processing Time Variability," TR #700, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY.

(2) Muth, E. J., "The Reversibility Property of Production Lines," Management Science, Vol 25, No 2 pp. 152-158

(3) Devore, J. L., Probability & Statistics for Engineering and the Sciences, Brooks/Cole Publishing Company; Monterey, California 1982.

(4) Malathronas, J. P., J. D. Perkins, R. L. Smith; "The Availability of a System of Two Unreliable Machines Connected by an Intermediate Storage Tank," IIE Transactions, Vol 15, No 3 pp. 195-201.