The quasi-simultaneous finishing of work orders on a flexible automated manufacturing cell in a job shop

INT. J. PROD. RES., 1990, VOL. 28, NO. 9, 1635 1655

The quasi-simultaneous finishing of work orders on a flexible automated manufacturing cell in a job shop

C. W. G. M. DIRNE†

The introduction of flexible automated machine tools changes some of the characteristics of the production system in component manufacturing shops. Not only are machines made more universal by the use of flexible production automation, but it also is becoming more and more possible to change the set up of a machine without loss of productive hours. However, because of a limitation in the number of (often expensive) fixtures available per type of fixture, work orders are often finished more or less simultaneously, rather than in a sequential manner.

This paper considers the logistic implications of these changes. In particular, the consequences of the quasi-simultaneous finishing of work orders for the throughput time will be analyzed. The paper shows that the work order lot size is a major factor in the determination of the throughput time per work order.

1. Introduction

The use of flexible automation in the manufacturing process is growing considerably. Studies have shown that in a few years time the use of flexibly automated manufacturing systems has tripled in countries like the USA, France, FRG and UK (see e.g. Bessant and Haywood (1986) and Spur and Mertins (1982)).

In this paper we are especially interested in the logistic implications of the use of flexibly automated manufacturing systems in component manufacturing shops. The reason for this limitation is that previous studies have shown that assumptions about the environment are very important for problem definition (Dirne 1987).

As we will see, a new phenomenon appears in the sequencing of work orders. Instead of finishing a work order completely before starting with a new one, several work orders are operated on more or less simultaneously. This phenomenon is the central subject of our study. In Section 2 we will define the various types of flexible automation in component manufacturing shops from a logistic point of view. For each type the major logistic consequences will be mentioned. The next section concentrates on the above mentioned phenomenon of the quasi-simultaneous finishing of work orders. Some analytical approximations will be made. In order to study the problem in more detail and to validate the conclusions from the analytical approximations and intuitive hypotheses, some simulation experiments have been executed. Section 4 reports about the results of these experiments. In section 5 the major conclusions are given.

2. Flexible automation in the manufacturing process

In order to be able to categorize the various types of flexible automation in component manufacturing shops, we will have a closer look at the goods flow.

Revision received January 1990.

† Research Assistant, Eindhoven University of Technology, Faculty of Industrial Engineering, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

Generally speaking we may distinguish three categories of transformation of the status of a workpiece:

(1) transformation in physical appearance (e.g. in shape, size or composition);
(2) transformation of place or position; and
(3) transformation of time.

The first kind of transformation could be a real change in physical appearance (e.g. a machining operation), or merely a change in the information about the physical appearance (i.e. a testing operation). The second category of transformation can be split up into transporting activities (being transformations of place between two locations, e.g. two machines, that should be distinguished from a production control point of view) and material handling activities (being transformations of place or position at one location, e.g. the fixturing of a workpiece on a pallet). Also the last category of transformations (transformation of time) can be split up into two different transformation types, depending on whether this time transformation is explicit controlled or not. In the former case we can speak of inventory in controlled stock points under the responsibility of a goods flow control function (Bertrand 1985 a). In the latter case we may speak of storage of work in progress as a result of an imbalance between two production stages. Figure 1 presents a simplified goods flow for a component manufacturing shop using the concepts mentioned above. Using this goods flow, we can distinguish a number of types of flexible automation in the manufacturing process that are interesting from a production control point of view. As we will see, each type has some consequences for the production control problem. A major issue in this problem is the work order throughput time (i.e. the time elapsing between the release of the work order to the component manufacturing shop and the finishing of the last

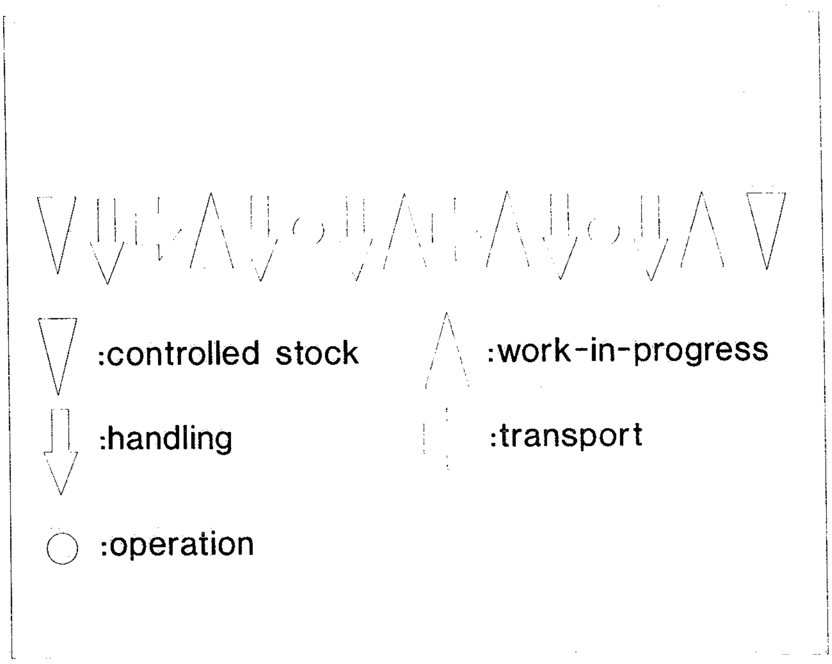


Figure 1. Transformations in a simplified goods flow.

operation on the work order). The mean throughput time of a work order i at a work centre j in a job shop can be expressed as the sum of the time spent on the machine (denoted by s_{ij}) and the waiting time at the work centre (denoted by w_j). Usually s_{ij} can be expressed as:

$$s_{ij} = u_{ij} + Q_i v_{ij} \tag{1}$$

where

u_{ij} = set up time for order i at work centre j;
Q_i = lot size of order i;
v_{ij} = processing time of one part of order i at work centre j.

We know from queueing theory that w_j is linearly related to the average of all s_{ij} at work centre j (denoted by s_j):

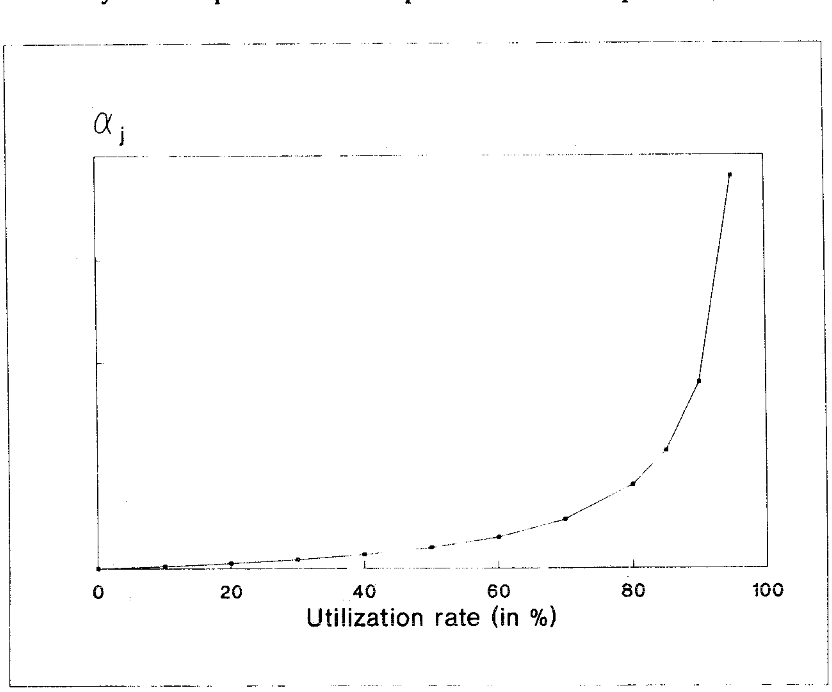
$$w_j = \alpha_j s_j \tag{2}$$

where α_j is the mean number of work orders waiting at centre j (Bertrand 1985 b).

One important factor influencing α_j is the utilization rate of the work centre. In Fig. 2 a typical relation between α_j and the utilization rate of work centre j in a job shop is drawn.

We can distinguish the following different types of flexibility automated manufacturing systems (see also Fig. 3).

Manufacturing Centre (*MC*): a computer controlled machine tool capable of performing more types of operations (instead of being specialized in only one type of operation). An MC enables the reduction of the number of operations in a routing. This leads not only to a simplification of the production control problem, but also to a



igure 2. Relation between utilization rate and α_j in a job shop.

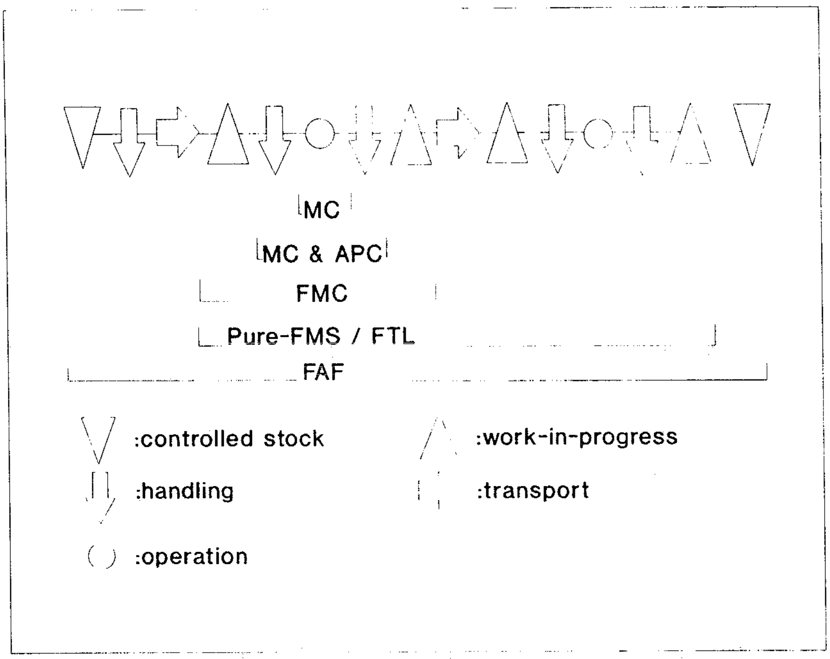


Figure 3. Types of flexibly automated manufacturing systems in component manufacturing shops.

reduction in the number of set ups to be made per work order. Very likely such a reduction will lead to a reduction in the total set up time required per work order ($\Sigma_j u_{ij}$). We know from (1) and (2) that both the time spent on machines ($\Sigma_j s_{ij}$), and the waiting time at work centres ($\Sigma_j w_j$) will be reduced. Therefore, assuming no change in α_j, the throughput time will be reduced proportionally.

Apart from throughput time reduction because of a reduction in the number of operations, a further reduction may be realized due to the fact that by using universal machines we may be able to create parallel resources. Again, we know from previous studies (Kleinrock 1976) that this may reduce α_j and thus the throughput time considerably.

MC with Automatic Pallet Changer (APC): a manufacturing centre that is expanded with an automatic workpiece changer, thus enabling automation of a part of the handling activities as well; since workpieces often are placed on a pallet, an automatic pallet changer is frequently being used. Together with an APC often an Automatic Tool Changer (ATC) and a tool magazine located at the machine are included.

In the case of an MC with an APC it becomes possible to perform set ups on the pallet changer and thus in front of the machine instead of on the machine itself. By doing so a further machine set up time reduction will be established. While the machine is operating on a workpiece, a second workpiece can be loaded on a pallet on the APC. As soon as the machine has finished the operation, these two workpieces will be exchanged. The machine starts on operating the second workpiece and the operator can unload the finished workpiece and load another unfinished workpiece. Due to the set up time reduction the throughput time will be reduced and it may become possible to reduce the work order lot size. This lot size reduction in its turn again will lead to a

ORDER CHARACTERISTICS

Order:	A	B	C	D
Arrival time:	0	0	3	6
Workpiece:	a	b	c	d
Operation time / workpiece:	1	1	1	1
Lot size:	4	3	2	1

WORK ORDER SEQUENCE

Time: 0 1 2 3 4 5 6 7 8 9 10

Sequence MC:	a	a	a	a	b	b	b	c	c	d
Sequence MC&APC:	a	b	a	b	a	b	a	c	d	c

Figure 4. Sequencing example for the case of MC and MC with APC.

further throughput time reduction at the MC. However, one has to be careful that the reduction in throughput time at the MC is not lost at other workstations. Because of the increased number of set ups to be made on these other workstations, the utilization rates of these workstations may increase and thus affect total throughput time in a negative way (see Fig. 2).

The introduction of an APC may also result in a new phenomenon. Since fixtures often are rather expensive, the number of fixtures available per type of fixture can be limited. If this number is limited to only one, work orders will not any longer be in operation one after the other, but more or less simultaneously! Due to the mechanism of the changing of pallets as described above, the machine will alternately operate on workpieces of two work orders. This phenomenon is illustrated in the example of Fig. 4 and is the central subject of this paper. We will refer to it as the quasi-simultaneous finishing of work orders.

Extra consideration should be given to the possibility of creating new constraints caused by the combination of work orders in hand (e.g. a limitation in the number of tool pockets available, see Stecke (1983)).

Flexible automated Manufacturing Cell (*FMC*): one or more interconnected MCs that are not only expanded with APC and ATC equipment, but also with a pallet pool.

This enables the storage of several pallets on the system. Each workpiece/pallet combination visits only one MC, after which the workpiece can be unloaded. A single FMC contains only one MC, a multiple FMC more.

An FMC makes it possible to reduce the throughput time even further. This manufacturing system is less dependent on the continuous presence of an operator. In fact, it is able to produce unmanned for several hours (depending on the number of pallets available, the manufacturing time per workpiece and the number of workpieces placed on one pallet). This 'extra' capacity results in a reduction in throughput time and may create some volume flexibility. However, not all the workpieces may be qualified for unmanned production (both from a technical and logistical point of view). If that is the case, this extra constraint should be taken into consideration.

An FMC enables work orders to be sorted according to urgency: each time a pallet is moved from the machine a pallet containing workpieces for an urgent work order can be given highest priority. Further, the above mentioned phenomenon of the quasi-simultaneous finishing of work orders plays an even more important role for these systems. The number of work orders in operation at the same time can grow as large as the number of pallets available!

Finally, in the case of a multiple FMC it may be possible to exchange tools between the machines automatically. Since this tool exchange may take some time, it might be wise to take this into consideration when sequencing the work orders (ElMaraghy 1985).

Pure Flexible automated Manufacturing System (*pure-FMS*) *and a Flexible automated Transfer Line* (*FTL*): a number of computer controlled machines connected by APC, ATC and pallet storage equipment in such a way that several routings between the machines are possible. In the case of an FTL, the number of routings is limited due to the physical constraints of the transportation system.

In the case of a pure-FMS or an FTL, the production control problem may become more complicated due to balancing and blocking problems. This is a complicated phenomenon, though well known in literature (see e.g. Buzacott (1984) and Stecke and Solberg (1985)). Also it might be necessary to have a closer look at the transporting system and its constraints and utilization rate. These consequences are added to the ones mentioned above.

Flexible Automated Factory (*FAF*): one or more pure-FMSs or FTLs connected with an Automatic Storage and Retrieval System (AS/AR) that enables the automatic transport and storage to and from controlled stock points.

An FAF finally includes the warehouses in the automated manufacturing system. The main difficulties in doing so are the handling and accuracy of (logistic) information and the limitation in picking capacity.

We may conclude that some of the effects are predictable, while others are now being investigated in several other studies. This study will be concentrated on the phenomenon of working on several work orders at the same time (quasi-simultaneous), assuming a restriction on the number of fixtures available per type of fixture and a certain work order lot size.

3. Quasi-simultaneous finishing of work orders

In the previous section we have seen that with the introduction of an automated pallet changer and the use of a pallet pool, the way work orders are sequenced may be changed. Due to a limitation in the number of available fixtures per type of fixture and a certain lot size per work order, machines may operate alternately on workpieces of two or more work orders. In this section we will define this problem more precisely and give some thoughts about the possible effects.

3.1. *The problem*

The problem can be defined as follows. A flexible automated manufacturing system is part of a larger production department with job shop characteristics. Jobs or work orders containing several workpieces arrive at the flexible automated manufacturing system in a more or less random way, due to the fact that they may come from several different workstations. The number of fixtures available per type of fixture is limited (and often only one). Figure 5 illustrates the way workpieces are loaded and unloaded in case the number of fixtures available per type of fixture is limited to one. At time t_0 a special fixture 1 and one or more workpieces of a job A are loaded on one pallet. This set of workpieces fixtured together on one pallet will be called a *runbatch*. The pallet will be loaded with the first runbatch of job A as soon as the pallet becomes free, job A has highest priority and the necessary tools for job A are or can be loaded in the tool magazine. Due to the fact that for most jobs the job lot size will be larger than the runbatch lot size (in this example the job contains three runbatches), and due to the limitation in the number of fixtures, the operator has to wait on the return of the first runbatch in order to be able to load the next runbatch of job A (time t_1). If necessary and available, another pallet might be loaded with the first runbatch for one or more operations in a different type of fixture (e.g. fixture 2). The first runbatch is finished at time t_2. The job is finished as soon as all necessary operations on its runbatches are finished and the runbatches are unloaded (at time t_3). It is important to stress the fact

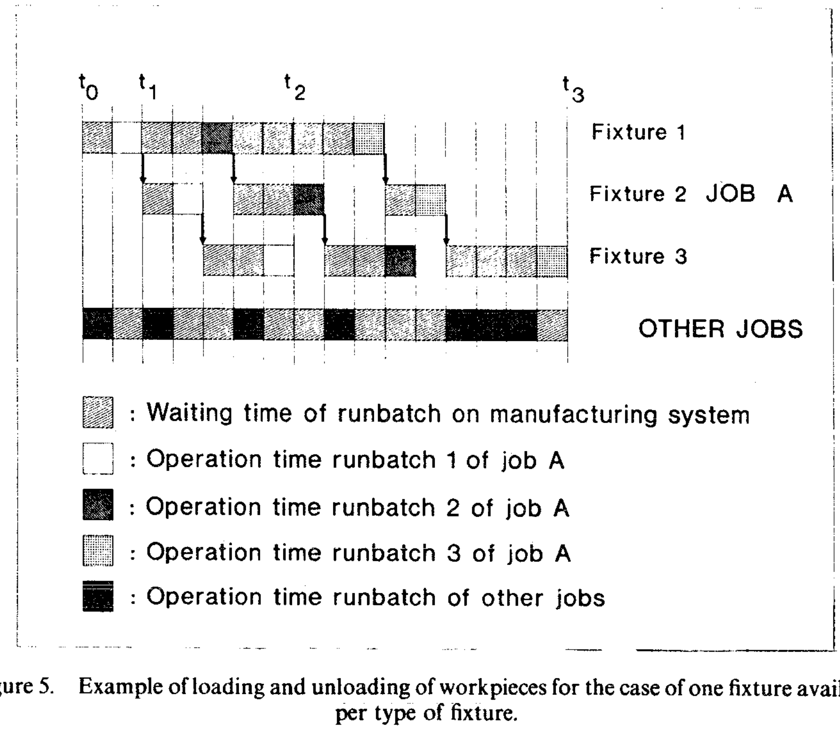


Figure 5. Example of loading and unloading of workpieces for the case of one fixture available per type of fixture.

that the manufacturing system contains one or more pallets loaded with runbatches that often will belong to different jobs.

What effects will quasi-simultaneous finishing of work orders have on throughput time? Both a longer average throughput time and a larger variation in throughput time are undesirable from the point of view of production planning and control.

3.2. *The model*

We assume a Poisson arrival process. For clarity we will limit our study to jobs with only one operation to be performed, thus eliminating the balancing and blocking problems of, for instance, a pure-FMS and the possibility of having operations to be performed in different types of fixtures. We also assume that the number of tool pockets available is sufficient, that the number of fixtures available per type of fixture equals one and that loading/unloading and handling times are negligible compared with the operation times. Sequencing decisions are based on the FIFO priority rule, where the arrival time of the job is used to decide for which job a pallet should be reserved, and the time of loading of the runbatch is used inside the manufacturing system to decide which runbatch should be operated on next. The remaining model is primarily aimed at the study of the main phenomenon of interest. The model is schematically drawn in Fig. 6.

3.3. *The approximation*

We define job class i as the set of jobs with the same number of runbatches (q_i) and the same operation time distribution. Let $\mathbf{c}_i$ be the time a runbatch of a job belonging to class i spends in the manufacturing system. Furthermore, let N be the number of pallets, let m be the number of machines in the system and let $\mathbf{n}$ be the number of jobs in and in front of the system. We may assume that $N \geqslant m \geqslant 1$. The expected operation time

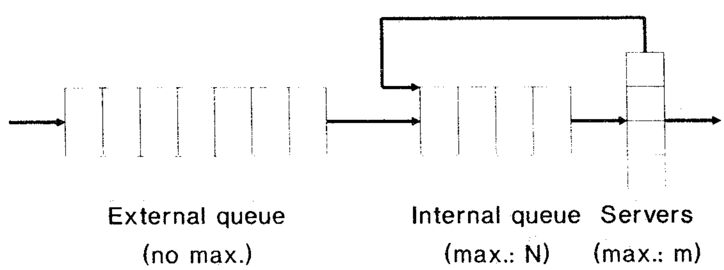


Figure 6. The model.

required per runbatch of job of class i is denoted by $E(\mathbf{o}_i)$, whereas the expected operation time over all runbatches will be denoted by $E(\mathbf{o})$. We assume an average arrival rate of λ.

Assuming a number of jobs in the system, we can get an estimate of $\mathbf{c}_i$ (denoted by $E(\mathbf{c}_{in})$):

$$E(\mathbf{c}_{in}) = \begin{cases} E(\mathbf{o}_i) & (n \leqslant m), \\ E(\mathbf{o}_i) + E(\mathbf{o})(n-m)/m & (m \leqslant n \leqslant N), \\ E(\mathbf{o}_i) + E(\mathbf{o})(N-m)/m & (n \geqslant N). \end{cases} \tag{3 a}$$

An estimate of $\mathbf{c}_i$ can be given using the relation:

$$E(\mathbf{c}_i) = \sum_{n=1}^{\infty} \{E(\mathbf{c}_{in}) p(\mathbf{n}=n|n \geqslant 1)\}, \tag{3 b}$$

with $p(\mathbf{n}=n|n \geqslant 1)$ being the probability of n jobs being present in or in front of the system, given the fact that $n \geqslant 1$ (n is an integer). Combining (3 *a*) and (3 *b*) we get:

$$E(\mathbf{c}_i) = E(\mathbf{o}_i) + E(\mathbf{o}) \sum_{n=m}^{N-1} \{(n-m/m) p(\mathbf{n}=n|n \geqslant 1)\} + E(\mathbf{o})(N-m/m) \sum_{n=m}^{\infty} p(\mathbf{n}=n|n \geqslant 1).$$

We know that $p(\mathbf{n}=n|n \geqslant 1) = p(\mathbf{n}=n)/p(\mathbf{n} \geqslant 1) = p(\mathbf{n}=n)/\{1-p(\mathbf{n}=0)\}$. Thus:

$$E(\mathbf{c}_i) = E(\mathbf{o}_i) + (E(\mathbf{o})/1 - p(\mathbf{n}=0)) \times \left\{ \sum_{n=m}^{N-1} (n/m) p(\mathbf{n}=n) + \sum_{n=N}^{\infty} (N/m) p(\mathbf{n}=n) - \sum_{n=m}^{\infty} p(\mathbf{n}=n) \right\} \tag{4}$$

In the case of $m=1$ (a single FMC or an MC with APC), we get:

$$E(\mathbf{c}_i) = E(\mathbf{o}_i) + E(\mathbf{o}) \{E(\mathbf{x})/\rho - 1\} \tag{5}$$

where

$$E(\mathbf{x}) = \sum_{n=1}^{N} n p(\mathbf{n}=n) + \sum_{n=N+1}^{\infty} N p(\mathbf{n}=n),$$

$$\rho = 1 - p(\mathbf{n}=0) \qquad (= \lambda E(\mathbf{q}) E(\mathbf{o})).$$

In the case of $m > 1$ (multiple FMC) and identical machines we may use eqn. (5) if:

(1) all runbatches of a job will go to the same machine, e.g. because the necessary tools are only loaded in the tool magazine of one machine and no automatic exchange of tools is possible;
(2) no preference is given in advance to a certain machine.

If we are able to get an estimate of $E(\mathbf{c}_i)$, we can estimate the total time a job of class i will have runbatches on the system by

$$E(\mathbf{t}_{i,1}) = q_i E(\mathbf{c}_i), \tag{6}$$

where $\mathbf{t}_{i,1}$ is the time between loading of the first runbatch and unloading of the last runbatch of a job of class i.

The total estimated throughput time of a job ($E(\mathbf{t}_i)$) will be

$$E(\mathbf{t}_i) = E(\mathbf{t}_{i,1}) + E(\mathbf{t}_{i,2})$$

where $E(\mathbf{t}_{i,2})$ is the expected waiting time of a job of class i in front of the manufacturing system. In fact, the expected waiting time in front of the system is, under the assumption of a FIFO priority rule, independent of i. Therefore we may write $E(\mathbf{t}_2)$ instead of $E(\mathbf{t}_{i,2})$, thus

$$E(\mathbf{t}_i) = q_i E(\mathbf{c}_i) + E(\mathbf{t}_2) \tag{7}$$

The expected throughput time over all jobs ($E(\mathbf{t})$) will then be:

$$E(\mathbf{t}) = \sum_{i=1}^{I} \{E(\mathbf{t}_i) y_i\} \tag{8}$$

where

y_i = percentage of jobs belonging to class i;
I = number of job classes.

Where eqn. (5) can be used, we can get a good estimate of $E(\mathbf{t}_i)$ in real life by measuring $E(\mathbf{x})$ (i.e. the mean number of pallets occupied) and $E(\mathbf{t}_2)$. However, in the case $m > 1$ and where there is the possibility of exchanging tools or a redundancy in tools, we have to calculate $p(\mathbf{n} = n)$. Avi-Ithzak and Heyman (1973) give an approximation of $E(\mathbf{t})$ by calculating $p(\mathbf{n} = n)$, assuming that $\mathbf{c}_n$ (the time a certain job will have runbatches in the system if n jobs are present) has a negative exponential distribution. In the case of identical machines we may write:

$$E(\mathbf{c}_n) = \begin{cases} E(\mathbf{o}) & (n \leqslant m), \\ (n/m)E(\mathbf{o}) & (m \leqslant n \leqslant N), \\ (N/m)E(\mathbf{o}) & (n \geqslant N). \end{cases}$$

We know from (Avi-Ithzak and Heyman 1973):

$$p(\mathbf{n} = n) = \begin{cases} p(\mathbf{n} = 0)\lambda^n/(\gamma_1 \gamma_2 \cdots \gamma_n) & (n = 1, 2, \ldots, N), \\ p(\mathbf{n} = N)(\lambda/\gamma_N)^{n-N} & (n = N, N+1, \ldots), \end{cases}$$

where γ_n is the average rate of completions of jobs, or:

$$\gamma_n = \begin{cases} n/(E(\mathbf{q})E(\mathbf{c}_n)) & (n \leqslant N), \\ N/(E(\mathbf{q})E(\mathbf{c}_n)) & (n \geqslant N). \end{cases}$$

Therefore:

$$p(\mathbf{n} = n) = \begin{cases} p(\mathbf{n} = 0)\lambda^n E(\mathbf{q})^n E(\mathbf{o})^n/n! & (n = 0, 1, \ldots, m), \\ p(\mathbf{n} = 0)\lambda^n E(\mathbf{q})^n E(\mathbf{o})^n/(m!\, m^{n-m}) & (n = m, m+1, \ldots). \end{cases} \tag{9}$$

We know that

$$\sum_{n=0}^{\infty} p(\mathbf{n} = n) = 1$$

thus:

$$\begin{aligned} p(\mathbf{n} = 0) &= \left\{ \sum_{n=0}^{m-1} (\lambda E(\mathbf{q})E(\mathbf{o}))^n/n! + \sum_{n=m}^{\infty} (\lambda E(\mathbf{q})E(\mathbf{o}))^n/(m!\, m^{n-m}) \right\}^{-1} \\ &= \left\{ \sum_{n=0}^{m-1} (A^n/n!) + mA^m/((m-A)m!) \right\}^{-1}, \end{aligned} \tag{10}$$

where $A = \lambda E(\mathbf{q})E(\mathbf{o})$.

In order to calculate $E(\mathbf{t})$ we may use (Avi-Ithzak and Heyman 1973):

$$E(\mathbf{t})=(1/\lambda)\sum_{n=0}^{\infty} np(\mathbf{n}=n) \tag{11}$$

In order to calculate $E(\mathbf{t}_2)$ we may use (Avi-Ithzak and Heyman 1973):

$$E(\mathbf{t}_2)=\gamma_N p(\mathbf{n}=N)/(\gamma_N-\lambda)^2 \tag{12}$$

3.4. *Conclusions and discussion*

From eqn. (7) we may conclude that the expected throughput time of a job depends on the number of runbatches per job. From eqn. (11) a conclusion could be that the throughput time averaged over all jobs is independent of both the distribution of the number of runbatches per job and the number of available pallets.

However, this second conclusion is in contradiction to our intuition, which is that the more constant the distribution of $\mathbf{q}$ is, the higher the average throughput time will be. This is due to the fact that in the case of a constant $\mathbf{q}$, the last runbatch of a job not only has to wait until all runbatches of previous jobs have been finished (as in 'normal' FIFO-queues), but also may have to wait for some of the runbatches of jobs that have arrived later (see Fig. 7). In the case of a highly variable distribution of $\mathbf{q}$, jobs with a

ORDER CHARACTERISTICS

Order:	A	B
Arrival time:	0	2
Workpiece:	a	b
Operation time / workpiece:	1	1
Lot size:	4	4

WORK ORDER SEQUENCE

Time:	0–1	1–2	2–3	3–4	4–5	5–6	6–7	7–8
Sequence 1 (sequential):	a	a	a	a	b	b	b	b
Sequence 2 (simultaneous):	a	a	b	a	b	a	b	b

AVERAGE THROUGHPUT TIME:

Sequence 1: 5

Sequence 2: 6

Figure 7. Example of throughput time consequences of new sequence for the case of constant number of runbatches.

ORDER CHARACTERISTICS

Order:	A	B
Arrival time:	0	2
Workpiece:	a	b
Operation time / workpiece:	1	1
Lot size:	6	2

WORK ORDER SEQUENCE

Time:	0–1	1–2	2–3	3–4	4–5	5–6	6–7	7–8
Sequence 1 (sequential):	a	a	a	a	a	a	b	b
Sequence 2 (simultaneous):	a	a	b	a	b	a	a	a

AVERAGE THROUGHPUT TIME:

Sequence 1: 6

Sequence 2: 5.5

Figure 8. Example of throughput time consequences of new sequence for the case of a variable number of runbatches.

small number of runbatches may compensate for jobs with a large number of runbatches (see Fig. 8). In fact, in the case of a geometric distribution of $\mathbf{q}$ and a negative exponential distribution of $\mathbf{o}$ it can be proven that the average throughput time where $N=\infty$ is equal to the case where $N=1$ (see Appendix and Kleinrock 1976).

In both cases (i.e. $\mathbf{q}$ being constant or highly variable) we may expect the standard deviation of the throughput time to grow with a larger number of pallets. If $\mathbf{q}$ is constant, this can easily be seen. The deviation in waiting time of the last runbatch of a job not only depends on the deviation in amount of work caused by the other runbatches of the same job and the remaining runbatches of previous jobs (as in the one pallet situation), but also on deviation in the amount of work caused by a number of jobs that have arrived after the job (which, among others, depends on the number of pallets). Since these two amounts are independent stochastic quantities, we may add their variances. The larger the number of pallets, the higher the variance of the second quantity will be and thus the higher the variance in waiting time will be. In the case of a highly variable $\mathbf{q}$, we can see from eqn. (7) that jobs with a small number of runbatches will have a shorter throughput time than jobs with a large number of runbatches. In fact, this last category of jobs may have to wait for a very long time before the last runbatch is finished.

3.5. *The use of simulation*

In order to check the conclusions and assumptions on average and standard deviation of throughput times, we will use simulation. The model described above is translated into a simulation program that runs on a IBM PS/2 Model 70 using Simulation (based on Simula). The next section will describe this simulation experiment and discuss the results.

4. The simulation experiment

The effects of finishing work orders in a quasi-simultaneous way will be demonstrated by some simulation experiments. The model of Fig. 6 will be used as a basis for the simulation experiments. Jobs arrive according to a Poisson arrival process with an average rate of arrival of λ. Upon arrival it is determined how many runbatches the job will contain and what the operation time per runbatch will be. The simulated manufacturing system contains only one machine. We finally assume the operation time per runbatch to be negative exponentially distributed with an average of 0·1.

In the first experiment we will illustrate the influence of q_i on the expected throughput time of a job. The number of runbatches per job in this experiment will follow the distribution of Fig. 9.

We will vary the number of pallets (1, 2, 4 or 10) and the utilization rate ($\pm 80\%$ or 90%; the exact utilization rate can only be measured after the simulation due to randomness), thus creating eight simulation runs. The simulation of a system with only one pallet is used as a reference, since this resembles the situation where all runbatches of a work order have to be finished before runbatches of a new work order can be loaded in the system.

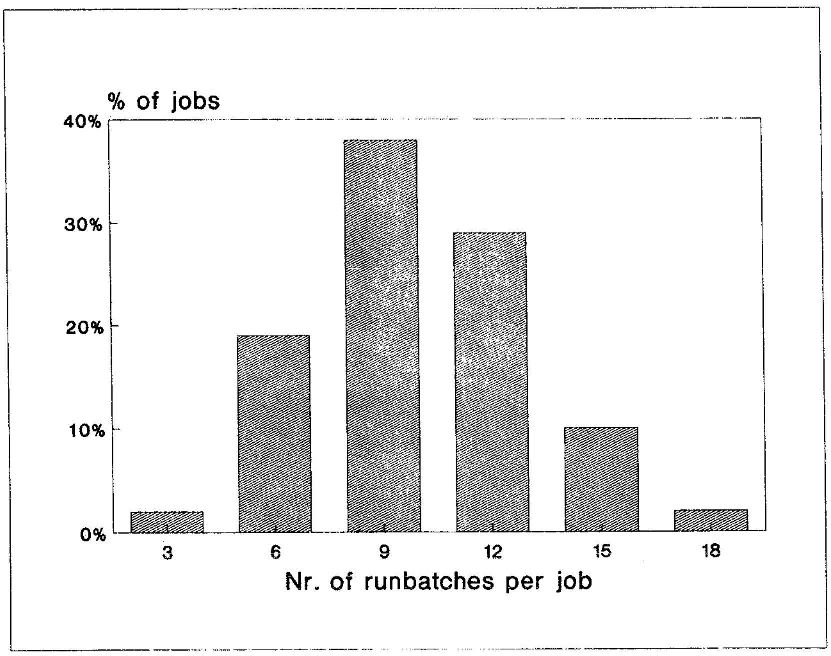


Figure 9. Distribution of number of runbatches per job in first experiment.

Table 1 presents the results of the simulation experiment. In Fig. 10 the results for the average throughput time per class of jobs are drawn. We may conclude from the table and the figure that the expected throughput time ($E(\mathbf{t}_i)$) indeed depends on the number of runbatches per job. The higher the utilization rate and the number of pallets are, the more dominant the runbatch related part of the average throughput time ($E(\mathbf{t}_i) - E(\mathbf{t}_{i,2})$) will be. The expected external waiting time ($E(\mathbf{t}_2)$) of course is independent of the number of runbatches per job.

The second experiment involves the influence of the quasi-simultaneous way of finishing work orders on the total average throughput time ($E(\mathbf{t})$) and the standard deviation in throughput time ($\sigma(\mathbf{t})$). As we have seen in the previous section, we expect this influence to depend on the distribution of the number of runbatches per job. Therefore, we have repeated the experiment described above using different distributions for the number of runbatches per job (with the same expected number of runbatches). Table 2 presents the results of simulation runs using a constant number of runbatches per job, whereas Table 3 gives the results of simulation runs using a geometric distribution. The results are shown more clearly in Figs. 11 and 12, depicting respectively the results for the average throughput time and the results for the standard

Number of pallets	$E(x)$		Lot size 3	6	9	12	15	18	All
Utilization rate 78·8%									
		Entries:	104	893	1807	1396	491	102	4793
1	0·79	$E(\mathbf{t}_i)$	3·46	4·37	4·69	5·13	5·07	5·91	4·80
		$E(\mathbf{t}_{i,2})$	3·17	3·78	3·77	3·97	3·68	3·98	3·81
		$\sigma(\mathbf{t}_i)$	4·68	4·98	4·81	5·09	5·01	5·82	4·98
2	1·44	$E(\mathbf{t}_i)$	3·25	4·36	4·91	5·58	5·74	6·89	5·09
		$E(\mathbf{t}_{i,2})$	2·70	3·27	3·26	3·45	3·20	3·51	3·30
		$\sigma(\mathbf{t}_i)$	4·54	4·94	4·87	5·23	5·24	6·22	5·05
4	2·47	$E(\mathbf{t}_i)$	2·82	4·33	5·26	6·31	6·83	8·43	5·57
		$E(\mathbf{t}_{i,2})$	1·96	2·44	2·44	2·63	2·47	2·96	2·50
		$\sigma(\mathbf{t}_i)$	4·31	4·93	5·05	5·62	5·85	7·18	5·32
10	4·24	$E(\mathbf{t}_i)$	2·23	4·28	5·85	7·60	8·64	10·99	6·38
		$E(\mathbf{t}_{i,2})$	0·85	1·08	1·03	1·19	1·06	1·40	1·09
		$\sigma(\mathbf{t}_i)$	3·49	4·79	5·63	6·94	7·77	9·89	6·23
Utilization rate 89·7%									
		Entries:	116	1006	2067	1587	548	117	5441
1	0·90	$E(\mathbf{t}_i)$	7·20	8·32	8·61	9·01	8·99	10·26	8·72
		$E(\mathbf{t}_{i,2})$	6·90	7·74	7·70	7·85	7·54	8·32	7·73
		$\sigma(\mathbf{t}_i)$	7·46	7·80	7·54	7·83	7·70	8·45	7·71
2	1·72	$E(\mathbf{t}_i)$	6·98	8·30	8·88	9·52	9·78	11·39	9·06
		$E(\mathbf{t}_{i,2})$	6·42	7·15	7·16	7·28	7·00	7·82	7·18
		$\sigma(\mathbf{t}_i)$	7·32	7·77	7·57	7·93	7·84	8·73	7·76
4	3·19	$E(\mathbf{t}_i)$	6·36	8·30	9·35	10·47	11·20	13·35	9·69
		$E(\mathbf{t}_{i,2})$	5·35	6·15	6·16	6·28	6·09	6·96	6·19
		$\sigma(\mathbf{t}_i)$	7·18	7·79	7·71	8·26	8·24	9·34	7·97
10	6·58	$E(\mathbf{t}_i)$	5·39	8·20	10·37	12·58	14·44	18·01	11·08
		$E(\mathbf{t}_{i,2})$	3·35	3·73	3·80	3·95	3·81	4·53	3·84
		$\sigma(\mathbf{t}_i)$	6·53	7·62	8·29	9·59	10·19	12·02	8·84

Table 1. Results from first simulation experiment.

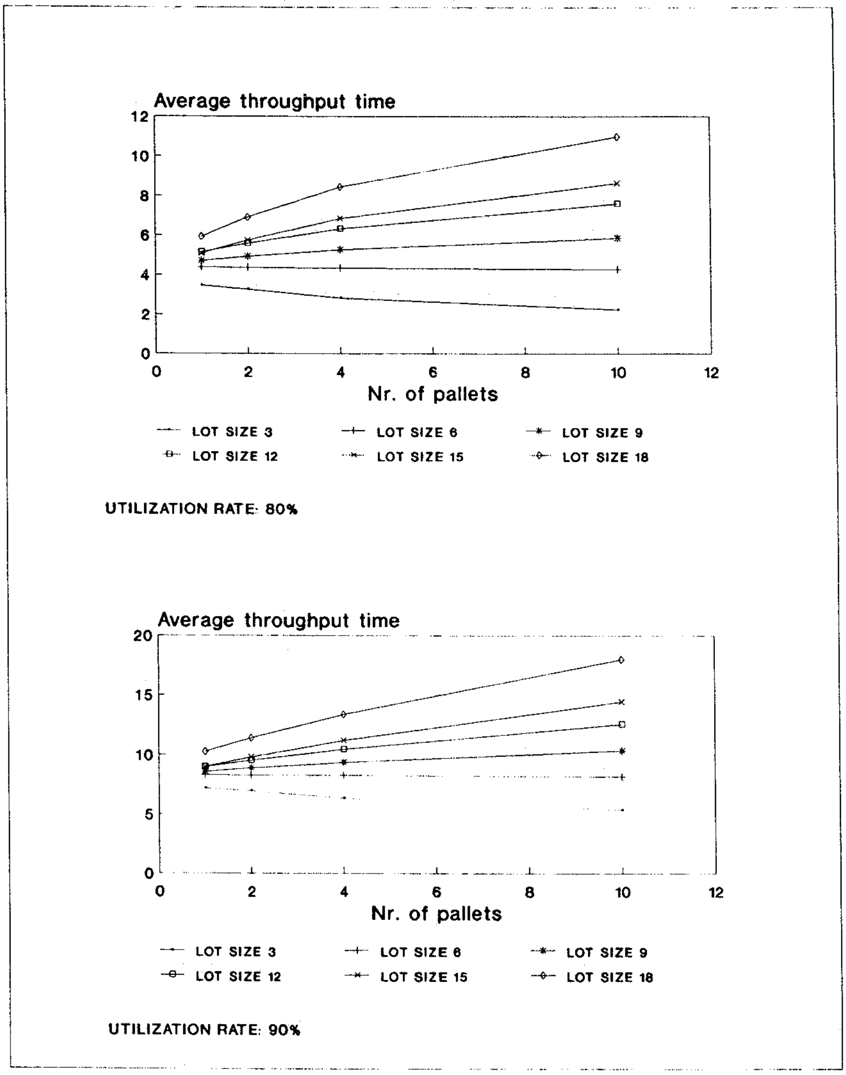


Figure 10. Average throughput time as function of number of pallets in first simulation experiment (utilization rate resp. 80% and 90%).

Utilization rate	Number of pallets	$E(x)$	$E(\mathbf{t}_i)$	$E(\mathbf{t}_{i,2})$	$\sigma(\mathbf{t}_i)$	Entries
79·2%	1	0·79	4·42	3·42	3·95	1901
	2	1·45	4·76	2·93	4·06	1901
	4	2·51	5·30	2·12	4·34	1899
	10	4·27	6·19	0·82	5·28	1897
88·6%	1	0·89	7·13	6·13	5·96	2122
	2	1·69	7·53	5·62	6·02	2122
	4	3·11	8·21	4·70	6·24	2121
	10	6·04	9·67	2·83	7·29	2120

Lot size = 10.

Table 2. Results in the case of constant number of runbatches per job.

Utilization rate	Number of pallets	$E(x)$	$E(\mathbf{t}_i)$	$E(\mathbf{t}_{i,2})$	$\sigma(\mathbf{t}_i)$	Entries
79·0%	1	0·79	7·23	6·24	8·06	1905
	2	1·45	7·22	5·39	8·06	1905
	4	2·49	7·15	4·02	8·37	1904
	10	4·25	6·85	1·51	9·35	1905
88·1%	1	0·88	11·06	10·06	10·32	2117
	2	1·69	11·05	9·14	10·34	2117
	4	3·09	11·05	7·56	10·64	2118
	10	6·07	10·89	4·03	6·86	2119

Average lot size = 10.

Table 3. Results in the case of geometric number of runbatches per job.

deviation in throughput time. As can be seen from the figures, the quasi-simultaneous finishing of work orders may have consequences for the average throughput time (the more constant the distribution of runbatches per job is, the more the average throughput time may grow) and will have (undesirable) consequences for the standard deviation in throughput time. The higher the utilization rate is and the larger the number of pallets are, the stronger these effects will be.

5. Conclusions

The results from the simulation study in this paper show that the quasi-simultaneous finishing of jobs has some important effects on the throughput time of the jobs. We have seen that the expected throughput time of a job heavily depends on its number of runbatches, especially in case of a high utilization rate and a large number of pallets. In order to translate these results into conclusions for real life situations, we have to relax some of the assumptions made in Section 3. In the case of only one pallet

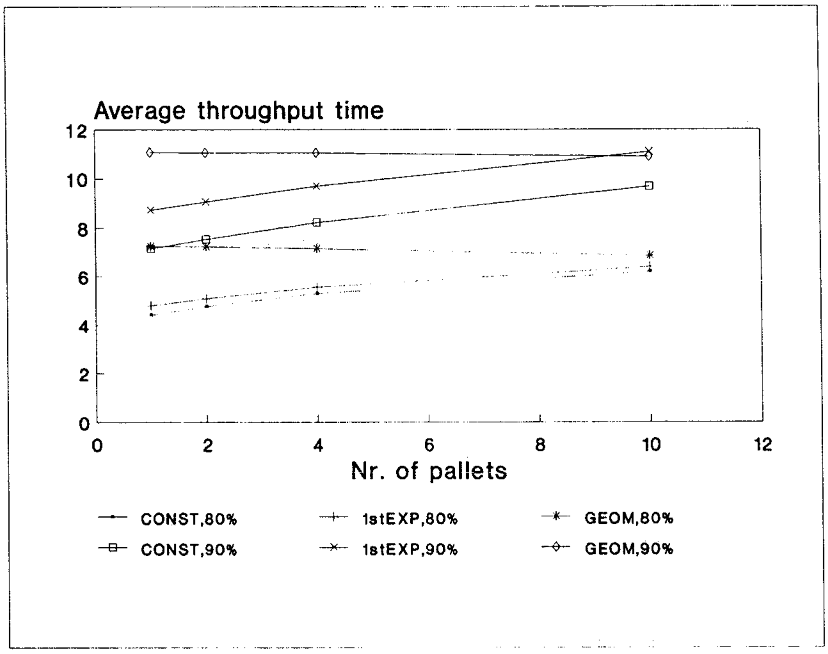


Figure 11. Average throughput time as function of number of pallets.

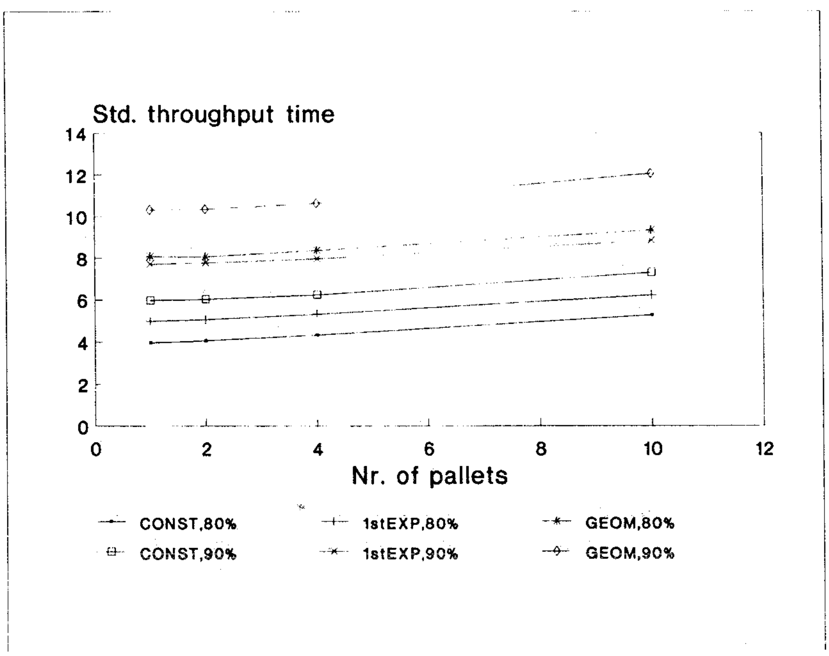


Figure 12. Standard deviation in throughput time as function of number of pallets.

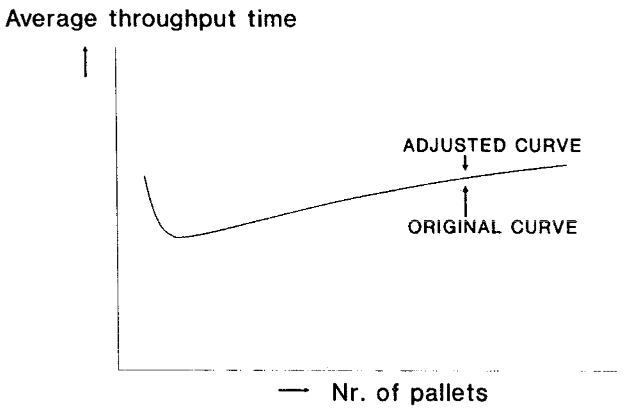


Figure 13. Shape of average throughput time curve in real life.

available on the manufacturing system, loading and unloading times cannot be neglected. As we have seen in Section 2, the possibility of loading and unloading pallets in front of the machine instead of actually on the machine itself reduces the average throughput time considerably. Therefore, we may not conclude from Fig. 11 that with a small variation in number of runbatches per job the optimal number of pallets on the system equals one. For real life situations the curves from Fig. 11 should be adapted according to Fig. 13. However, we may conclude that the optimal number of pallets (both from the point of view of average throughput time and of the variance in throughput time) in most cases will be small, unless some investments are made in fixturing tools in order to be able to put more than one runbatch per job at the same time on the manufacturing system.

On the other hand, a large number of pallets is desirable in the case of unmanned production. The larger this number is, the more capacity can be gained from unmanned night shifts. This leads to the conclusion that during manned production not all pallets should be used, whereas during the night as many pallets as possible should be used. These extra pallets should be loaded (and unloaded) during the manned shifts. Further research is required to validate this conclusion and translate the conclusions into practical decision rules. A further generalization of the conclusions is also required.

Acknowledgments

The author wishes to thank Professor J. W. M. Bertrand for his inspirational and supporting ideas and for the time spent on discussing the contents of this paper.

Appendix

In this Appendix we will demonstrate the difference in average throughput time between a single machine manufacturing system containing only one pallet and a system containing an infinite number of pallets, assuming the number of runbatches per job to be geometrically distributed, the jobs to arrive according to a Poisson arrival process with an average arrival rate of λ and the operation time per runbatch to be negative exponential distributed.

Case of only one pallet

In the case of a system containing only one pallet (see Fig. 14), we can calculate the average throughput time using the Pollaczek–Khinchin mean value formula (Kleinrock 1976). Let **z** be the operation time required for a whole job, **q** be the number of runbatches per job and **o** be the operation time per runbatch. Since **q** and **o** are independent we get:

$$E(\mathbf{z}) = E(\mathbf{q})E(\mathbf{o}), \tag{A 1}$$

$$E(\mathbf{z}^2) = 2E^2(\mathbf{q})E^2(\mathbf{o}) \tag{A 2}$$

Using the Pollaczek–Khinchin formula, (A 1) and (A 2) we get for the expected number of jobs present:

$$E(\mathbf{n}) = \{\lambda E(\mathbf{q})E(\mathbf{o})\}/\{1 - \lambda E(\mathbf{q})E(\mathbf{o})\} \tag{A 3}$$

Using Little's result (Kleinrock 1976) we may obtain the expected average throughput time:

$$E(\mathbf{t}) = E(\mathbf{n})/\lambda. \tag{A 4}$$

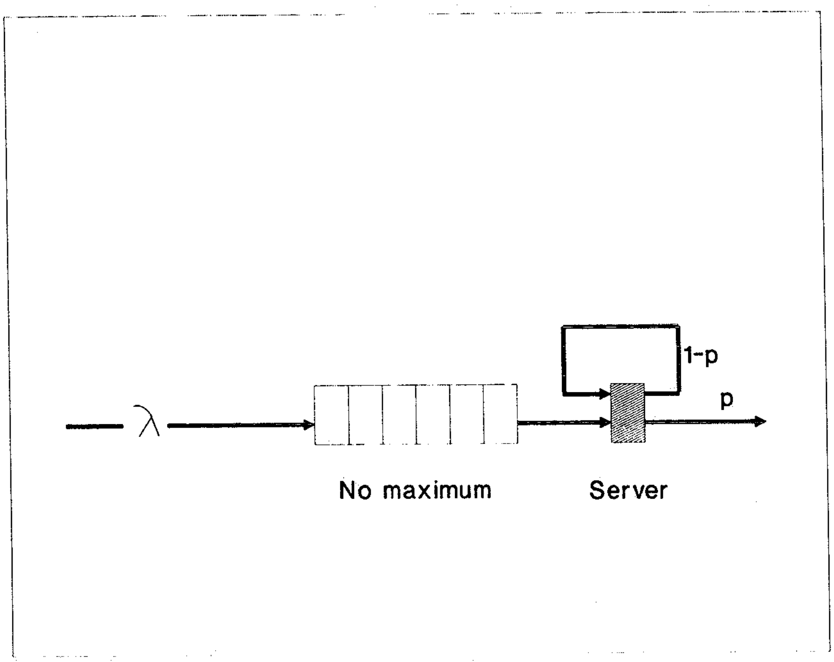


Figure 14. Model of case with one pallet.

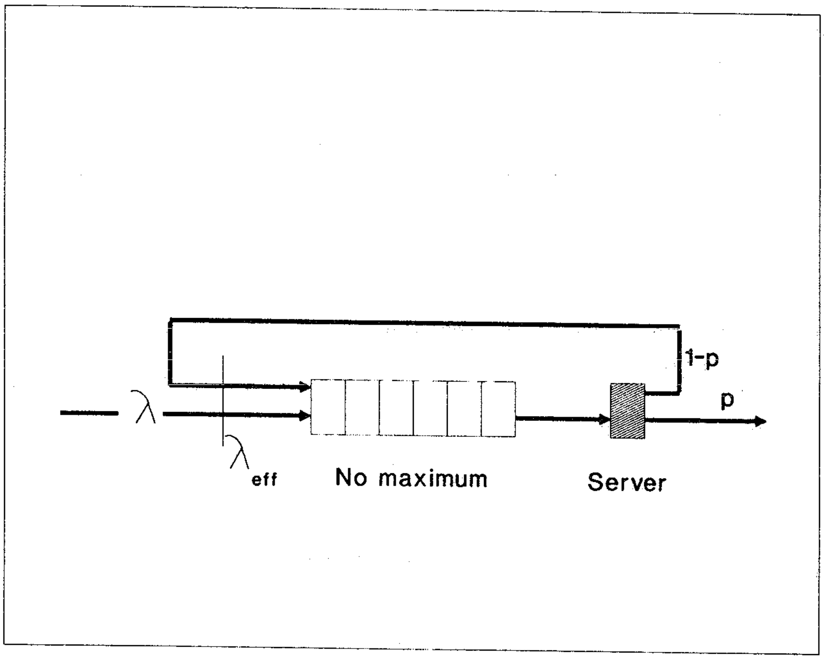


Figure 15. Model of case with infinite number of pallets.

Case of infinite number of pallets

In the case of an infinite number of pallets we can calculate $E(\mathbf{t})$ by realizing that the effective input in the system (with an average input rate of λ_{eff}, see Fig. 15) again is a Poisson process. λ_{eff} can be calculated by using the principle of 'what comes in must come out' (assuming a steady-state solution):

$$\lambda_{eff} = \lambda + (1-p)\lambda_{eff}, \quad \text{thus} \quad \lambda_{eff} = \lambda/p \tag{A 5}$$

where $p = E^{-1}(\mathbf{q})$.

We may obtain $E(\mathbf{n})$ using the results for an $M/M/1$ queue:

$$E(\mathbf{n}) = \lambda_{eff}E(\mathbf{o})/(1-\lambda_{eff}E(\mathbf{o}))$$
$$= \lambda E(\mathbf{q})E(\mathbf{o})/\{1-\lambda E(\mathbf{q})E(\mathbf{o})\} \tag{A 6}$$

As one can see $E(\mathbf{n})$ calculated by eqn. (A 3) is equal to $E(\mathbf{n})$ calculated by eqn. (A 6), thus resulting in an equal throughput time (see eqn. (A 4)).

References

AVI-ITHZAK, B., and HEYMAN, D. P., 1973, Approximate queuing models for multiprogramming computer systems. *Operations Research*, **21**, 1212–1230.

BERTRAND, J. W. M., 1985 a, Hierarchical approach to structuring the production control. *Modelling Production Management Systems* (Amsterdam: Elsevier Science Publishers/IFIP).

BERTRAND, J. W. M., 1985 b, Multiproduct optimal batch sizes with in-process inventories and multi work centers. *IIE Transactions*, **17**, 157–163.

BESSANT, J., and HAYWOOD, W., 1986, Experiences with FMS in the UK. *Managing Advanced Manufacturing Technology* (Kempston, Bedfordshire: IFS/Berlin: Springer Verlag).

BUZACOTT, J. A., 1984, Modelling flexible manufacturing systems. *Operational Research '84* (Amsterdam: North Holland/IFORS).

DIRNE, C. W. G. M., 1987, The impact of FAMS on overall production control structures. *Computers in Industry*, **9**, 337–351.

ELMARAGHY, H. A., 1985, Automated tool management in flexible manufacturing. *Journal of Manufacturing Systems*, **4**, 1–13.

KLEINROCK, L., 1976, *Queueing Systems, Vol. II: Computer Applications* (New York: Wiley).

SPUR, G., and MERTINS, K., 1982, Flexible manufacturing systems in Germany, conditions and development trends. *FMS: Proceedings of the First Conference* (Kempston, Bedfordshire: IFS).

STECKE, K. E., 1983, Formulation and solution of nonlinear integer production planning problems for flexible manufacturing systems. *Management Science*, **29**, 273–288.

STECKE, K. E., and SOLBERG, J. J., 1985, The optimality of unbalancing workloads and machine group sizes in closed queueing networks of multiserver queues. *Operations Research*, **33**, 882–910.