Throughput of Multiple Part-Type Systems:
A Subjective Linear Measure

Avi Herbon

Dept. of Interdisciplinary Studies - Logistics, Bar Ilan University, Israel

Eugene Khmelnitsky and Irad Ben-Gal

Dept. of Industrial Engineering, Tel-Aviv University, Ramat-Aviv 69978

Abstract

The term *throughput*, which is commonly used as a performance measure of various production systems, has not been uniquely defined for multiple part-type systems. In some cases, the analytical procedures that were developed to maximize throughput of multiple part-type systems are difficult to evaluate and justify. Moreover, in some cases an inaccurate definition of this term motivated incorrect operational concepts. This paper discusses some of the problems in the traditional definition of throughput and suggests a new, more suitable definition for a multiple-product manufacturing system.

Key words: FMS, Performance measure, Productivity, Workload

1. Introduction

Modern flexible manufacturing systems (FMSs) are capital intensive and require an efficient use of their resources. In measuring the effectiveness and the size of FMSs, the *throughput* is one of the most important and commonly used performance measures together with inventory level, due date, and customer satisfaction. Several potential advantages are known to be associated with an increased throughput level: higher profits can be obtained; delays in orders can be decreased; and the need to expand existing resources can be reduced. Moreover, a larger throughput provides additional production flexibility in responding to sudden changes in resources and demands.

For a single part-type system, throughput is defined as *the number of parts produced in a unit of time* (e.g., see Gudmundsson and Goldberg, 1999). In this commonly accepted definition, the basic properties of throughput are quantity and time. In such a system, the throughput measure is commonly considered within a framework that optimizes the system performance. Dessouky *et al*. (1995), for example, discussed a scheduling deterministic problem of flexible assembly lines aiming at maximizing the throughput and minimizing the WIP. Since the authors considered only one type of assembly line with a single part type and a single route, the maximization of the throughput, which was defined as the *number of parts produced in a given period*, resulted in an optimal scheduling policy. A procedure for estimating throughput in a more complex plant configuration with rework loops is developed by Li (2004).

Kao and Sanders (1995) aimed at maximizing throughput using quality inspection policies. They measured throughput by the *number of "good" parts in a time unit* and developed a decision-making model for selecting equipment, determining the necessity of an inspection, and evaluating the error recovery. Although their definition was proposed only to a single part-type problem, it emphasizes another accepted property of throughput, namely, that throughput is usually based only on conforming parts. Such a definition of the throughput associates it with another key performance measure of manufacturing systems: "satisfying the customer needs".

In contrast to the above examples, the term *throughput* has not been uniquely defined for multiple part-type systems. Lacking such a proper definition, one cannot accurately evaluate the performance of modern manufacturing systems, which usually produce multiple part-types. The problem is further complicated when various definitions are adopted to measure the efficiency of implemented control and scheduling policies. Many scheduling rules for these systems tend to overlook the difference between what is *supposed* to be measured – the throughput – and what is actually being measured. Machine utilization, total machine work-load, average yield and other actual measures are often used as related estimates for the throughput. As a result, one might achieve a "better" scheduler and a "better" layout, yet, ends up with a smaller throughput – as illustrated by examples in following sections.

Down through the years, several approaches have been suggested for measuring the throughput of multiple part-type systems. Hodgson *et al.* (1987) suggested generalized scheduling control rules for maximizing throughput by using a Markov decision process. However, while seeking to maximize the "*expected number of loads* *delivered in the system per unit time"* over an infinite horizon, the authors ignored the difference among various part-types.

Other approaches for evaluating throughput in multiple part-types systems were based on the *quantities vector* that indicates the number of units of each part-type that are produced during a given time period. The quantities vector has been used to analyze the throughput in complex, open and closed queuing networks (Jackson 1963, Suri and Hildenbrant 1984, Suri and Sanders 1993). Comparing different systems by their quantities vector throughput is tricky when no system dominates the others, i.e., when no system produces more units of all part-types. This problem has led many researchers to transform the quantities vector to a scalar measure, as described next.

A popular practice for obtaining a scalar measure for the throughput is to assign weights to the different part-types with respect to their relative share (percentage) in the total workload. Tiwari *et al*. (1997) discussed the loading problem in FMS. Ignoring setups and handling costs, the authors considered a deterministic production environment with a single routing alternative for each job. Throughput was defined as the *"sum of the processing times of all produced parts"*. Although the overall workload of such system can be estimated quite accurately, the definition raises counter-intuitive phenomena. For example, in a case of a technological improvement, the number of produced parts is likely to increase in a given time interval, whereas the total workload and, therefore, the throughput according to the above definition, remain unchanged.

Measuring the workload in a multiple part-type FMS, where each part may be processed in several routes, is a particularly complex task. Workload estimation depends not only on system specifications, such as the routings, but also on external factors, such as the demand distribution. Chan (1999) aimed at maximizing the sum of all the production rates of all products over all the process plans. He normalized the production rates of different product types by assigning relative weights to each of which. The product type weight was calculated as the ratio between i) the sum of all production rates of that product type over all possible routes; and ii) the overall production rate of the FMS. Note that this procedure requires calculating all routing weights, albeit there is no guarantee that the objective will be obtained since it is unclear which routes will eventually be used.

In highly stochastic environments, the *workload estimation* problem is further complicated by additional sources of uncertainty, such as,

- different part-types that share common resources
- the number of routes which is usually too large to be considered on-line
- unexpected events, such as machine failures, that prevent an accurate off-line predictions of the workload
- processing times that can only be approximated off-line
- unknown a priori set-up times that significantly affect the workload
- scheduling policies that usually assign higher priorities to efficient machines and by that cause an unbalanced machines' workload.
- rerouting that often occurs online.

Arzi and Roll (1993), Arzi (1995) and Herbon (1998) addressed the problem of controlling an FMS that operates under a highly stochastic environment. The control objective of these research works was to maximize the throughput while minimizing orders tardiness. Once again, throughput was measured by the *weighted sum of all possible processing routes of all parts*. Although the proposed procedure assigned higher priority to machines that participate in various processing routes, their definition of the throughput carry the same weaknesses listed above.

Mukhopadhyay and Sahu (1996) discussed a deterministic tool allocation problem in FMS. They claimed that the maximization of the throughput is equivalent to the minimization of the makespan. Even though such equivalence seems logical, its utilization is problematic from practical reasons. First, an exact comparison of competing schedulers is almost impossible since their performance highly depends on the set of processed parts that changes over time. Second, when increasing the quantities to be processed, the throughput (makespan in this case) is not guaranteed to increase accordingly. Finally, when using the makespan-based definition for the throughput, one cannot distinguish between throughputs associated with different product mixes that are produced on a given time period.

Following the above discussion, this paper suggests a new definition for the throughput of multiple part-type systems. The new definition overcomes most of the above-mentioned difficulties. The rest of the paper is organized as follows. Various arguments for seeking an acceptable throughput definition are given in Section 2 by means of small illustrative examples. A conceptual discussion on the required properties of the throughput definition is presented in Section 3. Section 4 presents an example of the proposed throughput definition and shows that it confirms with common industrial principles and beliefs. Section 5 concludes the paper.

2. Consequences of non-accurate throughput definition

This section argues that traditional throughput definitions might result in counter-intuitive phenomena. Specifically, it is claimed that inaccurate definitions of the throughput for multiple part-type systems might encourage the wrong evaluation of production control performance, and even lead to inaccurate evaluation of technological improvements that affect the system efficiency. We emphasize some of these phenomena by means of illustrative examples that follow next.

Example 1*: *Discouraging the reduction in the production-time

Consider a manufacturing system that consists of two machines, M_1 and M_2, and produces respectively two part-types. In particular, producing part-type 1 requires t_1 minutes of machine M_1, while producing part-type 2 requires t_2 minutes of machine M_2. The system operates for T minutes.

Since there is a single route for each part-type and the production resources are independent, the system has the capacity to process $Q_1 = T/t_1$ parts of type 1 and $Q_2 = T/t_2$ parts of type 2. Measuring throughput by the "weighted workload" definition results in the following throughput: $Thr = Q_1 t_1 + Q_2 t_2 = 2T$.

Let us now suppose that the parts' processing times decrease as a result of a technological improvement (for example, replacing the old machines by new ones) that is represented by factor $0 < \alpha < 1$. Accordingly, the updated processing times after the improvement are $\alpha \cdot t_1$, and $\alpha \cdot t_2$. Operating the system for T minutes following the improvement yields the following quantities of parts: $Q_1^{after} = \frac{T}{\alpha \cdot t_1}$ of part-type 1, and $Q_2^{after} = \frac{T}{\alpha \cdot t_2}$ of part-type 2. Note, however, that despite the increase in the produced quantities, the measured throughput remains fixed and equals $Thr^{after} = 2T$. Thus, measuring the throughput by the workload leads toward a counter-intuitive situation where the same measured values are obtained before and after the implementation of the technological improvement. Since such a measure overlooks the fact that after the improvement more parts are being produced from both part-types, it does not encourage the desired increase in the process efficiency.

Example 2: Superiority of non-efficient production-control rules

Consider an ideal flexible manufacturing system, where each machine can process any part-type. The system operates for *T* minutes under two different scheduling rules, denoted by A and B. Each time that a machine becomes available, Scheduler A applies the LRA (Largest Relative Advantage) rule, as suggested in Roll *et al*. (1991), to increase the system efficiency by assigning the part with the relatively lowest processing time to the available machine (see Seciton 4 for further discussion). For illustration purpose, let us consider Scheduler B that applies the inefficient (hypothetic) rule – the SRA (Smallest Relative Advantage) – which iteratively assigns the part with the relatively highest processing time to the available machine. In such a case, the averaged throughputs, as measured by the weighted workload of both schedulers A and B, are $Thr_A = \sum_{i=1}^{I} Q_i^A \bar{t}_i^A$ and $Thr_B = \sum_{i=1}^{I} Q_i^B \bar{t}_i^B$, where $\bar{t}_i^A$ and $\bar{t}_i^B$ denote the average processing time of part-type *i* over all the machines that produced such a part in a given period, using, respectively, scheduler A or B, and Q_i^A, Q_i^B denote the produced quantities of part type *i* by the respective scheduler. Note from the definitions of the LRA and SRA rules that, on the average, $\bar{t}_i^A < \bar{t}_i^B$ whereas $Q_i^A > Q_i^B$, thus, one cannot guarantee that $Thr_A > Thr_B$. In other words, the workload definition of throughput may lead to a situation where the inefficient SRA rule results in a higher throughput measure than the efficient LRA rule.

Example 3: A non-consistent definition

Another objective definition for the throughput was introduced by Arzi (1995) that considered a multiple-cell FMS operating under stochastic conditions. The production system included several products, multiple routes for each part-type, machine failures, handling failure and random stream of orders. The workload of each part-type was calculated off-line by a heuristic procedure and the throughput was measured by the total system workload. In order to evaluate the effectiveness of proposed heuristics with respect to an 'objective' measure, the author calculated an upper bound for the throughput via simulations. This upper bound, denoted by *Z*, was obtained from the following linear model:

$$Max\ Z$$

s.t. 1) $\sum_{k=1}^{K} q_{i,k} = Q_i \cdot Z$, $\forall i$,

2) Feasibility constraint

3) Capacity constraint,

where $q_{i,k}$ is a decision variable that indicates the recommended number of parts of type *i* to be processed in cell *k* aiming at maximizing the throughput; Q_i is the number of parts of type *i* that were actually produced in the simulated system (known when the simulation is completed) Note that maximum throughput strongly depends on the values of Q_i obtained from a particular simulation run. This means that different production control schemes that result in different inputs Q_i potentially lead to different 'objective' throughputs for the same system. Moreover, the product mix associated with the maximum throughput in this model, must be no less than Q_i for each *i*, i.e., $\sum_{k=1}^{K} q_{i,k} \geq Q_i$ for all the part-types (*Z*=1 is obtained from the actual simulation run). This requirement significantly constrains the selection of a product mix, and therefore, the maximum possible throughout is unlikely to be obtained by solving this problem. For example, if a non-efficient scheduler oriented towards a specific part-type *j* is implemented at the simulation stage, then finding the optimal throughput by this model is even more restrictive. In such a case, if product *j* is highly resource-consuming, the optimization is likely to result in a low throughput.

Example 4: Profit-based and objective-based definitions

A market-based definition for the throughput was proposed by Goldratt (1986), who suggested measuring the throughput by the "*profit rate of processed parts*". Even though such a definition targets the bottom-line of production-control, it encourages the production of the set of the most profitable parts at a given period of time. This set, however, is not necessarily identical to the set of largest produced quantities in a given time period. Throughput, according to this definition, may increase or decrease solely due to price changes in the markets, regardless of operational actions or the system capabilities. Such a profit-oriented definition is quite distant from the conventional concept of throughput that is oriented in the literature to part quantities and production speed.

3. Suggested Properties of the Throughput Definition

Following the above examples, this section outlines the required properties for a definition of the throughput of multiple part type systems. These required properties aim at maintaining the common concepts of the term throughput, yet to overcome the above mentioned inconsistencies.

1. Throughput should be a function of part quantities, Q_i, actually produced on a given planning horizon *T*, i.e., $f(Q_1,...,Q_n,T)$, where *n* is the number of part-types. This definition is in agreement with all previous throughput definitions found in the literature.
2. Throughput accumulated at a given time interval [0,*T*] should reflect the summation of throughputs that were obtained at its sub-intervals. That is, $f(Q_1,Q_2,..,T) = f(Q_1^a,Q_2^a,..,\rho T) + f(Q_1^b,Q_2^b,..,(1-\rho)T)$, where Q_i^a are quantities produced on interval $a=[0,\rho T]$, Q_i^b are quantities produced on interval $b=[\rho T,T]$, $\forall 0 \le \rho \le 1$, and $Q_i^a + Q_i^b = Q_i$ for all *i*. As a straightforward outcome of such requirements, the throughput should be a linear function of part quantities, Q_i, $f(Q_1,...,Q_n,T) = \sum_i \omega_i Q_i$.

3. The coefficients of the above linear function, ω_i, must be non-negative, otherwise, the production of a larger set of parts might reflect a decreased throughput measure, a counter-intuitive phenomenon.
4. The coefficients of the throughput function, ω_i, should be subjectively related to the production system and its operator's goals. This subjectivity assumes that in order to achieve a higher throughput in a production system, one has to utilize its resources (e.g., equipment) more efficiently. Otherwise, having objective coefficients ω_i might contradict this assumption, as illustrated by the following example. Consider two sets of parts, P_a and P_b. Associate P_a with a throughput measure, which is considered objectively higher than the throughput measure associated with P_b. Accordingly, in a specific production system, which is relatively more adapted in producing parts from P_b, the manufacturer has to utilize the equipment inefficiently in order to achieve a higher throughput. This contradicts the above assumption and results in a counter-intuitive phenomenon.
5. The coefficients of the linear function, ω_i, should be independent of the implemented scheduling procedure that seeks a higher throughput. Otherwise, a contradiction, such as the one presented in example 2, can occur, where inefficient procedures result in higher throughput measures.
6. Since the conventional definition of throughput is often associated with the speed of the production process, the coefficients of the linear function, ω_i, should be related to the processing times and the production routing flexibility. These coefficients should be independent of part attributes that are not directly related to processing speed and produced quantities. While production attributes such as profit, cost, due-date and quality reflect vital performance measures, they are not necessarily related to the common concept of throughput as reflected in the literatures.
7. Several methods were suggested for an appropriate selection of coefficients of a given objective function (Korhonen and Wallenius 1988, Thurston 1990, Brans and Mareschal 1994). The methods are based on subjective acquaintance of managers with product mixes and machine layouts. Although these procedures do not guarantee to obtain the exact values of the coefficient values, they provide reasonable results and are quite convenient from practical reasons. Note that on one hand, the presented properties do not leave a large gap for selecting the values

of ω_i , while on the other hand, small fluctuations in the values of ω_i often do not alter the operational decisions that are mainly based on the ratios among these coefficients, such as the selection between competing operational policies.

Note that once one accepts the first two properties, the throughput measure has to be defined by a linear function. Then, the problem of a correct definition of the throughput measure of a multiple part type system is reduced to the proper selection of the linear coefficients, as discussed by the rest of the listed properties. Moreover, having built the throughput measure, one can use it as an optimization criterion for a proposed scheduling system, i.e,

$$Max\ Thr = \sum_i \omega_i Q_i$$

s.t. 1) Feasibility constraint

2) Capacity constraint.

4. Discussion

This paper focuses on the definition of the throughput measure, which is one of the basic and frequently used measures in industrial engineering. The suggested definition of throughput as a linear function with subjectively chosen coefficients does not contradict the widely accepted concepts that are associated with the throughput, such as ‘processing speed’, ‘larger quantities’, and ‘increasing quantity of good parts’. We believe that the suggested definition of throughput directs a production system towards desirable performance, while maintaining the main traditional elements of the throughput measure.

The following short example illustrates how a user can practically calculate the coefficients of the linear throughput measure.

Consider a production system that produces three part-types. The system can follow three different processes. In each process the system produces two out of the three part-types. In the first process, the system produces part-types 1 and 2. The ratio between the processing times of the two part types in the system indicate that part 2 consumes twice as many resources as part 1 (e.g., processing time of part type 2 is twice as large as that of part type 1). In the second process, the system produces part-types 1 and 3, where part 3 process time is seven times larger than that of part 1. In the third process, the system produces part-types 2 and 3 and similar considerations led the user to conclude that part 3 process time is twice as larger as the process time of part-type 2. As a result, the user generates the following matrix of relative resource consumption of the parts that is related to all possible processes

$$r_{ij} = \begin{bmatrix} 1 & 2 & 7 \\ 1/2 & 1 & 2 \\ 1/7 & 1/2 & 1 \end{bmatrix}. \qquad (1)$$

Note that the matrix, which is based on pair-wise comparisons, is not consistent, since resource consumption of different part-types cannot be described by a single linear function. As a measure of consistency, Saati (1980) suggested the *consistency ratio (CR)*, which is based on the weighted distance between the principle eigenvector of the matrix and the matrix rank. In this case, $CR = \frac{3.035 - 3}{3 * 0.58} = 0.02$ which is considered as an acceptable consistency ratio. Following the *Analytic Hierarchy Process*, suggested by Saati (1980), the consistent coefficients ω_i form the vector of priorities, which is the normalized principle eigenvector of the matrix. In this case, these coefficients are $\omega_1 = 0.63, \omega_2 = 0.26,$ and $\omega_3 = 0.11$. As a result, the throughput of a set Q_i, i=1,2,3, produced in a given time *T*, will be measured by the following function,

$$Thr = \omega_1 Q_1 + \omega_2 Q_2 + \omega_3 Q_3 = 0.63 \cdot Q_1 + 0.26 \cdot Q_2 + 0.11 \cdot Q_3 .$$

Now, for comparison purpose, let us discuss the meaning and the value of the suggested throughput measure with respect to the shortcomings/inconsistencies of previous definitions, as illustrated by the examples in Section 2.

Technological improvement (see example 1 in Section 2). Suppose that the production technology has improved and the rate of the three processes presented has increased by a factor of $\frac{1}{\alpha}$ $0 < \alpha < 1$. As a result, the new production quantities, $Q_i^{after} = Q_i / \alpha \quad i = 1,2,3$ have increased by the same ratio. Since the relative resource consumption of the parts (as measured by the relative ratios of processing times), r_{ij}, remains unchanged, the coefficients ω_i remain fixed, and the new throughput equals

$$Thr^{after} = 0.63 \cdot \frac{Q_1}{\alpha} + 0.26 \cdot \frac{Q_2}{\alpha} + 0.11 \cdot \frac{Q_3}{\alpha} = \frac{Thr^{before}}{\alpha}.$$

Thus, as expected and in contrast to the workload-based definitions of the throughput, the suggested definition does reflect the effects of technological improvements.

Superiority of more efficient scheduling algorithms (see example 2 in Section 2). is clearly expressed in the suggested measure, since, again, the coefficient ω_i are not influenced by the change in scheduling algorithm, while the quantities produced are larger for better algorithms. For example, let us keep the relationships between the processing times in the example above, and consider the processing times (in minutes) of part *i* on machine *m* tp be represented by the following matrix:

$$t_{im} = \begin{vmatrix} 4 & 3 & \infty \\ 8 & \infty & 5 \\ \infty & 21 & 10 \end{vmatrix},$$

where, the *im* entry corresponds to the processing time of the product i (row) on the *m*-th machine (column). Then, applying the LRA and SRA rules results in the following schedules:

	Scheduler LRA	Scheduler SRA
Chosen part type by M1	1	2
Chosen part type by M2	1	3
Chosen part type by M3	3	2

In particular, machine 1 has to choose between part types 1 and 2. With respect to part type 1 it has an advantage of (-1) minutes per part over machine 2, while with respect to part type 2 it has an advantage of (-3) minutes per part over machine 3. Thus, the LRA scheduling rule prefers part type 1 to be processed on machine 1. Similarly, the other machines choose product types with respect to their largest relative advantage. Contrary to the LRA rule, the SRA rule chooses the part types with smallest relative advantage. When operating the system during a time period of 60 minutes, the suggested measure results in:

$$Thr_{LRA} = 0.63 \cdot \left(\frac{60}{4} + \frac{60}{3} \right) + 0.26 \cdot 0 + 0.11 \cdot \frac{60}{10} = 22.71.$$

$$Thr_{SRA} = 0.63 \cdot 0 + 0.26 \cdot \left(\frac{60}{8} + \frac{60}{5} \right) + 0.11 \cdot \frac{60}{21} = 5.38 \ .$$

Thus, the Thr_{LRA} > Thr_{SRA} as desired. Note however that the workload measure, as discussed in Example 2 in section 2, results in this example in:

$$Thr_{LRA} = \left(\frac{60}{4} + \frac{60}{3} \right) \cdot \left(\frac{3+4}{2} \right) + \frac{60}{10} \cdot 10 = 182.5 \ .$$

$$Thr_{SRA} = \left(\frac{60}{8} + \frac{60}{5} \right) \left(\frac{8+5}{2} \right) + \frac{60}{21} \cdot 21 = 186.75 \ .$$

This observation agrees with the conclusion made in Example 2 that the "non-efficient" hypothetic SRA rule may result in a higher throughput than the "efficient" LRA rule. At the same time, the suggested definition does not contradict the superiority of the LRA rule.

A non-consistent definition (see examples 3 and 4 in Section 2). The suggested measure also resolves the inconsistency described in Example 3, where the maximum throughput of a specific system directly depended on the scheduler that had been used in the past. In such a situation, there can be several different outcomes for the well-defined question: What is the maximum throughput of a specific system? In the suggested method, the choice of the parameters ω_i is independent of the scheduler, and the maximum throughput is obtained uniquely from maximization of the throughput measure. Similarly, the inconsistency described in Example 4 is resolved, since the parameters ω_i are independent of immediate (short-term) market fluctuations.

The above discussion highlights some additional points that should be considered. First, note that when the production system or the managerial environment changes (for example, due to non-symmetrical technological innovations, when the processing times of some products decrease, while the others remain unchanged), the user's perception of the produced parts changes as well, and that should lead to a re-evaluation of the coefficients ω_i. Second, note that the comparison between different production control and scheduling methods, with respect to the new definition of the throughput are subjective in its nature, and, therefore, can give preference to different schemes for different users.

5. Conclusions

This paper addresses the problem of a proper throughput definition for a multiple part-type system. While seeking to maintain the basic traditional concepts associated with the throughput measure, we suggest a new definition that does not lead to some of the inconsistencies that are associated with some of the conventional definitions of throughput, such as a workload measure.

Although any definition, including the one proposed here, is subjective by its nature, we tried to reach a less biased definition by first specifying a set of properties that should be satisfied. The specified set of properties in section 3 lead us to define the throughput measure as a linear function of the produced quantities at a given time period. The function subjective coefficients can be determined by several numerical methods, including the one exemplified in Section 4. It is not claimed that the presented definition of throughput is an ultimate one or a dominant one. Yet, it is believed that the suggested definition addresses certain phenomena that were somewhat overlooked by the traditional definitions.

References

Arzi, Y. (1995), "On-line scheduling in a multiple-cell flexible manufacturing system", *International Journal of Production Research*, Vol. 33 No. 7, pp. 3283-3300.

Arzi Y. and Roll, Y. (1993), "Real-time production control of an FMS in a produce-to-order environment", *International Journal of Production Research*, Vol. 31 No. 9, pp. 2195-2214.

Brans, P.J. and Mareschal, B. (1994), "The PROMCALC & GAIA decision support system for multiple-criteria decision aid", *Decision Support Systems*, Vol. 12, pp. 297-310.

Dessouky, M.M., Adiga, S. and Park, K. (1995), "Design and scheduling of flexible assembly lines for printed circuit boards". *International Journal of Production Research*, Vol. 33 No. 3, pp. 757-775.

Gudmundsson, D. and Goldberg K. (1999), "Tuning robotic part feeder parameters to maximize throughput", *Assembly Automation*, Vol 19 No. 3, pp. 216-221.

Herbon, A. (1998), "On-line production control of a flexible multiple-cell manufacturing system operating in a highly dynamic environment", *International Journal of Production Research*, Vol 36 No. 10, pp. 2771-2791.

Hodgson, T.J., King, R.E., Monteith, S.K. and Schultz S.R. (1987), "Developing control rules for an AGVS using Markov decision processes", *Material Flow*, Vol. 4 No. 1-2, pp. 85-96.

Jackson, J. R. (1963), "Jobshop-like queuing systems". *Management Science*, Vol. 10, pp. 131-142.

Kao, J.-F. and Sanders, J.L. (1995), "Analysis of operating policies for manufacturing cells", *International Journal of Production Research*, Vol. 33 No. 8, pp. 2223-2239 .

Korhonen, P. and Wallenius, J. (1988), "A Pareto race", *Naval Research Logistics*, Vol. 35, pp. 615-623.

Li, J. (2004), "Performance analysis of production systems with rework loops", *IIE Transactions*, Vol. 36, pp. 755-765.

Mukhopadhyay, S.K. and Sahu, S.K. (1996), "Priority-based tool allocation in a flexible manufacturing system", *International Journal of Production Research*, Vol. 34 No. 7, pp. 1995-2018.

Roll, Y., Karni, R. and Arzi, Y. (1991), "The Largest Relative Advantage (LRA) scheduling rule for flexible manufacturing", *Proceedings of the 11th International Conference on Production Research*, L. Ming (ed), Heifei, China, pp. 197-200.

Saati T. L. (1980), *The Analytic Hierarchy Process*, Mc-Graw Hill, New-York.

Suri, R. and Hildenbrant, R.R. (1984), "Modeling flexible manufacturing systems using mean value analysis", *Journal of Manufacturing Systems*, Vol. 3 No. 1, pp. 27-38.

Suri, R. and Sandres, J. (1993), "Performance evaluation of production networks", *Handbook in OR&MS*, Vol. 4, Graves S. C. *et al*., Eds., Elsevier Science Publishers.

Thurston, D.L. (1990), "Multiple-attribute utility analysis in design management", *IEEE Transactions on Engineering Management*, Vol. 37 No. 4, pp. 296-301.

Tiwari, M.K, Hazaria, B., Vidyarthi, N.K, Jaggi, P. and Mukhopadhyay, S.K. (1997), "A heuristic solution approach to the machine loading problem of an FMS and its Petri net model", *International Journal of Production Research*, Vol. 35 No. 8, pp. 2269-2284.