

To appear in the *International Journal of Production Research*
Vol. 00, No. 00, 00 Month 20XX, 1–32

Multi-job production systems: definition, problems, and product-mix performance portrait of serial lines

Pooya Alavian^a, Peter Denno^b, and Semyon M. Meerkov^{c*}

^a*Integrative Systems + Design, University of Michigan, Ann Arbor, Michigan, USA;* ^b*Systems Integration Division, Engineering Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland, USA;* ^c*Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan, USA.*

(Received 00 Month 20XX; accepted 00 Month 20XX)

This paper pursues two goals: (a) Define a class of widely used in practice flexible manufacturing systems, referred to as Multi-Job Production (MJP) and formulate industrially motivated problems related to their performance. (b) Provide initial results concerning some of these problems pertaining to analysis of the throughput and bottlenecks of MJP serial lines as functions of the product-mix.

In MJP systems, all job-types are processed by the same sequence of manufacturing operations, but with different processing time at some or all machines. To analyze MJP with unreliable machines, we introduce the work-based model of production systems, which is insensitive to whether single- or multi-job manufacturing takes place. Based on this model, we investigate the performance of MJP lines as a function of the product-mix. We show, in particular, that for the so-called conflicting jobs there exists a range of product-mixes, wherein the throughput of MJP is larger than that of any constituent job-type manufactured in a single-job regime. To characterize the global behavior of MJP lines, we introduce the Product-Mix Performance Portrait, which represents the system properties for all product-mixes and which can be used for operations management. Finally, we report the results of an application at an automotive assembly plant.

Keywords: Flexible manufacturing systems; Serial lines; Unreliable machines; Product-mix; Throughput and bottleneck analysis; Performance portrait.

1. Introduction

1.1 *Multi-job production systems: description and definition*

Multi-job production (MJP) is a class of flexible manufacturing systems, intended to produce different products (or *job-types*) within the same production system. These systems are widely used in product assembly, e.g., in automotive assembly plants, engine and battery plants, computer and appliance assembly, etc.

To illustrate MJP operation, consider an automotive assembly plant manufacturing two car models, A and B . In each area of the plant, i.e., body shop, paint shop, and final assembly, each job-type follows the same sequence of manufacturing operations. Let $\mathbf{r} = (r_A, r_B)$ be the *product-mix*, where r_A is the fraction of automobiles A to be manufactured and $r_B = 1 - r_A$ is that of B . The jobs are released one-by-one into the body shop in a sequence defined by the product-mix and the corresponding *build-schedule* and then proceed to the paint shop and final assembly. For instance, a segment of a release sequence may be $\cdots BAAABABAAB \cdots$, generated by the product-mix with $r_A = 2/3$ and $r_B = 1/3$. The jobs are transported from one operation to another (typically,

*Corresponding author. Email: smm@umich.edu

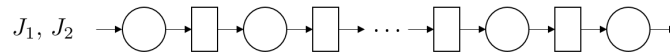


Figure 1. MJP serial line

by conveyors, which serve also as buffers) in the sequence of release. Each job-type is processed by the machines (operations or stations) with zero (or practically zero) setup time, but requires different *processing time* at some or all machines.

Based on the above, the class of MJP systems is defined as follows:

- (I) The required system performance is specified in terms of the product-mix, which may be changing frequently (e.g., on a daily basis).
- (II) The jobs are released one-by-one (without batching) according to the product-mix and a corresponding build-schedule.
- (III) All jobs undergo identical sequence of manufacturing operations, but require different amount of work at some or all operations. Therefore, the processing time of each machine is deterministic but job-dependent.
- (IV) The machines are unreliable and experience random breakdowns.
- (V) The setup times are zero.
- (VI) In-process buffers are non-dedicated (i.e., store different job-types in the sequence of arrival).
- (VII) The jobs are processed on the first-come, first-served basis.

MJP systems can be classified into two groups: *serial lines* and *assembly systems* (see Figures 1 and 2, where the circles represent the machines, the rectangles are the buffers, and J_1 and J_2 denote the two job-types being manufactured). While the serial lines may (and, often do) comprise assembly operations, it is assumed that either no starvation by subassemblies take place or the probabilities of starvations are given along with other system parameters.

In MJP assembly systems, the subassembly components are manufactured by subassembly lines, which are parts of the overall system at hand and which, as shown in Figure 2, may operate either in a *single-job production* (SJP) or MJP regime and have either dedicated finished goods buffers (FGB) for different job-types or non-dedicated ones. Therefore, in addition to the above, MJP assembly systems are characterized by the following:

- MJP assembly systems consist of a *main assembly line* (MA) and *subassembly lines* (SA_i); while the former operates in MJP regime, the latter may be either MJP or SJP.
- Subassembly lines may have dedicated or non-dedicated FGB.
- Job release in the main and subassembly lines may be coordinated (i.e., synchronized) in accordance with the build schedule or not; the former case is referred to as *build to schedule* and the latter *build to finished goods buffer*.

It should be pointed out that the term MJP is not a standard one. Obviously, MJP systems can be viewed as part of the general field of flexible multi-product manufacturing. However, since the term flexibility is used in many different connotations (including machine tool flexibility, material handling flexibility, routing flexibility, scheduling flexibility, etc.), to emphasize that we address here only production systems with flexibility in the product-mix, we use the term MJP.

MJP systems are also related to a class of flexible production systems referred to as mixed-model assembly, wherein different models of a product are assembled by the same production system. While this scenario is similar to MJP, the main assumption and the problems addressed are different. Specifically, in the mixed-model area the machines are assumed to be reliable with no breakdowns. As a result, the problems addressed are line balancing/workforce assignment and model sequencing (see Subsection 1.6 for a review). In contrast, in MJP the machines are assumed to be unreliable with random breakdowns, and the problems addressed are related to the study of

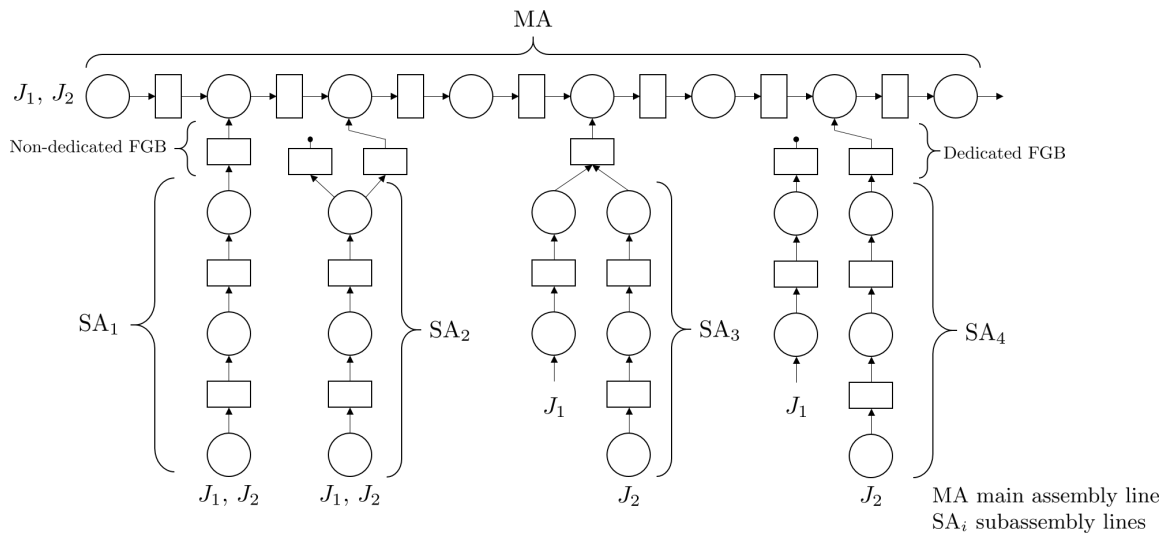


Figure 2. MJP assembly system

resulting stochastic systems, in particular, throughput and bottlenecks analyses as functions of the product-mix. Given the above, we find it beneficial to carve out the class of MJP systems defined by assumptions (I)-(VII).

1.2 Goals of the paper

In practice, performance of MJP systems is often marred by substantial losses: throughput losses of up to 15% is a typical occurrence. Recovering these losses is impeded by the lack of theory for MJP systems analysis and improvement. Therefore, the *main goals* of this paper are:

- Formulate a set of problems related to analysis and continuous improvement of MJP systems; the intention is to attract the attention of the research community to real-world open problems, solution of which would contribute substantially to both theory and practice of MJP.
- Provide solutions for some of these problems, as they pertain to MJP serial lines.
- Report the results of a six-month application project, wherein the techniques obtained to-date have been applied to develop a continuous improvement project for the underbody production system at an automotive assembly plant.

As it follows from Subsection 1.1, the main distinguishing feature of MJP in comparison with SJP is the product-mix. Since the current literature does not offer analytical tools for performance analysis of MJP systems as a function of the product-mix, the *secondary goals* of this paper are:

- Develop a method for MJP serial lines analysis, taking into account the product-mix (along with the machine and buffer parameters, habitually used in SJP systems).
- Evaluate the accuracy of this method using calculations and simulations.
- Based on this method, investigate the throughput and bottlenecks of MJP serial lines as functions of the product-mix.

Note that while this paper addresses, due to space limitations, only MJP serial lines, a similar development can be carried out for assembly systems as well (using, for example, the approach utilized in Li and Meerkov (2009) for SJP systems).

1.3 MJP problems formulation

Based on a number of continuous improvement projects carried out in the automotive industry (assembly and battery plants) during the last three years, we formulate the following problems of immediate importance for MJP operations management:

- *Performance analysis*: Given the machine, buffer, material handling system, and job-type parameters as well as the product-mix, calculate the system performance characteristics (e.g., the attainable total throughput, throughputs of various job-types, work-in-process, probabilities of machine blockages and starvations) as functions of the product-mix.
- *Continuous improvement*: Given the machine, buffer, material handling system, and job-type parameters, determine the production bottlenecks (i.e., the machines that affect the total throughput in the strongest manner) as a function of the product-mix, indicate a way to alleviate them, and quantify the resulting performance improvement.
- *Quality*: Given the machine, buffer, material handling system, and job-type parameters, determine the quality bottlenecks (i.e., the machines that impede the throughput of non-defective jobs in the strongest manner) as a function of the product-mix, indicate a way for their alleviation, and quantify the resulting improvement.
- *Leanness*: Determine the smallest capacity of the in-process buffers, which is necessary and sufficient to obtain the desired total throughput.
- *Product-mix optimization*: Given an MJP serial line or assembly system, quantify the product-mix leading to the total throughput maximization; note that while this optimal product-mix may not be utilized in all situations (e.g., due to the nature of customer demand), in some cases it offers a possibility for efficiency improvement.
- *Product-mix assignment*: Assuming that a company has several plants capable of manufacturing the products in question and given the desired company-wide product-mix, assign a product-mix to each plant, so that the overall throughput is maximized, while the required company-wide product-mix is maintained.
- *Job sequencing*: Develop a method for selecting a build-schedule, which achieves the maximum throughput corresponding to the assigned product-mix, satisfying sequence-constrained conditions, if any. This problem is widely discussed in the mixed-model assembly literature (see Subsection 1.6), and the results reported therein may be viewed as a foundation for MJP scheduling.

While the above problems refer equally to both serial lines and assembly systems, the ones below are assembly-specific:

- *Product-mix feasibility*: Given an MJP assembly system, quantify the set of feasible product-mixes. This problem arises since subassembly lines often have different production capacities as compared with that of the main assembly; this limits the attainable product-mixes in the assembly system as a whole.
- *Robustness analysis*: Investigate robustness of MJP assembly systems with respect to starvations by subassemblies. In particular, determine the optimal structure of subassembly lines (e.g., dedicated vs. non-dedicated FGBs, build-to-schedule vs. build-to-FGB, SJP vs. MJP regimes, etc.) leading to starvation minimization of the main line.

As mentioned in Subsection 1.2, this paper addresses only a small subset of these problems, using the serial MJP lines as an instance (see Subsection 1.5 for the specific problems addressed and results obtained).

1.4 Approach

The approach of this paper is based on a novel *work-based model*, instead of the traditional *part-based model*, of production systems. This implies that, unlike the traditional approach, where the analyses are carried out in terms of ‘parts produced’, in this paper the research is carried out in terms of the ‘work produced’, which is insensitive to whether a single- or multi-job manufacturing takes place. Given the work produced, the throughput of each job-type and other performance metrics can be calculated using the product-mix.

More precisely, in the work-based model (see Section 2 for details), the machines are defined by the amount of work they can carry out per unit of time. The jobs are defined by the amount of work they require at each machine. For instance, a welding operation is defined by the number of welds it can carry out per unit of time, and the jobs are specified by the number of welds they require per job-type. The reliability characteristics of the machines in the work-based model remain the same as in the part-based case, i.e., defined by distributions of up- and downtimes. The buffer capacity in the work-based model also remains the same as in the part-based case.

This model provides a foundation for solving the problems addressed in this paper.

1.5 Technical contributions of the paper

This paper reports the following results for MJP serial lines with exponential machines:

- A method for calculating MJP lines performance metrics as a function of the product-mix. This method is based on reducing MJP to SJP, manufacturing a virtual job whose work-requirements are defined by the product-mix and the original job-types involved. Simulations/calculations show that the accuracy of this method is similar to that of the SJP case (reported in Li and Meerkov (2009)).
- A method for production bottlenecks identification as a function of the product-mix. Here the bottleneck is defined as the machine whose work-capacity, weighted by the virtual job work-requirement, has the largest effect on the total throughput. Since the virtual job work-requirements depend on the product-mix, the bottleneck machine is also a function of the product-mix.
- Throughput of MJP vs. SJP. We show that if the jobs are conflicting (in the sense of having different machines as BNs when manufactured in SJP regime), there exists a range of product-mixes, where the total throughput of MJP is larger than that of SJP of any constituent job-type. This phenomenon takes place because SJP overloads respective bottlenecks, whereas MJP with the “right” product-mix leads to a more balanced work allocation.
- Product-mix performance portrait of MJP serial lines. To represent the global behavior of MJP lines, we introduce the Product-Mix Performance Portrait (PP), which characterizes the system throughput and bottlenecks for all possible product-mixes. This portrait is intended to help managing MJP lines with frequently changing product-mixes. To enable applications, we have developed a software tool, which calculates the PPs for given machines, buffers, and jobs parameters.
- Application. We used the above results in a six-month continuous improvement project for the underbody assembly system at an automotive assembly plant. Analyzing the throughput part of the resulting PP, it has been shown that the system cannot meet the daily throughput target for any of the required product-mixes. Analyzing the bottleneck part of PP, improvement measures have been suggested, resulting in the desired system throughput for all required product-mixes. These suggestions have been favorably accepted by the plant management.

1.6 *Related literature*

The MJP systems considered in this paper are related to three technical areas: serial lines, flexible multi-product systems, and mixed-model assembly. Since the literature on serial lines is summarized in over a dozen of monographs (see, for instance, Viswanadham and Narahari (1992), Askin and Standridge (1993), Papadopoulos, Heavy, and Browne (1993), Perros (1994), Gershwin (1994), Altioik (1997), Papadopoulos et al. (2009), Curry and Feldman (2009), Li and Meerkov (2009)), we overview here only the two latter areas and comment on their relationship with MJP.

The literature on flexible multi-product manufacturing, is quite extensive, with the burst of activity in 1980-2000 (see the reviews by Dupont-Gateland (1982), Browne et al. (1984), Sury (1985), Sethi and Sethi (1990), Buzacott and Yao (1996), DeToni and Tonchia (1998), Beach et al. (2000), Buzacott and Mandelbaum (2008), Boysen, Fliedner, and Scholl (2009)). By the end of this period, it became clear that the ideas of flexibility have not been adopted in industry as widely as originally expected, with the exception of flexibility in assembly. This is, perhaps, because of the fact that flexibility in assembly is relatively easy to implement (due to innately short or even zero set-up times), while flexibility in machining requires sophisticated and expensive equipment in order to ensure sufficiently short set-ups. Nevertheless, for the sake of completeness, we briefly overview below the results obtained to-date in various areas of research on production flexibility.

The publications on flexible multi-product systems can be classified into two groups: design and analysis. Representative papers on design include Kusiak (1985), Stecke (1985), Kimemia and Gershwin (1985), Dallery and Frein (1986), Heragu and Kusiak (1988), Dallery and Stecke (1990), Tetzlaff (1990) and several chapters of the monographs by Tempelmeier and Kuhn (1993) and Buzacott and Shanthikumar (1993). The main issues addressed are machine layout, flexible material handling, structures of inventory storage, flexible planning and scheduling, etc. The approaches are typically based on optimization. The results obtained provide guidance for designing flexible machining and assembly operations.

The literature on performance analysis is less extensive. Namely, publications by Yao (1983), Buzacott (1984), Yao and Buzacott (1986), and Tempelmeier, Kuhn, and Tetzlaff (1989) use queuing theory methods to investigate the performance of flexible manufacturing systems and quantify the effects of flexibility. Papers by Krieg and Kuhn (2002), Colledani, Matta, and Tolio (2005), and Colledani et al. (2008) use a decomposition approach to compute the throughput and buffer occupancy in linear and nonlinear multi-product manufacturing systems. Finally, Li and Huang (2005), Zhao, Li, and Huang (2014), Zhao and Li (2014), and Zhao and Li (2015) use the aggregation approach to investigate the throughput and bottlenecks of multi-product serial lines. This literature differs from the MJP problems considered in this work largely because it does not address the issues related to the product-mix and its effect on the performance metrics.

The literature on mixed-model assembly is also quite extensive, with activities starting in the 60s and continuing until today. As it follows from its name, it addresses systems manufacturing different models of a given product. The main problems considered in this literature are devoted to line balancing/workforce assignment and production scheduling (sequencing). The former addresses design issues related to the architecture of the production line, the number of work stations, the tasks assigned to each of them, and the number of operators and their task assignments so that the desired system cycle time is satisfied. The latter is concerned with operational issues related to sequencing model release so that the production of each model is as close to the desired product-mix as possible. In some papers, these two problems are addressed simultaneously (see, for instance, one of the first papers in the field by Thomopoulos (1967), and a paper by Merengo, Nava, and Pozzetti (1999)); in most cases, however, they are addressed separately. Below is a brief overview of this literature.

The line balancing/workforce assignment literature considers problems with various objective functions and constraints. Typical objective functions are line efficiency, number of stations, line cycle time, workforce amount and its efficiency. Typical constraints arise from precedence relations,

operation capacities, space/topological considerations, ergonomic issues, etc. The main surveys of this literature are van Zante-de Fokkert and de Kok (1997), Boysen, Fliedner, and Scholl (2008), and Battaia and Dolgui (2013). Recent representative papers on workforce assignment are Battaia et al. (2015), Bruecker et al. (2015), and Moreira et al. (2017). An example of manufacturing cell design so that workforce efficiency is maximized can be found in Klampfl, Gusikhin, and Rossi (2005). The mathematical approaches in this literature are typically based on optimization (often mixed integer programming) combined with various heuristics (simulated annealing, Tabu search, genetic algorithms, swarm intelligence, etc.). Although the research in this area is still active today (see, for instance, Delice et al. (2017), and Roshani and Nazemi (2017)), it can be viewed as a mature engineering field with a substantial practical importance.

The mixed-model scheduling literature also presents numerous techniques based on different goals and criteria involved (see the review by Boysen, Fliedner, and Scholl (2009)). Some of them are focused on leveling the schedule (Miltenburg (1989), Kubiak and Sethi (1991), Bard, Shtub, and Joshi (1994)), other intend to minimize work overload (Li, Gao, and Sun (2012)), still others utilize cyclic scheduling. The cyclic scheduling is a method for scheduling small batches so that the throughput is maximized, if the cycles are performed in perpetuity (see the reviews by Dawande et al. (2005), Brauner (2008), and Levner et al. (2010)). The approaches here are Gantt charts, PERT, and optimization algorithms coupled with heuristics. The application of cyclic scheduling to mixed-model assembly is based on the notion of minimum product set (MPS), i.e., the smallest set of products, which respects the desired model-mix. For instance, if the desired model-mix is 500 of model A and 200 of model B, the MPS is (5, 2). Given the MPS, one can use cyclic scheduling to obtain the optimal sequencing. It has been shown that this approach outperforms other approaches as far as the throughput is concerned (see Sawik, 2012).

Throughout the above-mentioned literature (with a few exceptions, see Bowman and Muckstadt (1993), Karabati and Tan (1998), Tan and Karabati (2000), and Lee and Lee (2005)), it is assumed that the operations (or “machines”) are reliable and experience no breakdowns. While this may be true in well-organized manual assembly (e.g., final assembly in automotive plants), in automated or partially automated systems (e.g., body or paint shops) this is not the case – breakdowns do occur and substantially impede systems’ performance. A contrasting assumption of MJP formulation is that the machines are unreliable, with random breakdown and repair times. The problems considered in MJP reflect this assumption: As indicated in Subsection 1.3, they are centered on the throughput, work-in-process, and bottleneck analysis of the resulting stochastic systems as functions of the product-mix. The latter feature – dependency on the product-mix – makes the MJP formulation different from the mixed-model assembly and novel in the area of performance analysis of stochastic production systems as well.

In summary, the relationship among the three areas considered above can be viewed as follows: The solution of the balancing/workforce assignment problem is the input to the mixed-model sequencing problem; similarly, the solution of the sequencing problem can be viewed as the input to MJP problems. Thus, advancements in each area would lead to advancements in flexible multi-product manufacturing as a whole.

1.7 Paper outline and abbreviations/notations

The remainder of this paper is structured as follows: Section 2 introduces the work-based model. In Section 3, this model is used for performance and bottleneck investigation of MJP serial lines. In Section 4, qualitative and quantitative behavior of the throughput and bottlenecks as functions of the product-mix is investigated. Section 5 introduces the notion of the Product-Mix Performance Portrait and illustrates its behavior for systems with two and three job-types. An application of the results obtained to an underbody assembly system at an automotive assembly plant is described in Section 6. Finally, Section 7 formulates the conclusion and topics for future research. All proofs

are included in the Appendix.

Throughout this paper, the following abbreviations and notations are used:

Abbreviations: BN – bottleneck; JPH – jobs per hour; MJP – multi-job production; PP – performance portrait; PSE – production systems engineering; SJP – single-job production.

Notations: b – buffer; BL – probability of blockage; c – machine capacity; e – machine efficiency; λ – machine breakdown rate; M – number of machines in a serial line; m – machine; μ – machine repair rate; N – buffer capacity; $\mathbf{r} = [r_1, \dots, r_S]$ – product-mix; S – number of job-types; smc_i – system-modified capacity of m_i ; ST – probability of starvation; τ – machine cycle time; TP_j – throughput of job-type j ; TP_v – throughput of virtual job; tp_i – stand-alone throughput of m_i ; W_i – work-capacity of m_i ; w_{ij} – work-requirement of job j at m_i ; WIP_i – work-in-process in b_i .

2. Work-Based Model

The work-based model of MJP serial lines (see Figure 1) is defined by the following assumptions:

- (i) Each machine $m_i, i = 1, \dots, M$, is characterized by its work-capacity, W_i (in units of work/min).
- (ii) Each job-type, $J_j, j = 1, \dots, S$, is characterized by its work-requirements $w_{ij}, i = 1, \dots, M; j = 1, \dots, S$, (in units of work/job), i.e., by the vector of work-required, $\mathbf{w}_j = [w_{1j}, \dots, w_{Mj}]$. The set-up time of each job-type is zero.
- (iii) The jobs are released one-by-one according to a given product-mix $\mathbf{r} = [r_1, \dots, r_S], \sum_{j=1}^S r_j = 1$, where r_j is the fraction of job-type J_j to be manufactured. The release sequence is formed by selecting each job-type j with probability $r_j, j = 1, \dots, S$.
- (iv) The buffers are not dedicated.

While these features of the model are novel, the remaining ones follow standard conventions used in serial lines modeling and analysis (as, for instance, in Li and Meerkov (2009)):

- (v) Machines are characterized by the breakdown and repair rates, λ_i and μ_i (in units of 1/min), respectively; this implies that the machines are exponential with the average up- and down-time given by $T_{up,i} = \frac{1}{\lambda_i}$ and $T_{down,i} = \frac{1}{\mu_i}$, and with machine efficiency $e_i = \frac{\mu_i}{\lambda_i + \mu_i}$.
- (vi) The first machine is not starved and the last machine is not blocked.
- (vii) Specific technical conventions, under which the MJP lines are analyzed in this paper, are:
 - the machines obey the blocked-before-service assumption;
 - the breakdowns are time-dependent;
 - the flow model description is used;
 - the job release also follows the flow model convention.

Discussion: (a) The machine work-capacity, W_i , is defined by the technological operation it carries out. In addition to the welding operation mentioned in Subsection 1.4, it can be the feed rate of a cutting instrument in turning, milling or drilling; the rate of etching or material deposition in semiconductor manufacturing; the number of assembly steps carried out in a robotic or manual assembly operation per unit of time, etc. The units of work in job work-requirements, w_{ij} , are the same as in the corresponding machines (but in terms of work per job-type, rather than work per unit of time).

(b) As it follows from (i) and (ii), the time necessary to process a job-type j on machine i (i.e., the *cycle time* of machine i for processing job j) is

$$\tau_{ij} = \frac{w_{ij}}{W_i}, \quad i = 1, \dots, M; j = 1, \dots, S. \quad (1)$$

While in the part-based model the cycle time is an independent variable, (1) indicates that in the

work-based model it is not: w_{ij} and W_i are the independent variables. This allows to investigate the effect of the job work-requirements on the system's throughput and bottleneck.

(c) The random job release of assumption (iii) is introduced for simplicity. The issues of performance under other job release policies (e.g., cyclic scheduling) are beyond the scope of this paper and will be addressed in future work.

(d) Also for simplicity, we use the exponential machine reliability model of assumption (v); other reliability models, e.g., Weibull, gamma, log-normal can be considered as well (using the approximations developed in Li and Meerkov (2009)).

(e) The technical conventions mentioned in (vii) are introduced to simplify the analysis and can be substituted by others discussed in the literature, e.g., those used in Gershwin (1994).

(f) Assumptions (i)-(vii) define open (non-palletized) serial lines. If the line under consideration is closed (i.e., palletized), the results obtained based on model (i)-(vii) provide an upper bound of performance for the corresponding closed line. Indeed, since pallets may lead only to impediment of the open system performance (by starving the first machine for pallets or blocking the last machine by the pallet return buffer), a closed line cannot perform better than the corresponding open line. In the future, we plan to extend the results of this paper to MJP closed lines (using the approach reported in Li and Meerkov (2009)).

(g) Model (i)-(vii) can be used for analysis of SJP systems as well. In this case, $S = 1$ and $w_{ij} = w_i$. While several results in this direction are available online¹, they are not included here due to space limitations.

The performance metrics of production systems in the framework of work-based model (i)-(vii) are as follows: $TP_j, j = 1, \dots, S$ – the average number of jobs of type j produced by the last machine per unit of time; $TP = \sum_j TP_j$ – the average total number of jobs produced by the last machine per unit of time; $WIP_i, i = 1, \dots, M - 1$ – the average number of jobs in buffer b_i ; $ST_i, i = 2, \dots, M$ – the probability that machine m_i is starved; $BL_i, i = 1, \dots, M - 1$ – the probability that machine m_i is blocked. The methods for evaluating these performance metrics and identification of bottlenecks as functions of the product-mix are described next.

3. Performance and Bottleneck Analyses of MJP Serial Lines In the Framework of Work-based Model

3.1 Performance analysis

3.1.1 Procedure

Consider an MJP serial line defined by assumptions (i)-(vii) and denote its performance metrics as $TP = \sum_{j=1}^S TP_j$, WIP_i , ST_i , and BL_i . To evaluate these characteristics, the following three-stage procedure is introduced:

Stage I: Given the work-requirements $w_{ij}, i = 1, \dots, M, j = 1, \dots, S$, define the work-requirements of the *virtual job* at machine m_i as the work imposed on m_i under a given product-mix $[r_1, \dots, r_S]$, i.e.,

$$w_{i,v} := \sum_{j=1}^S r_j w_{ij}. \quad (2)$$

Stage II: Consider the virtual SJP line consisting of the machines and buffers of the original MJP line, but manufacturing the virtual job. Denote this line as SJP_v and its performance metrics

¹<http://web.eecs.umich.edu/~simm/publications/MJP/MJPfull.pdf>

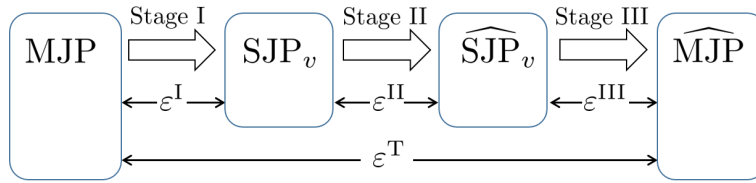


Figure 3. Flow diagram for MJP serial line performance evaluation

as TP_v , $WIP_{i,v}$, $ST_{i,v}$, and $BL_{i,v}$. Evaluate these performance metrics using the recursive aggregation procedure described in Li and Meerkov (2009) by expressions (11.40)-(11.49) (with the machine capacity c_i replaced by $W_i/w_{i,v}$). As a result, SJP_v is approximated by \widehat{SJP}_v with the performance metrics \widehat{TP}_v , $\widehat{WIP}_{i,v}$, $\widehat{ST}_{i,v}$, and $\widehat{BL}_{i,v}$.

Stage III: Calculate the estimates of the performance metrics of the original MJP line according to

$$\begin{aligned} \widehat{TP}_j &= r_j \widehat{TP}_v, j = 1, \dots, S, & \widehat{WIP}_i &= \widehat{WIP}_{i,v}, i = 1, \dots, M-1, \\ \widehat{ST}_i &= \widehat{ST}_{i,v}, i = 2, \dots, M, & \widehat{BL}_i &= \widehat{BL}_{i,v}, i = 1, \dots, M-1. \end{aligned} \quad (3)$$

3.1.2 Accuracy

The above procedure is illustrated in Figure 3. Each of its stages may introduce errors in the performance metrics estimates. We denote these errors as ε_X^I , ε_X^{II} , and ε_X^{III} , where the subscript 'X' stands for one of the four performance metrics. In order to quantify the errors of Stage III, we evaluate the errors denoted as ε_x^T , where 'T' stands for the total error between the metrics of MJP and \widehat{MJP} , and compare ε_x^T with ε_x^I and ε_x^{II} .

For Stage I, these errors are defined as follows:

$$\begin{aligned} \varepsilon_{TP}^I &= \frac{|TP - TP_v|}{TP} \cdot 100\%, & \varepsilon_{WIP}^I &= \frac{1}{M-1} \sum_{i=1}^{M-1} \frac{|WIP_i - WIP_{i,v}|}{N_i} \cdot 100\%, \\ \varepsilon_{ST}^I &= \frac{1}{M-1} \sum_{i=2}^M |ST_i - ST_{i,v}|, & \varepsilon_{BL}^I &= \frac{1}{M-1} \sum_{i=1}^{M-1} |BL_i - BL_{i,v}|, \end{aligned} \quad (4)$$

The errors ε_X^{II} , ε_X^{III} , and ε_x^T are defined similarly.

Specific MJP lines, for which these errors have been evaluated, are as follows: The values of M and S have been selected from the sets

$$M \in \{2, 3, 4, 5\}, \quad S \in \{2, 3, 4\}. \quad (5)$$

For each pair (M, S) , 2000 MJP serial lines have been created by selecting their parameters randomly and equiprobably from:

$$\begin{aligned} T_{up,i} &\in [20, 100], e_i \in [0.80, 0.99], W_i \in [0.75, 1.25], w_{ij} \in [1.0, 1.5], \\ N_i &= [k_i W_i T_{down,i}] + 1, \text{ where } k_i \in \{1, 2, 3, 4, 5\}, \\ r_j &\in [0.1, 0.9] \text{ so that } \sum_{j=1}^S r_j = 1, \end{aligned} \quad (6)$$

where $k_i = N_i \frac{\mu_i}{W_i}$ represents the number of average downtimes the buffer of capacity N_i protects machine i . Thus, the total of 24,000 lines have been constructed and evaluated using the following simulation procedure (for MJP and SJP_v): For each line, 20 discrete-event simulation runs have been carried out. Discrete-event model assumes jobs are released and processed individually one after another and their type is determined randomly according to the product-mix (as opposed to flow model in which a continuous flow of jobs is released and processed continuously). In each simulation run, the first 20,000 units of time were considered as warm-up period, and the subsequent 180,000 units of time were used to statistically evaluate TP_j^s , WIP_i^s , ST_i^s , and BL_i^s , where s is the index of the simulation run. This leads to the expected values denoted as TP_j , WIP_i , ST_i , and BL_i , with 95% confidence intervals less than 0.002 for TP , less than 0.1 for WIP_i , and less than 0.005 for ST_i and BL_i .

The results of these analyses are shown in Table 1. Examining these data, we formulate:

Observation 3.1.

- Stage I induces practically no errors in all four performance metrics for all M and S considered.
- Stage II does introduce errors in all performance metrics. The errors in TP are two-to-four times smaller than those in WIP . The errors in BL and ST are practically identical. All the

Table 1. Average errors and confidence intervals
(a) For Stage I

M	2			3			4			5		
S	2	3	4	2	3	4	2	3	4	2	3	4
$\varepsilon_{TP}^I(\%)$	0.1 ± 0.004	0.1 ± 0.004	0.1 ± 0.009	0.1 ± 0.004	0.1 ± 0.009	0.1 ± 0.009	0.1 ± 0.004	0.1 ± 0.009	0.2 ± 0.009	0.1 ± 0.004	0.1 ± 0.004	0.2 ± 0.009
$\varepsilon_{WIP}^I(\%)$	0.5 ± 0.15	0.8 ± 0.18	0.8 ± 0.13	0.4 ± 0.08	0.6 ± 0.13	0.7 ± 0.12	0.4 ± 0.07	0.5 ± 0.11	0.6 ± 0.11	0.3 ± 0.06	0.5 ± 0.08	0.5 ± 0.09
ε_{BL}^I	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001
ε_{ST}^I	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001	0.001 ± 0.0001

(b) For Stage II

M	2			3			4			5		
S	2	3	4	2	3	4	2	3	4	2	3	4
$\varepsilon_{TP}^{II}(\%)$	0.1 ± 0.004	0.1 ± 0.004	0.1 ± 0.004	1.0 ± 0.076	0.8 ± 0.067	0.8 ± 0.067	1.4 ± 0.085	1.6 ± 0.089	1.7 ± 0.089	2.1 ± 0.103	2.0 ± 0.098	2.2 ± 0.107
$\varepsilon_{WIP}^{II}(\%)$	0.2 ± 0.016	0.2 ± 0.017	0.2 ± 0.018	4.1 ± 0.76	4.2 ± 0.775	4.3 ± 0.77	6.4 ± 0.903	6.5 ± 0.998	7.0 ± 1.025	8.0 ± 1.077	8.7 ± 1.104	8.0 ± 1.011
ε_{BL}^{II}	0.01 ± 0.001	0.01 ± 0.001	0.01 ± 0.001	0.02 ± 0.001	0.02 ± 0.001	0.02 ± 0.001	0.02 ± 0.002	0.02 ± 0.002	0.02 ± 0.002	0.03 ± 0.002	0.03 ± 0.002	0.03 ± 0.002
ε_{ST}^{II}	0.01 ± 0.001	0.01 ± 0.001	0.01 ± 0.001	0.02 ± 0.002	0.02 ± 0.002	0.02 ± 0.001	0.03 ± 0.002	0.03 ± 0.002	0.03 ± 0.002	0.03 ± 0.002	0.03 ± 0.003	0.03 ± 0.002

(c) Total

M	2			3			4			5		
S	2	3	4	2	3	4	2	3	4	2	3	4
$\varepsilon_{TP}^T(\%)$	0.1 ± 0.009	0.1 ± 0.004	0.2 ± 0.009	1.0 ± 0.080	0.9 ± 0.072	0.9 ± 0.067	1.5 ± 0.085	1.7 ± 0.094	1.8 ± 0.089	2.2 ± 0.107	2.1 ± 0.103	2.3 ± 0.107
$\varepsilon_{WIP}^T(\%)$	0.5 ± 0.152	0.8 ± 0.182	0.9 ± 0.139	4.4 ± 0.757	4.6 ± 0.783	4.8 ± 0.779	6.6 ± 0.904	6.8 ± 1	7.4 ± 1.024	8.2 ± 1.074	9.0 ± 1.105	8.3 ± 1.01
ε_{BL}^T	0.011 ± 0.001	0.010 ± 0.001	0.010 ± 0.001	0.018 ± 0.001	0.018 ± 0.001	0.017 ± 0.001	0.023 ± 0.002	0.023 ± 0.002	0.024 ± 0.002	0.028 ± 0.002	0.028 ± 0.002	0.027 ± 0.002
ε_{ST}^T	0.010 ± 0.001	0.011 ± 0.001	0.010 ± 0.001	0.020 ± 0.002	0.019 ± 0.002	0.019 ± 0.002	0.026 ± 0.002	0.026 ± 0.002	0.026 ± 0.002	0.029 ± 0.002	0.032 ± 0.003	0.029 ± 0.002

errors are increasing functions of M and practically independent of S . We note that these errors are similar to those observed in evaluating asynchronous exponential SJP lines (see Li and Meerkov 2009, Section 11.2).

- Stage III introduces practically no errors. This follows from the fact that the values of ε_X^H and ε_X^T are almost the same.

Along with quantifying the three stage method for performance analysis of MJP serial lines, this observation indicates that using the flow model of system description and job release (rather than discrete-event model) does not lead to substantial errors and, thus, can be used for MJP systems analysis.

3.2 Bottleneck definition and identification

3.2.1 Definition

The issue of bottlenecks in the framework of the work-based model is addressed as follows: Recall that the bottleneck in the framework of the part-based model is defined in Li and Meerkov (2009) as the machine m_i with the largest effect on the system throughput quantified as

$$\frac{\partial TP}{\partial c_i} > \frac{\partial TP}{\partial c_j}, \quad \forall j \neq i, \quad (7)$$

where $c_k = 1/\tau_k$ is the capacity of machine m_k , and τ_k is its cycle time. Since in the work-based model the virtual cycle time is $\frac{w_{k,v}}{W_k}$ and the only variable characterizing the machine is W_k , expression (7) becomes:

$$w_{i,v} \frac{\partial TP}{\partial W_i} > w_{j,v} \frac{\partial TP}{\partial W_j}, \quad \forall j \neq i. \quad (8)$$

We use this expression as the definition of the bottleneck in MJP systems. Note that (8) implies that MJP bottlenecks depend not only on the machines and buffers, but also on the job work-requirements and the product-mix. In other words, in the same system with the same jobs, different product-mixes may lead to different bottlenecks.

3.2.2 Identification method

To identify BNs defined by (8), we use the procedure described in Li and Meerkov (2009, Section 13.2) for bottlenecks in SJP lines. This method consists of the following:

- Calculate BL_i and ST_i for the line in question.
- Place these data under the block diagram of the line and assign arrows from m_i to m_{i+1} if $BL_i > ST_{i+1}$, and from m_{i+1} to m_i , if $ST_{i+1} > BL_i$ (see Figure 4).
- If there is only one machine with no emanating arrows, it is the bottleneck in the sense of (8).
- If there are multiple machines with no emanating arrows, the one with the largest severity is the Primary BN, where the severity is defined by:

$$\begin{aligned} S_1 &= |BL_1 - ST_2|, \\ S_i &= |BL_{i-1} - ST_i| + |BL_i - ST_{i+1}|, \quad i = 2, \dots, M-1, \\ S_M &= |BL_{M-1} - ST_M|. \end{aligned} \quad (9)$$

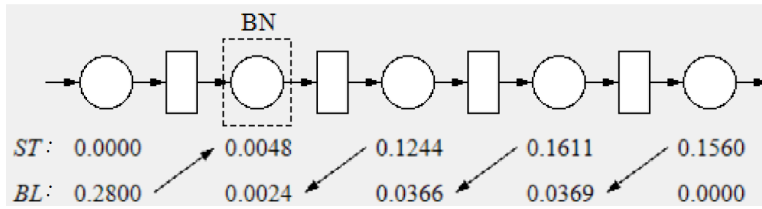


Figure 4. Illustration of BN identification using the arrow method

We use this method to determine bottlenecks of the first three systems of Figure 3 and denote them as \widehat{BN} , \widehat{BN}_v and \widehat{BN}_v , respectively. Note that the bottleneck of the fourth system, \widehat{MJP} , is the same as \widehat{BN}_v (because starvations and blockages of \widehat{SJP}_v are the same as those of \widehat{MJP} , see (3)). Also note that \widehat{BN} and \widehat{BN}_v are identified using (ST, BL) and (ST_v, BL_v) , respectively, while \widehat{BN}_v is determined by calculating $(\widehat{ST}_v, \widehat{BL}_v)$.

3.2.3 Accuracy

The discrepancy among the BNs of first three systems of Figure 3 are quantified by

$$\varepsilon_{BN}^I = \frac{\sum_{k=1}^K I^I(k)}{K} \times 100\%, \quad \varepsilon_{BN}^{II} = \frac{\sum_{k=1}^K I^{II}(k)}{K} \times 100\%, \quad \varepsilon_{BN}^T = \frac{\sum_{k=1}^K I^T(k)}{K} \times 100\%, \quad (10)$$

where $k \in \{1, \dots, K\}$ is the index of the line analyzed and $I^I(k)$ is the indicator function taking value 0 when $\widehat{BN} = \widehat{BN}_v$ and 1 otherwise. The indicator functions $I^{II}(k)$ and $I^T(k)$ are defined similarly (i.e., in terms of the discrepancies between \widehat{BN}_v and \widehat{BN}_v and between \widehat{BN} and \widehat{BN}_v , respectively).

We evaluate ε_{BN}^I , ε_{BN}^{II} , and ε_{BN}^T using the 24000 lines described in Subsection 3.1.2. The results are shown in Table 2. Part (a) of this table considers all 24,000 lines mentioned in Subsection 3.1.2. In some of these lines there is only one machine with no emanating arrows (i.e., a single BN). In others there are multiple machines with no emanating arrows and, thus, multiple BNs, one of which, with the largest severity (see (9)), is the Primary BN. To characterize each of these cases, Table 2(a) addresses systems with single and multiple BNs, while Table 2(b) considers only lines with a single BN. Based on Table 2, we formulate:

Observation 3.2.

- Stage I induces practically no errors in BN identification for all M and S considered.
- Stage II introduces significant errors in BN identification. These errors increase as a function of M and are almost independent of S . The errors of BN identification in MJP lines with a single BN are about 40% smaller than those in the case of combined single/multiple BNs.
- Since, ε_{BN}^{II} and ε_{BN}^T are almost the same, we conclude that Stage III introduces practically no errors in BN identification.

The methods developed in this section are used in the rest of this paper for performance analysis of MJP serial lines as a function of the product-mix.

Table 2. Errors of BN identification
(a) Combined single/multiple BNs

M	2			3			4			5		
S	2	3	4	2	3	4	2	3	4	2	3	4
ϵ_{BN}^I (%)	0.2	0.0	0.2	0.1	0.4	0.4	0.7	0.5	0.7	0.6	0.6	0.6
ϵ_{BN}^R (%)	1.0	0.8	0.8	3.2	3.5	3.3	7.6	6.5	7.3	9.6	10.7	11.3
ϵ_{BN}^T (%)	0.9	0.8	0.8	3.2	3.6	3.3	7.3	6.75	7.6	9.9	10.6	11.4

(b) Single BNs only

M	2			3			4			5		
S	2	3	4	2	3	4	2	3	4	2	3	4
ϵ_{BN}^I (%)	0.2	0.0	0.2	0.1	0.2	0.4	0.5	0.3	0.6	0.4	0.4	0.4
ϵ_{BN}^R (%)	1.0	0.8	0.8	2.3	2.4	2.5	4.5	3.9	4.5	4.8	6.4	6.7
ϵ_{BN}^T (%)	0.9	0.8	0.8	2.3	2.6	2.5	4.2	4.1	4.7	5.1	6.3	6.6
# lines	2000	2000	2000	1901	1900	1899	1770	1766	1771	1629	1615	1626

4. MJP Systems Throughput and Bottlenecks as Functions of Product-Mix

4.1 Preliminaries

The throughput, bottlenecks, and other characteristics of SJP serial lines as functions of machine and buffer parameters have been investigated in numerous studies (see, for instance, Li and Meerkov 2009, and references therein). Since, as it is shown in Section 3, performance analysis of MJP systems can be reduced to analysis of SJP lines (manufacturing virtual jobs), the results of these studies are applicable to MJP lines as well. Thus, the only open issue is the behavior of the throughput and bottlenecks as functions of the product-mix. The current section is intended to investigate this issue. We begin with the case of $S = 2$ job-types and then generalize the results for $S > 2$. Note that for $S = 2$, the product-mix is defined by a scalar $r = r_1$ (since $r_2 = 1 - r_1$) and, therefore, $TP_v(r) = TP_v(r)$.

For $\mathbf{N} = \mathbf{0}$ and $\mathbf{N} = \infty$, (where $\mathbf{N} = [N_1, \dots, N_{M-1}]$ and $\mathbf{0}$ and ∞ indicate zero and infinite buffers, respectively), the function $\widehat{TP}_v(r) = TP_v(r)$ can be evaluated quite easily – without using the aggregation procedure of Li and Meerkov (2009, Section 11.2). Indeed, for $\mathbf{N} = \infty$, $TP_v(r)$ can be calculated as

$$TP_v(r, \mathbf{N} = \infty) = \min\{tp_{1,v}(r), \dots, tp_{M,v}(r)\}, \quad (11)$$

where $tp_{i,v}(r)$ is the virtual *stand-alone throughput* of machine i given by

$$tp_{i,v}(r) = \frac{e_i W_i}{w_{i,v}(r)} = \frac{e_i W_i}{r w_{i1} + (1 - r) w_{i2}}. \quad (12)$$

For $\mathbf{N} = \mathbf{0}$, $TP_v(r)$ is

$$TP_v(r, \mathbf{N} = \mathbf{0}) = \min\{smc_{1,v}(r), \dots, smc_{M,v}(r)\}, \quad (13)$$

where $smc_{i,v}(r)$ is the so-called virtual *system-modified capacity* of machine i given by

$$smc_{i,v}(r) = \frac{W_i}{w_{i,v}(r)} \prod_{i=1}^M e_i = \frac{W_i}{r w_{i1} + (1 - r) w_{i2}} \prod_{i=1}^M e_i. \quad (14)$$

Clearly, $\widehat{TP}_v(r, \mathbf{N})$ for all other \mathbf{N} 's is upper- and lower-bounded by (11) and (13), respectively, i.e.,

$$TP_v(r, \mathbf{N} = \mathbf{0}) \leq \widehat{TP}_v(r, \mathbf{N}) \leq TP_v(r, \mathbf{N} = \infty). \quad (15)$$

In this section, we first analyze the behavior of (11) and (13) as functions of the product-mix and then address the case of finite buffers in more details.

It turns out that both qualitative and quantitative properties of (11) and (13) depend on the relationship between the jobs work-requirements. To characterize this relationship, consider an MJP serial line producing two job-types, J_1 and J_2 , with work-requirements, w_{i1} and w_{i2} , $i = 1, \dots, M$, respectively, and with product-mix r . Let the bottleneck of this line be the machine denoted as BN_{J_1} when $r = 1$, and BN_{J_2} when $r = 0$. Note that in all systems of Figure 3, the bottleneck for $\mathbf{N} = \infty$ is the machine with the smallest stand-alone throughput, and for $\mathbf{N} = \mathbf{0}$ the machine with the smallest system-modified capacity.

Definition 4.1. Given a serial MJP line defined by assumptions (i)-(vii), jobs J_1 and J_2 are called *non-conflicting* if $BN_{J_1} = BN_{J_2}$. Otherwise the jobs are *conflicting*.

In Subsections 4.2 and 4.3 below, we characterize the behavior of the throughput and bottlenecks of MJP lines for non-conflicting and conflicting jobs, respectively.

4.2 Throughput and bottlenecks of MJP serial lines with non-conflicting jobs

4.2.1 The case of two job-types

Theorem 4.1. Consider an MJP serial line defined by assumptions (i)-(vii) and producing two non-conflicting jobs, J_1 and J_2 , with $BN_{J_1} = BN_{J_2} = m_k$. Then, if all buffers are of infinite or zero capacity,

- (a) m_k is the $BN_v(r)$ for all $r \in [0, 1]$;
- (b) $TP_v(r)$ can be calculated using

$$TP_v(r) = \frac{1}{\frac{r}{TP_{J_1}} + \frac{1-r}{TP_{J_2}}}. \quad (16)$$

- (c) $TP_v(r)$ is:

- strictly monotonically increasing if $TP_{J_1} > TP_{J_2}$;
- strictly monotonically decreasing if $TP_{J_1} < TP_{J_2}$;
- constant if $TP_{J_1} = TP_{J_2}$.

Proof: See the Appendix.

For finite buffer capacity, \mathbf{N} , statements (a) and (c) of this theorem have been verified numerically. Specifically, we have constructed 25,000 non-conflicting MJP lines with five machines producing two job-types. The machine and job-type parameters have been selected randomly and equiprobably from the sets defined in (6). For each of these lines, $\widehat{TP}_v(r, \mathbf{N})$ and $\widehat{BN}_v(r, \mathbf{N})$ have been evaluated using the method of Section 3 for $r \in \{0, 0.01, 0.02, \dots, 1\}$. As a result, we obtain the following:

Observation 4.1. Among the 25,000 lines with non-conflicting jobs analyzed:

- $\widehat{BN}_v(r, \mathbf{N})$, is the same machine for all $r \in [0, 1]$ in 99.2% of cases.
- $\widehat{TP}_v(r, \mathbf{N})$ satisfies statement (c) of Theorem 4.1 in 92.8% of cases.

Statement (b) of Theorem 4.1 shows that $TP_v(r)$ for $\mathbf{N} = \mathbf{0}$ and $\mathbf{N} = \infty$ can be represented not only by (11) and (13), but by (16) as well. This representation allows to evaluate the throughput for any buffer capacity. Indeed, motivated by (16), introduce the following approximation for the

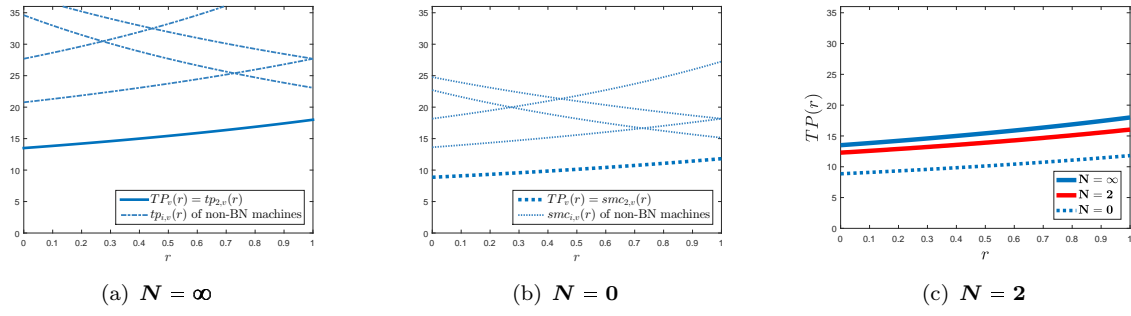


Figure 5. Throughput of a five-machine MJP line with non-conflicting jobs

throughput of a serial MJP line with arbitrary buffer capacity vector \mathbf{N} :

$$\widetilde{TP}_v(r, \mathbf{N}) = \frac{1}{\frac{r}{\widehat{TP}_{J_1}(\mathbf{N})} + \frac{1-r}{\widehat{TP}_{J_2}(\mathbf{N})}}, \quad (17)$$

where $\widehat{TP}_{J_1}(\mathbf{N})$ and $\widehat{TP}_{J_2}(\mathbf{N})$ denote the system throughput operating in the SJP regime, which can be evaluated using the method of Section 3. Note that while \widetilde{TP}_v represents the throughput for any $r \in [0, 1]$, it requires calculating only $\widehat{TP}_{J_1}(\mathbf{N})$ and $\widehat{TP}_{J_2}(\mathbf{N})$.

To investigate the accuracy of approximation (17), we used the same 25,000 MJP lines as before, along with their $\widetilde{TP}_v(r, \mathbf{N})$. In addition, we computed $\widetilde{TP}_v(r)$ using (17) and quantified its accuracy by

$$\varepsilon_{\widetilde{TP}} = \max_{0 < r < 1} \left\{ \frac{|\widehat{TP}_v(r, \mathbf{N}) - \widetilde{TP}_v(r, \mathbf{N})|}{\widehat{TP}_v(r)} \right\} \cdot 100\%. \quad (18)$$

As a result, we obtain:

Observation 4.2. Among the 25,000 lines with non-conflicting jobs analyzed, $\varepsilon_{\widetilde{TP}} \leq 1\%$ in 95.8% of cases and $\varepsilon_{\widetilde{TP}} \leq 2\%$ in 98.5% of cases.

Thus, expression (17) offers an efficient way for calculating the throughput of MJP lines with two non-conflicting jobs for any buffer capacity.

Example 4.1. Consider an MJP line with five identical machines, defined by $\lambda_i = 0.01$, $\mu_i = 0.09$, $W_i = 1$, producing two jobs, J_1 and J_2 , with the following work-requirements: $\mathbf{w}_1 = [1.3, 3.0, 2.3, 1.9, 1.9]$, $\mathbf{w}_2 = [1.9, 4.0, 1.6, 1.4, 2.6]$. For $\mathbf{N} = \infty$ and $\mathbf{N} = \mathbf{0}$, the jobs are non-conflicting and the common BN is m_2 . Thus, for $\mathbf{N} = \infty$, according to (11), $TP_v(r) = tp_{2,v}(r)$ for all $r \in [0, 1]$; the graph of $TP_v(r)$ is shown in Figure 5(a), along with the stand-alone throughputs of non-bottleneck machines. Similarly, for $\mathbf{N} = \mathbf{0}$, $TP_v(r) = smc_{2,v}(r)$ for all $r \in [0, 1]$; the graph of $TP_v(r)$ is shown in Figure 5(b), along with the system-modified capacities of other machines. Finally, when the capacity of all buffers is equal to 2, the graph of $\widetilde{TP}_v(r, \mathbf{N})$, calculated according to (17), is shown in Figure 5(c), along with its upper- and lower-bounds, $TP_v(r, \mathbf{N} = \infty)$, $TP_v(r, \mathbf{N} = \mathbf{0})$. Clearly, in this MJP system, even small buffers of capacity 2 lead to the throughput within 12% of that for $\mathbf{N} = \infty$ for all r . ■

4.2.2 The case of more than two job-type

Theorem 4.2. Consider an MJP serial line defined by assumptions (i)-(vii) and producing S pairwise non-conflicting jobs, J_1, \dots, J_S with $BN_{J_1} = \dots = BN_{J_S} = m_k$. Then, if all buffers are of infinite or zero capacity,

- (a) m_k is the $BN_v(\mathbf{r})$ for all product-mixes r_1, \dots, r_S , $\sum_{j=1}^S r_j = 1$;
- (b) $TP_v(\mathbf{r})$ can be calculated using

$$TP_v(\mathbf{r}) = \frac{1}{\frac{r_1}{TP_{J_1}} + \dots + \frac{r_S}{TP_{J_S}}}, \quad \sum_{j=1}^S r_j = 1, \quad (19)$$

where TP_{J_j} is the throughput of the line producing job-type J_j in SJP regime.

Proof: Similar to that of Theorem 4.1.

For finite buffer capacity, N , the TP can be evaluated using an expression analogous to (17):

$$\widetilde{TP}_v(\mathbf{r}) = \frac{1}{\frac{r_1}{\widehat{TP}_{J_1}(N)} + \dots + \frac{r_S}{\widehat{TP}_{J_S}(N)}}, \quad \sum_{j=1}^S r_j = 1, \quad (20)$$

where $\widehat{TP}_{J_j}(N)$, $j = 1, \dots, S$, is the throughput of the line in question manufacturing job-type J_j in SJP regime. These results offer a possibility for analyzing MJP serial lines with non-conflicting jobs for any number of job-types.

4.3 Throughput and bottlenecks of MJP serial lines with conflicting jobs

The behavior of MJP serial lines with conflicting jobs is more complex than that with non-conflicting ones. Therefore, we begin with the simplest case of two-machine lines and then generalize the results to systems with $M > 2$.

4.3.1 Two-machine lines

Theorem 4.3. Consider an MJP serial line defined by assumptions (i)-(vii) and producing two conflicting jobs, J_1 and J_2 , having $BN_{J_1} = m_1$ and $BN_{J_2} = m_2$. Then, if the buffer is of infinite or zero capacity,

- (a) $BN_v(r)$ has one switch on the interval $r \in [0, 1]$:

$$BN_v(r) = \begin{cases} m_2, & \text{if } r \in [0, r^*), \\ m_1, & \text{if } r \in (r^*, 1], \end{cases} \quad (21)$$

where r^* is the unique solution of $tp_1(r) = tp_2(r)$ for $N = \infty$ or $smc_1(r) = smc_2(r)$ for $N = 0$.

- (b) $TP_v(r)$ can be calculated using

$$TP_v(r) = \begin{cases} \frac{r^*}{\frac{r}{TP_v(r^*)} + \frac{r^*-r}{TP_{J_2}}}, & \text{if } 0 \leq r < r^*, \\ \frac{1-r^*}{\frac{r-r^*}{TP_{J_1}} + \frac{1-r}{TP_v(r^*)}}, & \text{if } r^* \leq r \leq 1. \end{cases} \quad (22)$$

- (c) This function of r is:

- non-monotonic if $\{w_{11} > w_{12}, w_{21} < w_{22}\}$, so that $TP_v(r^*) > \max\{TP_{J_1}, TP_{J_2}\}$;

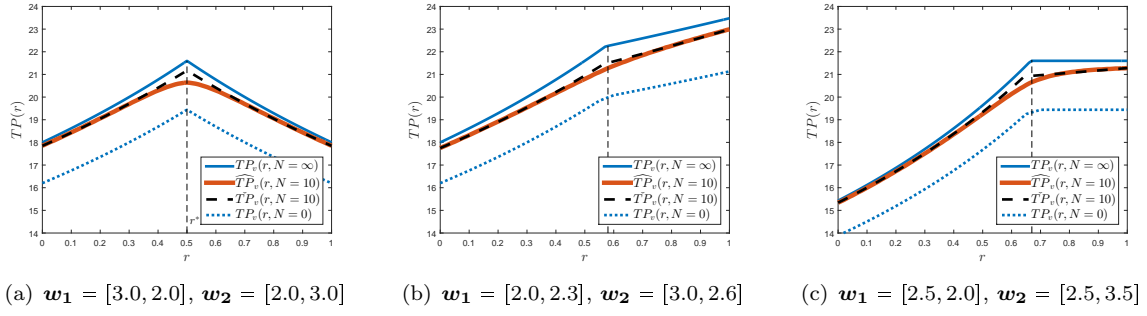


Figure 6. Patterns of throughput behavior for a two-machine line producing two conflicting jobs

- *strictly monotonically increasing* if $\{w_{11} > w_{12}, w_{21} > w_{22}\}$ and *decreasing* if $\{w_{11} < w_{12}, w_{21} < w_{22}\}$;
- *non-strictly monotonic* if $w_{i1} = w_{i2}$ for either $i = 1$ or 2 , so that $TP_v(r^*) = TP_{J_1}$ or $TP_v(r^*) = TP_{J_2}$.

Proof: See the Appendix.

The first bullet in part (c) implies that *there exists a range of product-mixes, where the throughput of MJP line is larger than the throughput of each individual job-type manufactured in SJP regime.* This phenomenon occurs because SJP regimes of J_1 and J_2 overload their respective BNs, whereas MJP with the “right” product-mix leads to a more balanced work allocation on both BNs and, thus, to a higher throughput.

For finite buffer capacity, statements (a) and (c) of Theorem 4.3 have been verified using the same approach as in Subsection 4.2. Specifically, we have constructed 25,000 two-machine MJP lines with two conflicting job-types and evaluated their $\widehat{TP}_v(r, N)$ and $\widehat{BN}_v(r, N)$ using the method of Section 3 for $r \in \{0, 0.01, 0.02, \dots, 1\}$. As a result, we obtained the following:

Observation 4.3. *Among the 25,000 lines with conflicting jobs analyzed:*

- $\widehat{BN}_v(r, N)$, $r \in [0, 1]$, *switches once in 99.1% of cases.*
- *If $w_{11} > w_{12}$ and $w_{21} < w_{22}$, then there exist r' and r'' such that $\widehat{TP}_v(r, N) > \max_r \{\widehat{TP}_{J_1}(N), \widehat{TP}_{J_2}(N)\}$ for all $r \in [r', r'']$ in 76% of cases.*
- *If the difference between job work-requirements is sufficiently large, so that $\frac{w_{11}-w_{12}}{0.5(w_{11}+w_{12})} > 0.1$ and $\frac{w_{22}-w_{21}}{0.5(w_{22}+w_{21})} > 0.1$, the inequality $\widehat{TP}_v(r, N) > \max_r \{\widehat{TP}_{J_1}(N), \widehat{TP}_{J_2}(N)\}$, $r \in [r', r'']$, takes place in 96.7% of cases.*

Statement (b) of Theorem 4.3 can also be generalized for arbitrary buffer capacity. This is accomplished as follows: Consider a two-machine MJP line with a buffer of capacity N and r^* defined by $tp_1(r) = tp_2(r)$. Using the method of Section 3, evaluate $\widehat{TP}_v(r, N)$ for the following four values of r : $r = 0$, $\frac{r^*}{2}$, $\frac{1+r^*}{2}$, and 1. Introduce the estimate of throughput, $\widetilde{TP}_v(r, N)$, $\forall r \in [0, 1]$, as follows:

$$\widetilde{TP}_v(r, N) = \min_{0 \leq r \leq 1} \left\{ \frac{0.5r^*}{\frac{r}{\widehat{TP}(0.5r^*, N)} + \frac{0.5r^*-r}{\widehat{TP}_{J_2}(N)}}, \frac{0.5(1-r^*)}{\frac{r-0.5(1+r^*)}{\widehat{TP}_{J_1}(N)} + \frac{1-r}{\widehat{TP}(0.5(1+r^*), N)}} \right\} \quad (23)$$

Quantifying the accuracy of (23) by error (18), we obtain:

Observation 4.4. *Among the 25,000 lines with conflicting jobs analyzed, $\varepsilon_{\widetilde{TP}} \leq 2\%$ in 95.6% of cases and $\varepsilon_{\widetilde{TP}} \leq 3\%$ in 99.9% of the cases.*

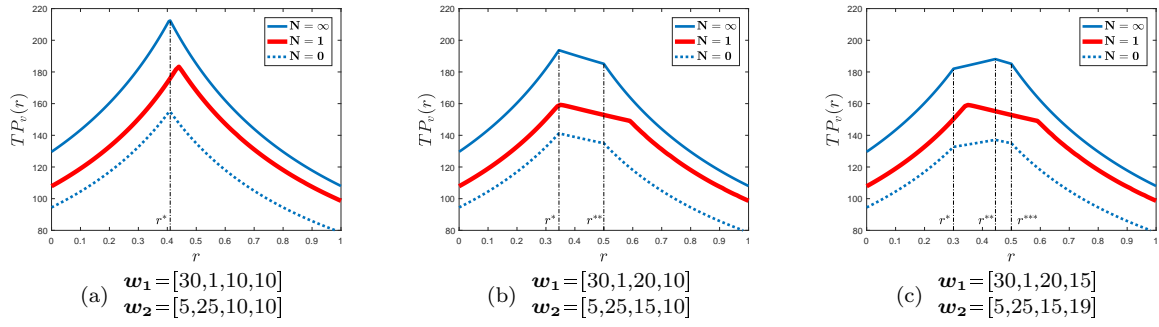


Figure 7. Patterns of throughput behavior for a four-machine line producing two conflicting jobs

Example 4.2. Consider a two-machine MJP line with $\lambda_i = 0.01$, $\mu_i = 0.09$, $W_i = 1$, $i = 1, 2$, and with two conflicting jobs having work-requirements corresponding to the three cases of part (c) of Theorem 4.3. For each set of the work-requirements, Figure 6 shows $TP_v(r)$, evaluated for $N = 0$ and $N = \infty$ using (22). As one can see, under the conditions of the first case, there exists a set R of r 's (which includes the point of the bottleneck switch r^*), such that $TP_v(r) > \max\{TP_{J1}, TP_{J2}\}, \forall r \in R$, and $TP_v(r^*)$ is 22% larger than TP_{J1} and TP_{J2} . Parts (b) and (c) of Figure 6 illustrate the behavior of $TP_v(r)$ for two other cases. The behavior of $\widehat{TP}_v(r, N = 10)$ (evaluated using the method of Section 3) and $\widetilde{TP}_v(r, N = 10)$ (calculated using (23)) are also shown in Figure 6. As one can see, $\widetilde{TP}_v(r, N = 10)$ provides an accurate approximation of $\widehat{TP}_v(r, N = 10)$, except for a small neighborhood of r^* . ■

4.3.2 $M > 2$ -machine lines

For MJP lines with more than two machines and conflicting jobs, numerous patterns of $TP_v(r)$ and BN_v behavior are possible. The theorem below quantifies these patterns.

Theorem 4.4. Consider an MJP serial line defined by assumptions (i)-(vii) with $M > 2$ machines producing two conflicting jobs, J_1 and J_2 . Then, if all buffers are of infinite or zero capacity,

(a) $BN_v(r)$ has at most $M - 1$ switches in the interval $r \in [0, 1]$; each machine can be a bottleneck only in a single interval of $[0, 1]$;

(b) $TP_v(r)$ has the following properties:

- if the number of switches of $BN_v(r)$ is $1 \leq K \leq M - 1$, then $TP_v(r)$, $r \in [0, 1]$ has $K + 1$ intervals of continuous differentiability; the $BN_v(r)$ switches occur at the values of r , where $TP_v(r)$ is non-differentiable (the value of K is referred to as the order of conflict);
- if $w_{BN_{J1,1}} > w_{BN_{J1,2}}$ and $w_{BN_{J2,2}} > w_{BN_{J2,1}}$, then there exist r' and r'' such that $TP_v(r) > \max\{TP_{J1}, TP_{J2}\}, \forall r \in (r', r'')$.

Proof: See the Appendix.

Note that a characterization of the throughput similar to (23) is possible for $M > 2$ as well. However, this characterization becomes too involved, and, therefore, is not pursued here.

Example 4.3. To illustrate Theorem 4.4, consider a four-machine MJP line with $\lambda_i = 0.01$, $\mu_i = 0.09$, $W_i = 1$, manufacturing two jobs with work-requirements indicated in Figure 7. Using (11) and (13) we evaluate $TP_v(r)$ for $N = \infty$ and $N = 0$. The results are shown in Figure 7, illustrating three types of the order of conflicts and the resulting patterns of $TP_v(r)$ behavior. For $0 < N < \infty$, $\widehat{TP}_v(r, N)$ can be evaluated for various $r \in [0, 1]$ using the method of Section 3. The results for $N = 1$ are shown in Figure 7 as well. ■

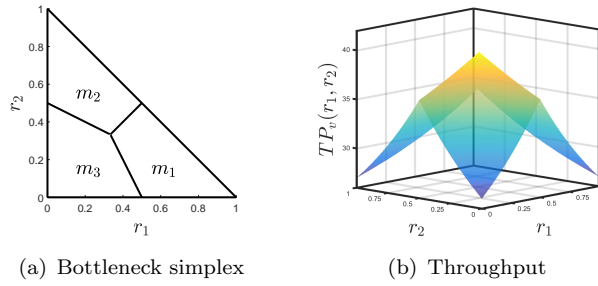


Figure 8. Illustration of bottleneck positions and the throughput in a serial line with three pairwise conflicting job-types

4.3.3 $S > 2$ job-types

In the case of $S > 2$ job-types, the throughput and bottlenecks are functions of $S - 1$ independent variables, r_1, \dots, r_{S-1} (with $r_S = 1 - \sum_{j=1}^{S-1} r_j$), i.e., $TP = TP(r_1, \dots, r_{S-1})$, $BN = BN(r_1, \dots, r_{S-1})$. This implies that both TP and BNs are defined not on the interval $[0, 1]$, as in the case of $S = 2$, but on a $(S - 1)$ -dimensional simplex in \mathbb{R}^{S-1} , referred to as the *product-mix simplex*. The following statement characterizes BNs in $S > 2$ case.

Theorem 4.5. *Consider an MJP serial line defined by assumptions (i)-(vii) with $M \geq 2$ machines, infinite or zero buffers and manufacturing $S \geq 3$ job-types, at least two of which are conflicting. Consider also the product-mix simplex $\mathbb{S} \in \mathbb{R}^{S-1}$, defined by $\sum_{j=1}^S r_j = 1$. Then, there are at most M non-intersecting simply connected open sets in \mathbb{S} , such that each machine is the BN in at most one of them.*

Proof: See the Appendix. Note that this proof provides also a way for calculating the open sets mentioned in the theorem.

If the buffers are finite, BNs and TP for any product-mix can be calculated using the method of Section 3.

Example 4.4. To illustrate Theorem 4.5, consider a five-machine MJP line with $\lambda_i = 0.01$, $\mu_i = 0.09$, $W_i = 1$, and infinite buffers, manufacturing three pairwise conflicting job-types with the following work-requirements:

$$\mathbf{w}_1 = [2, 1, 1, 1, 1], \quad \mathbf{w}_2 = [1, 2, 1, 1, 1], \quad \mathbf{w}_3 = [1, 1, 2, 1, 1].$$

For this system, the above-mentioned product-mix simplex and the open sets of Theorem 4.5 are shown in Figure 8(a). Clearly, in this system, m_4 and m_5 are never bottlenecks, while three other machines are bottlenecks for the indicated open sets. Note that in this case, the closure of the open sets constitutes a partitioning of \mathbb{S} . The resulting throughput as a function of r_1 and r_2 is shown in Figure 8(b). It indicates, in particular, that if $r_1 = r_2 = r_3 = 1/3$, the TP_v is 40.5 JPH, while for $r_1 = 0.6, r_2 = 0.1, r_3 = 0.3$, the TP_v is 16.7% smaller (33.75 JPH).

5. Product-mix Performance Portraits

In System Science, the global behavior of dynamical systems is often represented by a state-space portrait (SSP), (see Khalil 2002), which is a set of system trajectories for various initial conditions. The SSP succinctly represents the main properties of the system – its steady states, stability, limit cycles, and even strange attractors – in a single picture. Control system engineers often use SSP in order to design a controller, which forces the system operation in a desired regime.

Similar to SSP, it is possible to represent the global behavior of MJP systems by their portraits

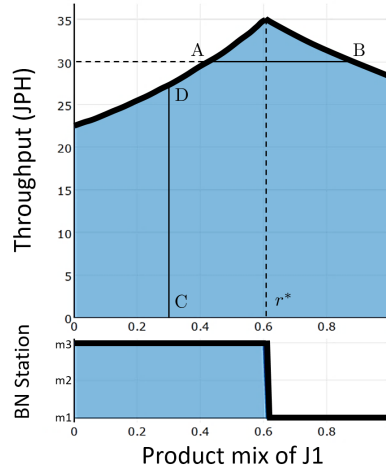


Figure 9. PP of MJP serial line with two conflicting job-types

with respect to the product-mix. We refer to this representation as the *Product-Mix Performance Portrait* or just *Performance Portrait* (PP). It consists of two graphs: the *throughput graph*, which shows \widehat{TP}_v as a function of the product-mix, and the *bottleneck graph*, which shows \widehat{BN}_v , also as a function of the product-mix. The purpose of this section is to discuss the PP and outline its utility for operations management and control.

We have created a software tool for calculating and displaying PPs. The calculations are based on the method of Section 3. Several screenshots of this tool are shown in Figures 9 and 10 for $S = 2$ and $S = 3$, respectively. Each of these figures is elucidated below.

Figure 9 represents the PP of the MJP serial line with the following parameters:

$$\begin{aligned} \lambda &= [0.05, 0.05, 0.06, 0.04, 0.09], & \mu &= [0.95, 0.95, 0.94, 0.96, 0.91], \\ \mathbf{W} &= [1, 1, 1, 1, 1], & \mathbf{N} &= [1, 2, 1, 3], \\ \mathbf{w}_1 &= [2, 1, 1, 1, 1], & \mathbf{w}_2 &= [1, 1.75, 1.75, 1, 1]. \end{aligned}$$

As indicated in this figure, it shows PP with respect to J_1 ; the PP with respect to J_2 is the symmetric image of the one shown. For the purposes of explanation, we have added to this PP two lines, AB and CD.

The shaded area in the throughput graph is the *feasibility domain*: for every r , it represents all attainable \widehat{TP}_v 's. The line AB indicates all product-mixes, for which \widehat{TP}_v of at least 30 JPH can be obtained. The line CD indicates all the throughputs, which can be attained for the product-mix corresponding to point C. Finally, r^* is the product-mix, for which $\widehat{TP}_v(r)$ is maximized.

The bottleneck graph represents the bottleneck machine for each product-mix. It indicates the most efficient way for system improvement (as far as the machines are concerned). For example, if the desired product-mix is $r_d \in [0, r^*)$, the most effective way of system improvement is to improve operation m_3 ; if $r_d \in (r^*, 1]$, m_1 should be improved.

For MJP systems with $S > 2$, an additional feature is introduced in PP to avoid the difficulties of three and higher dimensional representations. Specifically, for $S = 3$, we introduce a one-handle sliding bar, which enables to modify the ratio of non-primary job-types. Figure 10 provides an

illustration of such a PP for a system with parameters given by

$$\begin{aligned} \lambda &= [0.05, 0.05, 0.06, 0.04, 0.09], & \mu &= [0.95, 0.95, 0.94, 0.96, 0.91], \\ \mathbf{W} &= [1, 1, 1, 1, 1], & \mathbf{N} &= [1, 1, 1, 1], \\ \mathbf{w}_1 &= [1, 3, 1, 1, 2.1], & \mathbf{w}_2 &= [1, 1, 3, 1, 2.1], & \mathbf{w}_3 &= [2, 1, 2.5, 1, 2.1]. \end{aligned}$$

The rows of Figure 10 show the system performance for the primary job-type being, J_1 , J_2 , and J_3 , respectively, while the columns show the performance for various allocations of the non-primary job-types (0.2/0.8, 0.5/0.5, and 0.8/0.2, respectively), defined by the one-handle sliding bar at the bottom of each graph. The meaning and the significance of the throughput and the bottleneck graphs, remain the same as in the case of $S = 2$.

For the case of $S > 3$, performance portraits can be constructed in a similar manner. Namely, instead of using one-handle sliding bar, we use an $(S - 2)$ -handle sliding bar partitioned (by the handles) into $S - 1$ regions, each representing the ratios of non-primary r_j 's, whereas the primary one is the argument of the throughput and bottleneck graphs.

The utility of PP for operations management is that for any assigned product-mix, the manager can see which level of throughput can be achieved and, if necessary, which operation(s) should be most profitably improved, if a higher throughput is required to meet the daily production target.

6. Application

The methods developed in this paper have been applied in a six-month project devoted to analysis and potential improvements of a section of the underbody assembly system at an automotive assembly plant. This section, which we refer to as Line MA (where MA stands for the Main Assembly), has been the bottleneck of the body shop for a long time, consistently producing about 15% less than its daily target. The goal of the project was to identify the reasons for these losses and suggest steps for their elimination.

Line MA consists of 11 automated welding operations and a conveyor material handling system. Although it is capable of producing four different products, over 98% of the product-mix (which changes daily) is comprised of two job-types. Therefore, we assume $S = 2$. Operations 3, 4, 7, and 8 require subassemblies. While the system is closed (with respect to the carriers, transporting the jobs from one operation to another), during the six months of the study no blockage or starvation by carriers have been reported; therefore, Line MA is modeled as an open line.

The system performance has been monitored by the Production Monitoring System, which has been used to collect the data for this study. Based on these data, a mathematical model of Line MA has been constructed and validated (with the weekly error of throughput in the range of 1 to 9%, with the average 5.1%). Due to confidentiality reasons, we are displaying here modified data, although this modification has been carried out so that the qualitative features of the system at hand are preserved.

The modified machine and job parameters for Line MA are shown in Table 3 for Week 1; the data for other weeks are similar. The capacity of all buffers between the machines is 1. Based on these data, the bottlenecks of Line MA in SJP regimes are $BN_{J_1} = OP1$ and $BN_{J_2} = OP2$. Thus, jobs J_1 and J_2 are conflicting.

Using the data of Table 3, we obtain the PP of Line MA shown in Figure 11(a). As one can see, the maximum throughput is 60 JPH, reached for $r \in [0.25, 0.40]$. When r is close to 0 or 1, throughput is 46 JPH and 24 JPH, respectively. The bottlenecks are OP2 for $r \in [0, 0.23]$, OP6 for $r \in [0.23, 0.39]$, and OP1 for $r \in [0.39, 1]$.

The modified daily target for Line MA, as indicated by the thick horizontal line in Figure 11(a) (and in the subsequent figures), is 55 JPH, with the product-mix in the range $r \in [0.25, 0.50]$. As

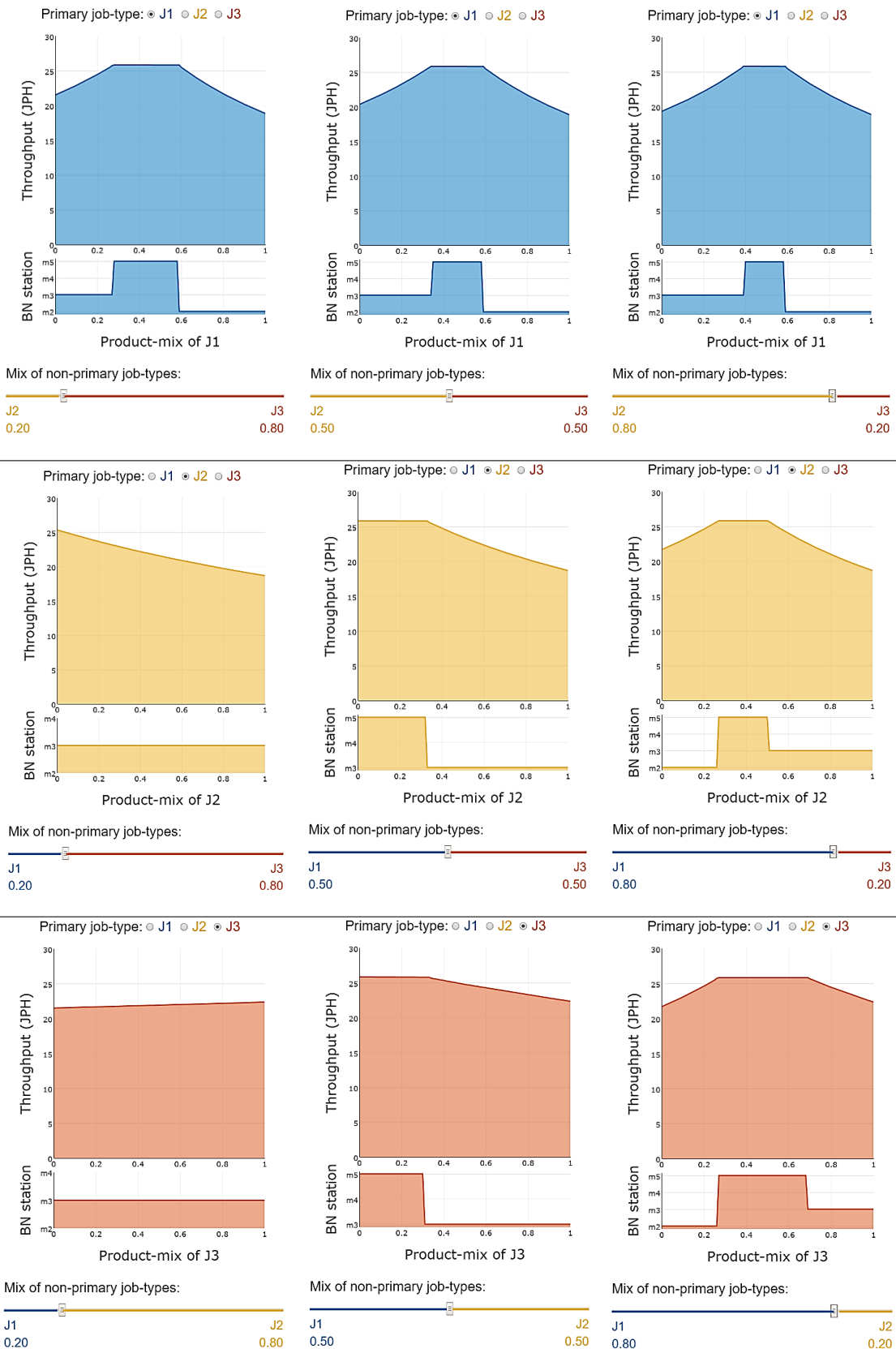


Figure 10. PP of MJP serial line with three job-types

Table 3. Line MA machines and jobs parameters for Week 1 (modified due to confidentiality reasons)

Operation i	1	2	3	4	5	6	7	8	9	10	11
$\lambda_i \left(\frac{1}{\text{min}}\right)$	0.007	0.023	0.015	0.001	0.062	0.083	0.037	0.078	0.009	0.002	0.006
$\mu_i \left(\frac{1}{\text{min}}\right)$	0.339	2.857	0.541	20.000	0.282	0.571	0.377	0.408	0.952	20.000	0.267
e_i	0.98	0.99	0.97	1.00	0.82	0.87	0.91	0.84	0.99	1.00	0.98
$\tau_{i1} = \frac{w_{i1}}{W_i}$ (sec)	150.0	1.0	39.5	30.3	30.6	44.9	45.5	42.8	40.7	23.8	33.9
$\tau_{i2} = \frac{w_{i2}}{W_i}$ (sec)	1.0	75.0	43.3	24.5	27.3	42.7	44.5	43.8	38.2	22.5	35.1

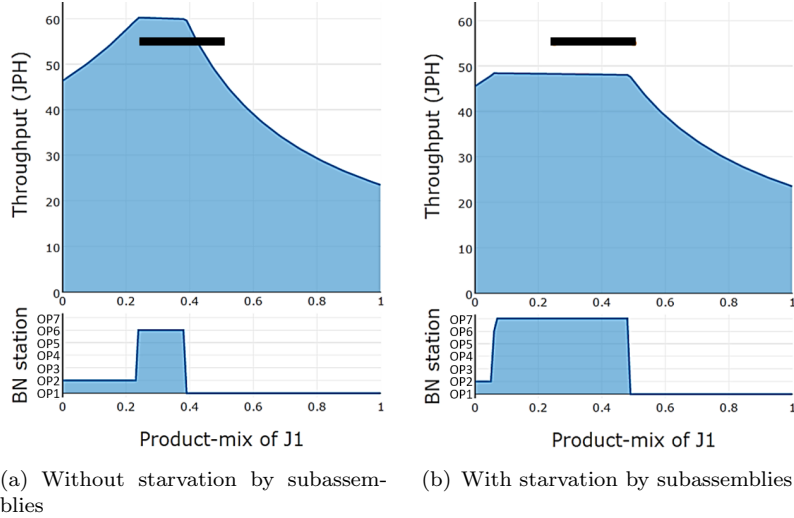


Figure 11. PP of Line MA

this line indicates, this target can be achieved only for $r \in [0.25, 0.42]$. Thus, for $r \in (0.42, 0.50]$ production is below the target, up to 15% (for $r = 0.50$).

These performance characteristics are achieved assuming that Line MA is not starved by subassemblies. In reality, however, these starvations do take place. Table 4 shows the probabilities of starvation by the subassemblies. To take them into account, we modify the breakdown rates of the respective operations of Line MA by introducing the starvation-induced efficiency, e'_i , of operation i as follows:

$$e'_i = \frac{\mu'_i}{\lambda_i + \mu'_i} = \frac{\mu_i}{\lambda_i + \mu_i} \cdot (1 - ST_i^{sub}), \quad (24)$$

where μ'_i is the adjusted breakdown rate and ST_i^{sub} is the probability of operation i starvation by its subassembly. From this relationship, the adjusted breakdown rate, μ'_i , is

$$\mu'_i = \frac{\lambda_i e_i (1 - ST_i^{sub})}{1 - e_i (1 - ST_i^{sub})}. \quad (25)$$

The values of μ'_i and e'_i are also shown in Table 4. Note that this approach is akin to that used in Sevastyanov (1962).

Using the data of Table 3 with μ_i and e_i substituted by μ'_i and e'_i from Table 4, we obtain the PP of Line MA with the starvations by subassemblies taken into account (see Figure 11(b)). As the horizontal line indicates, the throughput is 13-15% below the target for all $r \in [0.25, 0.50]$ and the bottleneck shifts to OP7. Thus, improvement of OP7 is necessary.

As it follows from Table 4, OP7 suffers from significant starvations by subassemblies. Examining the reasons for these starvations, it has been determined that they are mostly due to late delivery

Table 4. Starvations of MA by subassemblies and adjusted machine parameters

Operation i	1	2	3	4	5	6	7	8	9	10	11
ST_i^{sub}	–	–	0.02	0.02	–	–	0.18	0.04	–	–	–
e'_i	0.98	0.99	0.95	0.98	0.82	0.87	0.75	0.81	0.99	1.00	0.98
μ_i	0.339	2.857	0.309	0.0718	0.282	0.571	0.108	0.324	0.952	20.0	0.267

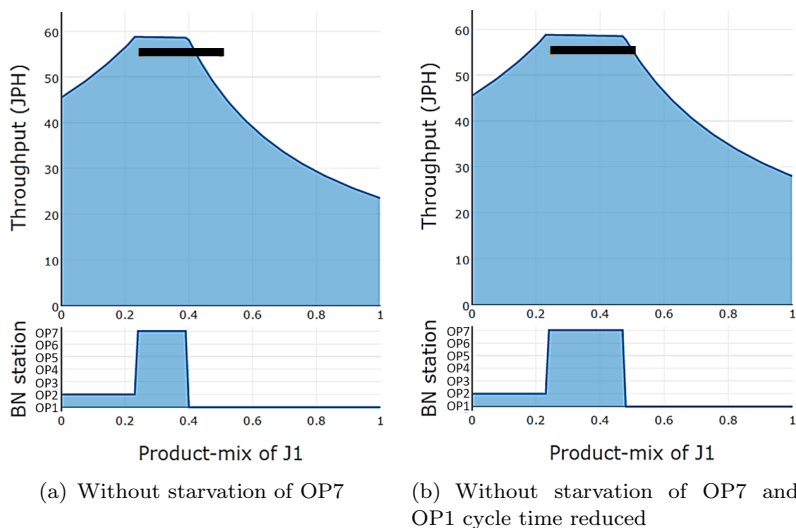


Figure 12. PP of improved Line MA

of the subassemblies by in-plant delivery trucks and by tardiness in manual loading operations. To investigate the effect of improvement of subassembly delivery and loading, we have calculated PP with $ST_7^{sub} = 0$, shown in Figure 12(a). As one can see from this PP (which is quite similar to that of Figure 11(a)), eliminating ST_7^{sub} allows for meeting the daily target with product-mix $r \in [0.25, 0.42]$. To meet the daily target for $r \in (0.42, 0.50]$, one must improve the corresponding bottleneck, which is OP1. Since OP1 has high efficiency and no starvation by subassemblies, the only venue of improvement is to reduce its cycle time. Reducing the cycle time for J_1 of OP1 by 15% allows the system to satisfy its daily target for all product-mixes, as shown in Figure 12(b). Note that cycle time of OP1 for J_2 does not need to be reduced.

The above recommendations have been communicated to the plant management and found their favorable acceptance.

7. Conclusions and Future Work

This paper addressed a class of multi-product flexible manufacturing systems, referred to as MJP, where all job-types are processed by the same sequence of machines, but with different processing time at some or all operations. The MJP system performance is defined in terms of a required product-mix, which may be changing on a daily basis.

For this class of systems, the paper introduced a set of industrially motivated problems of analysis and continuous improvement, solutions of which would allow to increase MJP systems efficiency. For some of these problems, pertaining to MJP serial lines with exponential machines, the paper provides solutions. Specifically, it investigates the behavior of the throughput and bottlenecks as functions of the product-mix. The results obtained lead to a possibility of calculating the so-called Product-mix Performance Portrait, which represents the system behavior for all feasible product-mixes and which can be used for operations management and improvement.

Obviously, these are just initial results, and numerous problems remain open. Some of them are as follows:

- Extension of the results obtained in this paper for open serial MJP lines in the following directions:
 - lines with release based on cyclic scheduling;
 - closed lines (with carriers);
 - lines with non-perfect quality machines;
 - lines with rework;
 - lean buffering design;
- Analysis and improvement methods for MJP assembly systems. (Note that the underbody manufacturing analyzed in Section 6 is, in fact, an assembly system; however, in the study reported here we reduced it to a serial line using the measured probabilities of its starvation by subassemblies.)
- Robustness properties of MJP assembly systems. As illustrated in Figure 2, there may be different configurations of subassembly lines supplying the main assembly. Which one of these configurations is the most robust with respect to various perturbations, e.g., machine downtime, release errors, etc.? Answering this question would lead to novel approaches to MJP assembly systems design.

Solutions of the problems mentioned above will lead to a relatively complete and practical theory of MJP systems analysis, design, and continuous improvement.

Acknowledgement

The authors are grateful to two anonymous reviewers for their helpful and constructive comments. We also thank the Editor-in-Chief for providing important references related to this work.

This work has been supported, in part, by the National Institute of Standards and Technology under the Award Number 70NANB16H017.

References

- Altiok, T. 1997. *Performance Analysis of Manufacturing Systems*. Springer-Verlag, New York, NY.
- Askin, R. G., and C. R. Standridge. 1993. *Modeling and Analysis of Manufacturing Systems*. Wiley, New York, NY.
- Bard, J. F., A. Shtub, and S. B. Joshi. 1994. "Sequencing mixed-model assembly lines to level parts usage and minimize line length." *Int. J. Prod. Res.* 32 (10): 2431–2454.
- Battaia, O., X. Delorme, A. Dolgui, J. Hagemann, A. Horlemann, S. Kovalev, and S. Malyutin. 2015. "Workforce minimization for a mixed-model assembly line in the automotive industry." *Int. J. Prod. Econ.* 170: 489–500.
- Battaia, O., and A. Dolgui. 2013. "A taxonomy of line balancing problems and their solution approaches." *Int. J. Prod. Econ.* 142: 259–277.
- Beach, R., A. P. Muhlemann, D. H. R. Price, A. Peterson, and J. A. Sharp. 2000. "A review of manufacturing flexibility." *Eur. J. Oper. Res.* 122: 41–57.
- Bowman, R. A., and J. A. Muckstadt. 1993. "Stochastic analysis of cycle schedules." *Oper. Res.* 41: 947–958.
- Boysen, N., M. Fliedner, and A. Scholl. 2008. "Assembly line balancing: Which model to use when?" *Int. J. Prod. Econ.* 111: 509–528.
- Boysen, N., M. Fliedner, and A. Scholl. 2009. "Sequencing mixed-model assembly lines: survey, classification, and model critique." *Eur. J. Oper. Res.* 192: 349–373.
- Brauner, N. 2008. "Identical part production in cyclic robotic cells: concepts, overview, and open questions." *Discrete Appl. Math.* 156: 2480–2492.

- Browne, J., D. Dubois, K. Rathmill, S. P. Sethi, and K. E. Stecke. 1984. "Classification of flexible manufacturing systems." *FMS Magazine* 114–117.
- Bruecker, P. De, J Van den Bergh, J. Belien, and E. Demeulemeester. 2015. "Workforce planning incorporating skills: state of the art." *Eur. J. Oper. Res.* 243: 1–16.
- Buzacott, J. A. 1984. "Modeling flexible manufacturing systems." In *Operational Research*, edited by J.P. Brans, 546–560.
- Buzacott, J. A., and M. Mandelbaum. 2008. "Flexibility in manufacturing and services: achievements, insights and challenges." *Flex. Serv. Manuf. J.* 20: 13–58.
- Buzacott, J. A., and J. G. Shanthikumar. 1993. *Stochastic Model of Manufacturing Systems*. Prentice Hall.
- Buzacott, J. A., and D. D. Yao. 1996. "Flexible manufacturing systems: A review of analytical models." *Manag. Sci.* 32: 890–905.
- Colledani, M., F. Gandola, A. Matta, and T. Tolio. 2008. "Performance evaluation of linear and non-linear multi-product mult-stage lines with unreliable machines and finite homogeneous buffers." *IIE Trans.* 40: 612–626.
- Colledani, M., A. Matta, and T. Tolio. 2005. "Performance of production lines with finite buffer capacity producing to different products." *OR. Spectr.* 27: 243–263.
- Curry, G. L., and R. M. Feldman. 2009. *Manufacturing Systems Modeling and Analysis*. Springer. Chinese translation 2011.
- Dallery, Y., and Y. Frein. 1986. "An efficient method to determine the optimal configuration of a flexible manufacturing system." *Ann. Op. Res.* 15: 207–225.
- Dallery, Y., and K Stecke. 1990. "On the optimal allocation of servers and workloads in closed queuing networks." *Op. Res.* 38: 694–703.
- Dawande, M., H. N. Geismar, S. P. Sethi, and C. Sriskandarajah. 2005. "Sequencing and scheduling in robotic cells: recent developments." *J. Sched.* 5: 387–496.
- Delice, Y., E. K. Aydogan, U. Ozcan, and M. S. Ilkay. 2017. "A modified particle swarm optimization algorithm to mixed-model two-sided assembly line balancing." *J. Intel. Manuf.* 28: 23–36.
- DeToni, A., and S. Tonchia. 1998. "Manufacturing flexibility: a literature review." *Int. J. Prod. Res.* 36: 1587–1617.
- Dupont-Gatelmand, C. 1982. "A survey off flexible manufacturing systems." *J. Manuf. Syst.* 1 (1).
- Gershwin, S. B. 1994. *Manufacturing Systems Engineering*. Prentice Hall, Englewood Cliff, NJ.
- Heragu, S. S., and A. Kusiak. 1988. "Machine layout problem in flexible manufacturing systems." *Op. Res.* 36: 258–268.
- Karabati, S., and B. Tan. 1998. "Stochastic cyclic scheduling problem in synchronous assembly and production lines." *J. Oper. Res. Soc.* 49: 1173–1187.
- Khalil, H. 2002. *Nonlinear Systems*. Prentice Hall.
- Kimemia, J., and S Gershwin. 1985. "Flow optimization in flexible manufacturing systems." *Int. J. Prod. Res.* 23: 81–96.
- Klumpfl, E., O. Gusikhin, and G. Rossi. 2005. "Optimization of workcell layouts in a mixed-model assembly line environment." *Int. J. Flex. Manuf.* 17: 277–299.
- Krieg, G., and H. Kuhn. 2002. "A decomposition method for multi-product with set-up times and lost sales." *IIE Trans.* 34: 613–625.
- Kubiak, W., and S. Sethi. 1991. "A note on "level schedules for mixed-model assembly lines in just-in-time production systems"." *Management Science* 37 (1): 121–122.
- Kusiak, A. 1985. "Flexible manufacturing systems: a structural approach." *Int. J. Prod. Res.* 23: 1057–1073.
- Lee, Y.-D., and T.-E. Lee. 2005. "Stochastic cyclic flow lines with blocking: Markovian models." *OR Spectrum* 27: 551–568.
- Levner, E., V. Kats, D. A. L. de Pablo, and T. C. E. Cheng. 2010. "Complexity of cyclic scheduling problems: a state-of-the-art survey." *Comput. Ind. Eng.* 59: 352–361.
- Li, J., J Gao, and L. Sun. 2012. "Sequencing minimum product sets on mixed-model U-lines to minimize work overload." *Int. J. Prod. Res.* 50 (18): 4977–4993.
- Li, J., and N. Huang. 2005. "Modeling and analysis of multiple product manufacturing systems with split and merge." *Int. J. Prod. Res.* 43: 4049–4066.
- Li, J., and S. M. Meerkov. 2009. *Production Systems Engineering*. Springer. Chinese translation, 2012.
- Merengo, C., F. Nava, and A. Pozzetti. 1999. "Balancing and sequencing manual mixed-model assembly lines." *Int. J. Prod. Res.* 37 (12): 2835–2860.

- Miltenburg, J. 1989. "Level schedules for mixed-model assembly lines in just-in-time production systems." *Management Science* 35 (2): 192–207.
- Moreira, M. C. O., R. Pastor, A. M. Costa, and C. Miralles. 2017. "The multi-objective assembly line worker integration and balancing problem of type 2." *Comp. Oper. Res.* 82: 114–125.
- Papadopoulos, H. T., C. Heavy, and J. Browne. 1993. *Queueing Theory in Manufacturing Systems Analysis and Design*. Chapman and Hill, London, UK.
- Papadopoulos, T. C., M. E. J. O'Kelly, M. I. Vidalis, and D. Spinellis. 2009. *Analysis and Design of Discrete Part Production Lines*. Springer.
- Perros, H. G. 1994. *Queueing Networks with Blocking*. Oxford University Press, Oxford, UK.
- Roshani, A., and F. G. Nazemi. 2017. "Mixed-model multi-manned assembly line balancing problem: a mathematical model and a simulated annealing approach." *Assembly Automation* 37: 34–50.
- Sethi, A. K., and S. P. Sethi. 1990. "Flexibility in manufacturing: a survey." *Int. J. Flex. Manuf. Syst.* 2: 289–328.
- Sevastyanov, B. A. 1962. "Influence of storage bin capacity on the average standstill time of a production line." *Theory of Probability and Application* 429–438.
- Stecke, K. E. 1985. "Designing, planning, scheduling, and control problems in flexible manufacturing systems." *Ann. Op. Res.* 3: 1–12.
- Sury, R. 1985. "An overview of evaluative models of flexible manufacturing systems." *Ann Oper. Res.* 3: 13–21.
- Tan, B., and S. Karabati. 2000. "Modeling and scheduling of an asynchronous cyclic production line with multiple parts." *J. Oper. Res. Soc.* 51: 1296–1308.
- Tempelmeier, H., and H. Kuhn. 1993. *Flexible Manufacturing Systems: Decision Support for Design and Operation*. Wiley.
- Tempelmeier, H., H. Kuhn, and U. Tetzlaff. 1989. "Performance evaluation of flexible manufacturing systems with blocking." *Int. J. Prod. Res.* 27: 1693–1709.
- Tetzlaff, U. 1990. *Optimal Design of Flexible Manufacturing Systems*. Physica-Verlag.
- Thomopoulos, N. T. 1967. "Line balancing-sequencing for mixed-model assembly." *Management Science* 14 (2): B59–B75.
- van Zante-de Fokkert, J. I., and T. G. de Kok. 1997. "The mixed and multi model line balancing problem: a comparison." *Eur. J. Oper. Res.* 100: 399–419.
- Viswanadham, N., and Y. Narahari. 1992. *Performance Modeling of Automated Manufacturing Systems*. Prentice Hall, Englewood Cliff, NJ.
- Yao, D. D. 1983. "Queueing Models of Flexible Manufacturing Systems." PhD diss., Department of Industrial Engineering, Univ. of Toronto.
- Yao, D. D., and J. A. Buzacott. 1986. "Models of flexible manufacturing systems with limited local buffers." *Int. J. Prod. Res.* 24: 107–118.
- Zhao, C., and J. Li. 2014. "Analysis of Multi-Product Manufacturing Systems with Homogeneous Exponential Machines." *IEEE Transactions on Automation Science and Engineering* 11: 828–838.
- Zhao, C., and J. Li. 2015. "Analysis and improvement of multiproduct Bernoulli serial lines: Theory and application." *IEEE Trans Syst., Man and Cybern.* 45: 1218–1230.
- Zhao, C., J. Li, and N. Huang. 2014. "Performance Evaluation of Multi-Product Manufacturing Systems with Asynchronous Exponential Machines." In *Proceedings of IEEE International Conference on Automation Science and Engineering*, edited by J.P. Brans, Taipei, Taiwan, 692–697.

Appendix A. Proofs

Proof of Theorem 4.1: For $N = \infty$, BN-machine for each job, j , is the machine with the smallest stand-alone throughput, given by

$$tp_{ij} = \frac{e_i W_i}{w_{ij}}, \quad i = 1, \dots, M, j = 1, 2. \quad (\text{A.1})$$

Since m_k is the common BN for J_1 and J_2 , $tp_{k1} < tp_{i1}$ and $tp_{k2} < tp_{i2}$, $\forall i \neq k$. Substituting (A.1) into these inequalities and inverting them, we have

$$\frac{w_{k1}}{e_k W_k} > \frac{w_{i1}}{e_i W_i}, \quad \forall i \neq k, \quad (\text{A.2})$$

$$\frac{w_{k2}}{e_k W_k} > \frac{w_{i2}}{e_i W_i}, \quad \forall i \neq k. \quad (\text{A.3})$$

Multiplying both sides of (A.2) by r , and (A.3) by $(1-r)$, and adding the inequalities, we obtain:

$$\frac{rw_{k1} + (1-r)w_{k2}}{e_k W_k} > \frac{rw_{i1} + (1-r)w_{i2}}{e_i W_i}, \quad \forall i \neq k, r \in [0, 1]. \quad (\text{A.4})$$

Since the numerators in (A.4) are the work contents of the virtual jobs (see (2)), the above inequality can be rewritten as

$$\frac{w_{k,v}(r)}{e_k W_k} > \frac{w_{i,v}(r)}{e_i W_i}, \quad \forall i \neq k, r \in [0, 1], \quad (\text{A.5})$$

which implies that

$$tp_{k,v}(r) < tp_{i,v}(r), \forall i \neq k, r \in [0, 1]. \quad (\text{A.6})$$

Clearly, expression (A.6) shows that m_k is the BN machine of the virtual job for all $r \in [0, 1]$. This proves part (a) of the theorem.

Using (11) and (12), part (b) is proved as follows:

$$\begin{aligned} TP_v(r) &= \min_{1 \leq i \leq M} \{tp_{i,v}(r)\} = \min_{1 \leq i \leq M} \left\{ \frac{e_i W_i}{w_{i,v}(r)} \right\} = \frac{e_k W_k}{w_{k,v}(r)} \\ &= \frac{e_k W_k}{rw_{k1} + (1-r)w_{k2}} = \frac{1}{r \frac{w_{k1}}{e_k W_k} + (1-r) \frac{w_{k2}}{e_k W_k}} = \frac{1}{\frac{r}{TP_{J1}} + \frac{1-r}{TP_{J2}}}. \end{aligned} \quad (\text{A.7})$$

Monotonicity properties of part (c) follow directly from the derivative of (16) with respect to r :

$$\frac{\partial TP_v(r)}{\partial r} = \frac{TP_{J1} - TP_{J2}}{TP_{J1} TP_{J2} \left(\frac{r}{TP_{J1}} + \frac{1-r}{TP_{J2}} \right)^2}. \quad (\text{A.8})$$

For $N = 0$, the bottleneck and throughput are evaluated using the same steps as above, replacing tp_{ij} by $smc_{ij} = W_i/w_{ij}$ and using (13) and (14). \blacksquare

Proof of Theorem 4.3: For $N = \infty$, $\text{BN}(r = 1) = m_1$ implies that $tp_{1,v}(1) < tp_{2,v}(1)$. Similarly, $\text{BN}(r = 0) = m_2$ implies $tp_{1,v}(0) > tp_{2,v}(0)$. Since, as it follows from (12), $tp_{i,v}(r)$ is a continuous function of $r \in [0, 1]$, there exist r^* , such that $tp_{1,v}(r^*) = tp_{2,v}(r^*)$. Solving for r^* , we obtain

$$r^* = \frac{\frac{w_{22}}{e_2 W_2} - \frac{w_{12}}{e_1 W_1}}{\frac{w_{11} - w_{12}}{e_1 W_1} + \frac{w_{22} - w_{21}}{e_2 W_2}}, \quad (\text{A.9})$$

which is unique as long as $w_{11} \neq w_{12}$ or $w_{21} \neq w_{22}$. Thus, TP_v is characterized by

$$TP_v(r) = \min_r \{tp_{i,v}(r)\} = \begin{cases} tp_{1,v}(r), & \text{if } r^* \leq r \leq 1, \\ tp_{2,v}(r), & \text{if } 0 \leq r < r^*. \end{cases} \quad (\text{A.10})$$

Using (8), it follows from (A.10) that BN of the line is m_1 for $r \in (r^*, 1]$, and m_2 for $r \in [0, r^*)$. This proves part (a).

To prove part (b), we generalize (12) for any $r_1 < r_2$ (rather than $r_1 = 0$ and $r_2 = 1$). This is accomplished as follows:

$$\begin{aligned} tp_{i,v}(r) &= \frac{e_i W_i}{r w_{i1} + (1-r) w_{i2}} = \frac{1}{\frac{r}{tp_{i1}} + \frac{1-r}{tp_{i2}}} = \frac{r_2 - r_1}{\frac{r(r_2 - r_1)}{tp_{i1}} + \frac{(1-r)(r_2 - r_1)}{tp_{i2}}} \\ &= \frac{r_2 - r_1}{\frac{rr_2 - rr_1}{tp_{i1}} + \frac{r_2 - r_1 - rr_2 + rr_1}{tp_{i2}}} = \frac{r_2 - r_1}{\frac{rr_2 - rr_1 + r_1 r_2 - r_1 r_2}{tp_{i1}} + \frac{r_2 - r_1 - rr_2 + rr_1 + r - r + r_1 r_2 - r_1 r_2}{tp_{i2}}} \\ &= \frac{r_2 - r_1}{(r - r_1) \left(\frac{r_2}{tp_{i1}} + \frac{1-r_2}{tp_{i2}} \right) + (r_2 - r) \left(\frac{r_1}{tp_{i1}} + \frac{1-r_1}{tp_{i2}} \right)} \\ &= \frac{r_2 - r_1}{\frac{r-r_1}{tp_{i,v}(r_2)} + \frac{r_2-r}{tp_{i,v}(r_1)}}. \end{aligned} \quad (\text{A.11})$$

The last equality in (A.11) is obtained by taking into account that $tp_{i,v}(r_1) = \frac{1}{r_1/tp_{i1} + (1-r_1)/tp_{i2}}$ and $tp_{i,v}(r_2) = \frac{1}{r_2/tp_{i1} + (1-r_2)/tp_{i2}}$.

Each $tp_{i,v}(r)$ in (A.10) can be rewritten using (A.11). Specifically, for $tp_{1,v}(r)$, set $r_1 = r^*$ and $r_2 = 1$, and for $tp_{2,v}(r)$, set $r_1 = 0$ and $r_2 = r^*$. This proves part (b).

For part (c), differentiating (12) with respect to r gives

$$\frac{\partial tp_{i,v}(r)}{\partial r} = \frac{e_i W_i (w_{i1} - w_{i2})}{(r w_{i1} + (1-r) w_{i2})^2}. \quad (\text{A.12})$$

As a result, if $w_{i1} > w_{i2}$, then $tp_{i,v}(r)$ is a decreasing function of r . Similarly, if $w_{i1} < w_{i2}$, then $tp_{i,v}(r)$ is an increasing function of r . Finally, if $w_{i1} = w_{i2}$, then $tp_{i,v}(r)$ is a constant. With conditions given in the first bullet of part (c), for $r \in [0, r^*]$, $TP_v(r) = tp_{2,v}(r)$ is an increasing function of r ; similarly, for $r \in [r^*, 1]$, $TP_v(r) = tp_{1,v}(r)$ is decreasing. Therefore, $TP_v(r)$ is non-monotonic. Furthermore, r^* yields maximum throughput, because of the monotonicity of each constituent part of $TP_v(r)$. This proves bullet one. Other bullets can be proved similarly.

For $N = 0$, the proof is similar to the above, using $smc_{i,v}(r)$ instead of $tp_{i,v}(r)$. The BN switch point in this case is

$$r^* = \frac{\frac{w_{22}}{W_2} - \frac{w_{12}}{W_1}}{\frac{w_{11} - w_{12}}{e_1 W_1} + \frac{w_{22} - w_{21}}{W_2}}. \quad (\text{A.13})$$

To prove Theorem 4.4, we need the following auxiliary statements: ■

Lemma A.1. *Every pair of functions $(tp_{i,v}(r), tp_{j,v}(r))$, defined by (12), has at most one intersection on $r \in [0, 1]$, unless the two functions are identical in the sense that $tp_{i,v}(r) = tp_{j,v}(r), \forall r \in [0, 1]$.*

Proof: If $tp_{i1} < tp_{j1}$ and $tp_{i2} < tp_{j2}$, then for all $r \in (0, 1)$, $\frac{r}{tp_{i1}} > \frac{r}{tp_{j1}}$ and $\frac{1-r}{tp_{i1}} > \frac{1-r}{tp_{j1}}$. Thus, $tp_{i,v}(r) < tp_{j,v}(r)$, and, therefore, $tp_{i,v}(r)$ and $tp_{j,v}(r)$ do not intersect in $(0, 1)$. Similar result takes

place when $tp_{i1} > tp_{j1}$ and $tp_{i2} > tp_{j2}$.

If $tp_{i1} < tp_{j1}$ and $tp_{i2} > tp_{j2}$, then solving $tp_{i,v}(r) = tp_{j,v}(r)$ yields

$$\frac{r}{tp_{i1}} + \frac{1-r}{tp_{i2}} = \frac{r}{tp_{j1}} + \frac{1-r}{tp_{j2}}, \quad (\text{A.14})$$

which has a unique solution $r^* = \frac{tp_{j2}^{-1} - tp_{i2}^{-1}}{tp_{i1}^{-1} - tp_{i2}^{-1} - tp_{j1}^{-1} + tp_{j2}^{-1}}$. If $tp_{i1} > tp_{j1}$ and $tp_{i2} < tp_{j2}$, solving $tp_{i,v}(r) = tp_{j,v}(r)$ yields similar results. ■

Lemma A.2. *The equality $tp_{i,v}(r) = \min\{tp_{1,v}(r), \dots, tp_{M,v}(r)\}$ takes place on at most one interval of $r \in [0, 1]$.*

Proof: As follows from Lemma A.1, since every pair of functions, $tp_{i,v}(r)$ and $tp_{j,v}(r)$, intersects only once, at, say, r^* , one of the following takes place:

(α) $tp_{i,v}(r) < tp_{a,v}(r)$, for $r \in [0, r^*)$, or

(β) $tp_{i,v}(r) < tp_{b,v}(r)$, for $r \in (r^*, 1]$.

If (α) takes place, $tp_{i,v}(r)$ cannot be the minimum for any $r \in (r^*, 1]$, because at least $tp_{a,v}(r)$ has smaller values in this range. Similar statement holds for (β). Now, consider machine i with $tp_{i,v}(r)$, for which no machine l with $tp_{l,v}(r) < tp_{i,v}(r), \forall r \in [0, 1]$ exists (if such machine exists, then $tp_{i,v}(r)$ is not the minimum). Let $tp_{a_1}(r), \dots, tp_{a_s}(r)$ be tp functions, which intersect with $tp_{i,v}$ at $r_{a_1} < \dots < r_{a_s}$, and satisfy item (α) above. Similarly, let $tp_{b_1}(r), \dots, tp_{b_t}(r)$ be tp functions, which intersect with $tp_{i,v}$ at $r_{b_1} < \dots < r_{b_t}$, and satisfy item (β). Then, as stated earlier, $tp_{i,v}$ cannot be the minimum in $\bigcup_{j=1}^s (r_{a_j}, 1] = (r_{a_1}, 1]$. It also cannot be the minimum in $\bigcup_{k=1}^t [0, r_{b_k}) = [0, r_{b_t})$. Thus, if $r_{b_t} < r_{a_1}$, then $tp_{i,v}(r) = \min\{tp_{1,v}(r), \dots, tp_{M,v}(r)\}$ only at the interval (r_{b_t}, r_{a_1}) . ■

Proof of Theorem 4.4: As it follows from Lemma A.2, there can be no more than M intervals in which different machines are the bottlenecks. Thus, at most $M - 1$ switch points exist. This proves part (a).

For part (b), let $r_1 < \dots < r_K$ be product-mixes at which bottlenecks switch from one machine to another and i_0, \dots, i_K the indices of the corresponding bottleneck machines (i.e., K switches and $K + 1$ bottlenecks). Defining $r_0 = 0, r_{K+1} = 1$, the bottleneck is given by

$$BN(r) = \begin{cases} m_{i_0}, & \text{if } r_0 \leq r < r_1, \\ m_{i_1}, & \text{if } r_1 < r < r_2, \\ \dots & \\ m_{i_K}, & \text{if } r_K < r \leq r_{K+1}, \end{cases}$$

which implies the throughput given by

$$TP_v(r) = \begin{cases} tp_{i_0}(r), & \text{if } r_0 \leq r < r_1, \\ tp_{i_1}(r), & \text{if } r_1 < r < r_2, \\ \dots & \\ tp_{i_K}(r), & \text{if } r_K < r \leq r_{K+1}. \end{cases}$$

Since, as it follows from (12), $tp_i(r)$ is continuously differentiable on $[0, 1]$ and at $r = r_k, k = 1, \dots, K, tp_{i_k}(r) = tp_{i_{k+1}}(r), TP_v(r)$ is piecewise differentiable on $[0, 1]$. This proves the first bullet of part (b).

Under conditions of the second bullet of part (b), $tp_{i_0,v}(r)$ is increasing on $[0, r_1]$ and $tp_{i_K,v}(r)$ is decreasing on $[r_K, 1]$. If $TP_{J_1} < TP_{J_2}$, then for any $r \in (0, r_1), TP_v(r) > TP_{J_2} > TP_{J_1}$. Similarly, if $TP_{J_1} > TP_{J_2}$, then for any $r \in (r_K, 1), TP_v(r) > TP_{J_1} > TP_{J_2}$. Finally, if $TP_{J_1} = TP_{J_2}$, for any $r \in (0, r_1) \cup (r_K, 1), TP_v(r) > TP_{J_1} = TP_{J_2}$. This proves the second bullet of part (b). ■

Proof of Theorem 4.5: Assume that all buffers are of infinite capacity. Define the set $\mathbb{S}_{k,i} = \{\mathbf{r} \in \mathbb{S} | tp_{k,v}(\mathbf{r}) < tp_{i,v}(\mathbf{r})\}$, which is a hyperplane in \mathbb{S} , obtained by solving the strict inequality $tp_k(\mathbf{r}) < tp_i(\mathbf{r})$, or, equivalently,

$$\frac{\sum_{j=1}^{S-1} r_j(w_{kj} - w_{kS}) + w_{kS}}{e_k W_k} > \frac{\sum_{j=1}^{S-1} r_j(w_{ij} - w_{iS}) + w_{iS}}{e_i W_i}. \quad (\text{A.15})$$

Then, the set \mathbb{S}_k , where m_k is the BN, is given by

$$\mathbb{S}_k := \bigcap_{\substack{i=1, \dots, M \\ i \neq k}} \mathbb{S}_{k,i}, \quad (\text{A.16})$$

which is the intersection of hyperplanes and, therefore, a simply connected convex open set (since its complement in \mathbb{S} is closed). Note that some of \mathbb{S}_k 's may be empty, implying that m_k is not a BN for any product-mix. Since at least two job-types are conflicting, there are at least two non-empty \mathbb{S}_k 's. It is possible to show (by contradiction) that \mathbb{S}_k 's are non-intersecting. Thus, each machine may be the BN in at most one simply connected subset of \mathbb{S} .

The proof for the case of zero buffers is identical to the above, using smc_k instead of tp_k . ■