

Analysis of assembly systems for interdeparture time variability and throughput

IHSAN SABUNCUOGLU¹, ERDAL EREL² and A. GURHAN KOK³

¹*Department of Industrial Engineering, and* ²*Faculty of Business Administration, Bilkent University, Ankara, Turkey 06533*
E-mail: Sabun@bilkent.edu.tr

³*Operations and Information Management Department, University of Pennsylvania, Philadelphia PA 19104, USA*

Received January 1999 and accepted April 2001

This paper studies the effect of the number of component stations (parallelism), work transfer, processing time distributions, buffers and buffer allocation schemes on throughput and interdeparture time variability of assembly systems. As an alternative to work transfer, variability transfer is introduced and its effectiveness is assessed. Previous research has indicated that the optimal throughput displays an anomaly at certain processing time distributions and, this phenomenon is now thoroughly analyzed and the underlying details are uncovered. This study also yields several new findings that convey important practical implications.

1. Introduction

In this paper, we consider the problem of designing an assembly system in which parts produced at two or more component stations are fed into an assembly station. In general, assembly systems are built from three main building blocks: (i) serial; (ii) merging; and (iii) splitting (or competing) configurations. Most of the existing work to date has been conducted on serial systems even though the merging and splitting configurations are commonly encountered in practice. In our study, however, we concentrate on the merging configuration and examine its various design characteristics.

The system under consideration is an assembly system containing two or more component stations and an assembly station with unlimited raw materials and finished goods inventory capacities. In the unbuffered system, there is no storage space between the component stations and assembly stations (Fig. 1). On the other hand, in a buffered system, there is storage space downstream of the component stations in which WIP inventory can be temporarily stored. In the unbuffered case, the assembly station does not start processing unless all items are transferred to the assembly station from the component stations (i.e., the assembly station does not start processing until a part from each feeder station becomes available). If a component station completes work early, it is blocked. On the other hand, the assembly station starves if one or more items are not ready from the component stations. Consequently, buffers reduce the incidence of blockage and starvation. Even though

the main emphasis of this paper is on an unbuffered configuration, we test the sensitivity of the results to the buffered case. Since the receiving area upstream of the component stations has an unlimited supply of raw materials and the shipping area downstream of the assembly station has an unlimited finished goods buffer capacity, the component stations never starve and the assembly station never becomes blocked. In common with the previous studies reported in the literature, all component stations are assumed to be identical and independent of each other.

In this study, we consider a system in which all the stations are perfectly reliable and have a 100% yield rate (i.e., no breakdowns, scraps, waste and absenteeism). We also assume that the system operates under a push mode. In other words, the component stations continuously process raw materials as long as they are not blocked and the assembly station operates as long as it is not starved. In contrast, demand triggers the production process in systems operating under the pull mode.

The system described above can be viewed as a simple system. However a thorough analysis of such a system can provide important insights into the better understanding of more complex real life systems. Also, this simple merging configuration constitutes a major building block of assembly systems encountered in practice. As previously noted by Powell and Pyke (1998), these systems are also common in several mixed model automated processing and manually operated assembly systems (e.g., ready-to-wear apparel production).

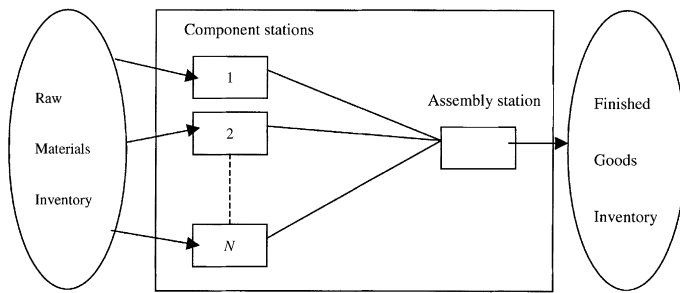


Fig. 1. Schematic view of an assembly system.

A key criterion in the design and operation of assembly systems has been the throughput which is measured as the number of units produced per unit time. Output variability (or interdeparture time variability) is also important especially in today's highly dynamic and stochastic environments, since a highly variable input or output process makes planning difficult and causes the performance to deteriorate significantly. For example, Hendricks (1992) and Hendricks and McClain (1993) study the output processes of serial production lines by observing the first and second moments of the interdeparture time distribution. Hence, practitioners designing such assembly systems should consider interdeparture time variability in addition to throughput. Here, interdeparture time variability is defined as the standard deviation of the time between consecutive departures of finished products from the system.

In this study, we consider both throughput and interdeparture time variability and analyze the effects of design factors such as parallelism (given by the number of component stations), processing time distributions, work and variability transfers from the assembly station to component stations, buffers and buffer allocation schemes. We also study the so called "intrinsic behavior of the optimal throughput for some processing time distributions" (Baker *et al.*, 1993; Rekhi *et al.*, 1995) and uncover the underlying details.

The rest of the paper is organized as follows. In Section 2, we summarize the relevant literature on assembly system design. In Section 3, we present the proposed approach, system considerations, and experimental design. The results of simulation experiments are presented in Section 4. In Section 5, we explain the intrinsic behavior. Later, we extend our analysis to the buffered case in Section 6. The paper ends with concluding remarks and suggestions for further research in Section 7.

2. Literature survey

There are only a few limited studies on this problem. They are briefly summarized in chronological order below.

Baker *et al.* (1990) examine the design of balanced assembly systems with variable processing times under two

loading mechanisms. The term "balanced" refers to identically distributed component and assembly station processing times. For all processing time distributions, the authors observe that the push mode results in a higher throughput than the pull mode since an assembly system utilizes the virtual buffer that exists in the assembly station under the push mode. The authors also analyze the effects of buffers on the throughput of assembly systems containing two feeder lines where each feeder is a serial line. They note that the results from the serial line research generally apply to such systems. In addition, the authors observe that a small buffer is sufficient to recover the significant portion of the lost capacity and that equal buffer allocation is desirable for these systems.

Later, Baker *et al.* (1993) examined the problem of allocating a fixed amount of work to stations in an assembly system operating under the push mode. For systems of up to four component stations with exponentially distributed processing times, Markov analysis was used, whereas simulation was used for other distributions and larger systems. Their basic finding is that the throughput can be improved by transferring work from the assembly station to component stations. The authors also study two specific unbuffered assembly systems for exponential and uniform processing time distributions. The first one is a system with two feeder lines where each feeder line is composed of two stations. Their results indicate that throughput is maximized by allocating more work to the initial stations and less work to the final stations of the feeder lines and assembly station. The second system involves two or more component stations in parallel. The results show that optimal throughput is a decreasing function of the number of the component stations. However, this phenomenon is not observed for the uniform distribution; instead, optimal throughput displays a small peak in a specific range. We call this anomaly "hump behavior" in this paper. In a later study (but published earlier), Baker (1992) presents a brief survey on serial lines and assembly systems. He notes that the above unexpected behavior might be due to the lower coefficient of variation (CV) of the uniform distribution.

Bhatnagar and Chandra (1994) examine the impact of different types of variability (processing time, unreliable stations and imperfect yield) on the throughput of assembly and competing systems. They find out that the CV of the processing times is a critical criterion for studying the impact of processing time variability in assembly systems.

Rekhi *et al.* (1995) studied assembly systems with up to 100 component stations for exponential, uniform, gamma and normal processing time distributions with different CV 's. The results indicate that as the number of component stations increases, the optimal throughput steadily deteriorates for the distributions that have small tails such as exponential and high- CV gamma distributions. However, they observe a hump behavior for those

distributions that have long tails such as normal or low- CV gamma distributions. The authors explain the hump behavior as being a result of the long tails of the above distributions. In this study, we further examine this behavior and uncover the underlying details.

In another study, Baker and Powell (1995) present a predictive model for the throughput of an assembly system with two component stations. They develop a distribution-free method to evaluate alternative system designs and claim that the algorithm performs well in terms of the accuracy of the predictions. Simon and Hopp (1995) analyze an assembly system with two component stations. There are finite buffers between each component and the assembly station. The authors develop a stochastic model to estimate the steady-state average throughput and inventory level performances.

Finally, Powell and Pyke (1998) analyze optimal buffer allocations for unbalanced assembly systems with two or three component stations. The authors suggest methods to find the optimal location for the first buffer. The results indicate that the optimal buffer location depends not only on the mean processing times, but also on the variability of the processing times.

3. Proposed study

As can be noted from the literature review, there are only a few studies that analyze the merging assembly configuration. Furthermore, the problem is only studied for the throughput measure by considering some design factors. For this reason, there is a need for a thorough analysis that address the following research issues:

1. An analysis of the interactions between various factors such as the number of component stations, processing time variability, work or variability transfer between stations for throughput.
2. An explanation of the hump behavior.
3. Repeating the analysis in 1 and 2 for interdeparture time variability.
4. Testing the sensitivity of the result to the buffered case.

We analyze these research issues in the following sections.

3.1. Experimental settings

We use four experimental factors with their levels given in Table 1: system size, transfer type, processing time distribution, and variability. System size, which determines the complexity of the system, is defined by the number of component stations. The previous studies on assembly systems indicate that the system performance is strongly affected by the system size. As seen in Table 1, we have 11 levels of N , ranging from the most basic system ($N = 2$) to a very large system ($N = 100$).

Table 1. Experimental factors and levels

| <i>Factors</i> | <i>Levels</i> |
|--------------------------------------|--|
| Number of component stations (N) | 2, 3, 4, 5, 10, 15, 20, 30, 40, 50, 100 |
| Transfer | Work Transfer (WT) with fixed PV Work Transfer (WT) with fixed CV Variability Transfer (VT) |
| Processing time distributions | Exponential, Lognormal, Uniform, Normal, Gamma Exponential (mean = 1, $CV = 1$) Lognormal (mean = 1, $CV = 0.3$) Lognormal (mean = 1, $CV = 0.6$) Uniform (mean = 1, $CV = 0.289$) Uniform (mean = 1, $CV = 0.5774$) |
| Processing time variability | Gamma (mean = 1, $CV = 0.289$) Gamma (mean = 1, $CV = 0.5774$) Normal (mean = 1, $CV = 0.289$) Normal (mean = 1, $CV = 0.4331$) Normal (mean = 1, $CV = 0.5774$) Normal (mean = 1, $CV = 0.8331$) |

In a merging configuration with an equal amount of work at each station, the assembly station acts as a bottleneck since it is dependent on component stations. In such a system, favoring the bottleneck station in order to improve the system performance can be achieved either by transferring the work and/or variability of the bottleneck station (assembly station) to other non-bottleneck stations (component stations) or imposing planned overtime work on the assembly station. Here, work transfer (WT) refers to shifting some work from the assembly station to the component stations by keeping the total work content of the system constant. Specifically, some portion of the mean processing time of the assembly station is transferred to the component stations. For example, consider a two-component station (i.e., one assembly station and two feeders) system with one unit of mean processing time and one unit of variance on each station. As depicted in Fig. 2a, the assembly station acts as a bottleneck in the system. One possibility to relieve the bottleneck station is to transfer some work, say 0.2 units of work, from the assembly to the component stations that results in 1.1, 1.1, and 0.8 units of work (mean processing time) in the component stations and assembly station, respectively.

Note that the work transfer can be achieved by either keeping the coefficient of variation (CV) or processing time variability (PV) constant (Fig. 2). In practice, the CV -fixed case arises when the variability is mainly caused by internal factors such as, worker qualification, part and process characteristics, whereas the variability in the PV -fixed case is mainly caused by external factors such as the nature of work environment, level of automation, etc.

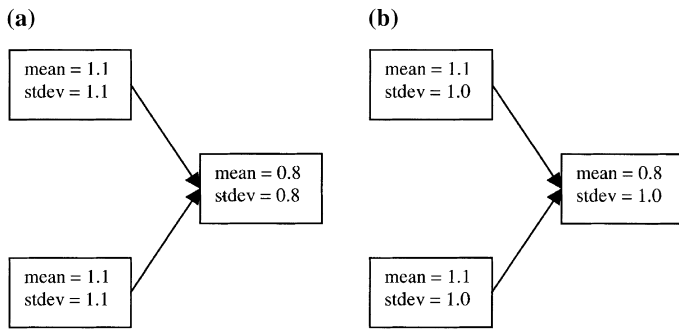


Fig. 2. The work transfer alternatives: (a) *WT* with a fixed *CV*; and (b) *WT* with a fixed *PV*.

An alternative way of relieving the bottleneck station is to transfer some portion of variability of the assembly station to component stations. This is called variability transfer (*VT*) and it shifts the variance of the assembly station to the component stations by keeping the sum of the variance of the processing times constant. In practice, the variability transfer can be achieved by allocating more experienced/trained workers to the assembly station.

We use the following three-element notation (*i/ii/iii*) to describe the main characteristics of the experimental setting.

- i. Type of transfer: Work transfer (*WT*) or variability transfer (*VT*).
- ii. Coefficient of variation (*CV*) or processing time variability (*PV*) is kept constant during the transfer.
- iii. Objective function: maximize throughput (*T*) or minimize interdeparture time variability (*IDTV*).

The factors in the first two entries have been discussed previously. Note that the second entry is left blank for variability transfer since neither *CV* nor *PV* is kept constant. The third entry in this notation is reserved for the performance measures (*T* and *IDTV*). If this entry is left blank, it indicates that an optimization is conducted for either *T* or *IDTV* by generating the entire search space (i.e., complete enumeration) and selecting the best performer. For example, the notation *WT/CV/-* of Sections 4.1.1 to 4.1.4 represents the case in which an optimal *T* is investigated as work is transferred from the assembly station to the component stations. Here, the optimal *T* is obtained by searching all possible ways of work transfer that yield the maximum *T* for a given number of component stations (*N*). As a result, Fig. 14a,b (optimal *T* versus *N*) is obtained by plotting the best *T* values extracted from Fig. 4a–c (*T* versus *WT*).

In this study, we consider five different distribution functions for the processing times. A list of these distributions and their variability levels are given in Table 1. These distribution functions and their parameters are selected from the literature based on previous theoretical and empirical studies (Knott and Sury, 1987; Baker *et al.*, 1990). We mostly used the lognormal distribution in our

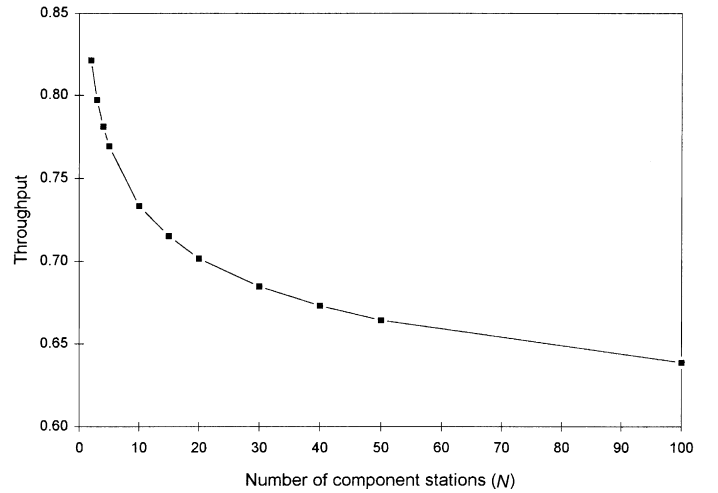


Fig. 3. The effect of parallelism (*N*) on throughput in the balanced case for lognormal times with *CV* = 0.289.

experiments because of its flexibility and positive skewness. It is also one of the most suggested distributions in the literature. The exponential distribution is used to see the effect of a high variance case. The other distributions (gamma, normal, and uniform) are employed to investigate the hump behavior. Note that we use the same distribution function for both the assembly and component stations. In the simulation experiments, processing times are generated independently for each component station from the same distribution function (i.e., random variates are generated at each component station whenever needed). At the assembly station, we use the same type of distribution function with different parameters due to work and variability transfer.

3.2. Simulation model

The simulation model is developed using SIMAN. The Welch approach (Law and Kelton, 1991) is used to determine the warm-up period (300 observations). Relevant statistics are collected in steady-state conditions using the replication/deletion method. Specifically, 20 replications each with a length of 2000 observations are taken with an initial condition of an empty state (i.e., the data analysis is performed for 40 000 observations in steady-state conditions). A total of 20 simulation replications are taken to achieve a satisfactory level (at most 1%) of the relative accuracy (half width of the confidence interval divided by the estimate of the quantity of interest). Specifically, we used the confidence interval approach for the paired-t test to assess the significance of the differences between results obtained at various experimental points. In our study, *IDTV* is measured by using the average of the sample standard deviations of 20 replications. Even though this is not an unbiased estimator of the popula-

tion standard deviation, it is a practical measure and produces a reliable estimate for the variability.

4. Computational results

In this section, we present the results of our simulation experiments. First, we examine the effects of the experimental factors (i.e., parallelism, work transfer, variability of processing times) on throughput. This is followed by the analysis for the interdeparture time variability (*IDTV*). We then compare work and variability transfers in terms of throughput and *IDTV*. We should also point out here that the results presented below are tested for statistical significance at $\alpha = 0.05$.

4.1. Results on throughput

The results on throughput (T) are itemized below. A few of these results are confirmations of previous findings; we present them in order to give a complete picture and validate our model.

4.1.1. Effect of parallelism (N)

Throughput of a balanced system (with the mean processing times set to one) decreases at a decreasing rate (Fig. 3) as N increases. This is due to the fact that, as parallelism increases, coupling in the system increases, synchronization of the components becomes more difficult and thus T decreases. This result is previously reported in the literature (Baker *et al.*, 1993). Note that the rate of decrease slows down as N increases, since the effect of adding one more station to a larger system (e.g., $N = 50$) is less significant than a smaller system (e.g., $N = 2$).

The above result has important practical implications since the decreasing behavior of T reinforces the *design for assembly* principle: designing products with the minimum number of component parts.

4.1.2. Effect of WT/CV -

When a certain amount of work is transferred from the assembly to the component stations by keeping CV constant (i.e., the variability of the processing times is proportional to the mean), T increases up to a level and then decreases, resembling a concave function (Fig. 4(a-c)). Note that the decreasing behavior is not observed for large N since the amount of work in the assembly station is not sufficient to raise the mean processing time of the component stations. Prior to work transfer, the assembly station acts as the bottleneck station. Hence, work transfer favors the bottleneck station and leads to an improvement in T . However, after a certain amount of work transfer, the component stations become a bottleneck and consequently any further work transfer leads to a decrease in T .

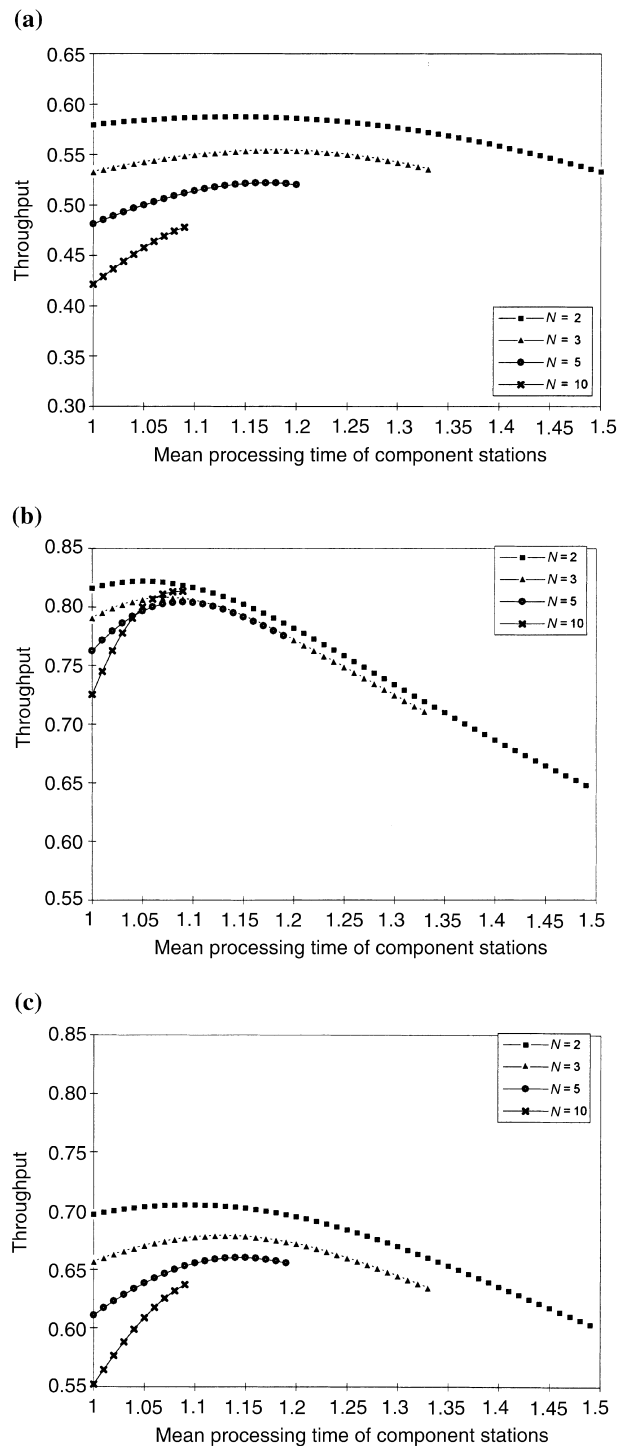


Fig. 4. The effect of WT/CV - on throughput for: (a) exponential processing times; (b) lognormal processing times with $CV = 0.3$; and (c) lognormal processing times with $CV = 0.6$.

Note also that the maxima of these curves correspond to optimal T values associated with N .

The practical implication of the above result is that a certain amount of work transfer from the assembly to the component stations helps to raise the throughput level without employing any additional resources (e.g., buffers).

4.1.3. Interaction effect of $WT|CV|$ - and parallelism

The effect of work transfer on T is pronounced for large N , since there is more improvement potential in such systems (see Fig. 4(a-c)). This means that the capacity lost due to coupling in the system can be regained by work transfer to a greater extent in large systems than in small systems.

4.1.4. Interaction effect of $WT|CV|$ - and CV

As depicted in Fig. 4(b and c), the rate of improvement in T is smaller in the high CV case (e.g., for $N = 10$, the improvement after 0.02 units of work transfer is 5.11% for a low CV while it is 4.44% in the high CV case). This is because, in the high CV case, the positive effect of work transfer becomes smaller due to increased in the system. In other words, to obtain the same percentage of improvement, one needs to make more work transfer in the high CV case than the low CV case. When the simulation experiments are repeated at very high CV (e.g., $CV = 1.5$), we observed the same behavior.

4.1.5. Effects of $WT|PV|$ - and interactions with parallelism and PV

As discussed earlier in the paper, the second type of work transfer is accomplished by keeping PV constant. This means practically that more experienced workers and/or faster machines are shifted to the component stations. When this situation is analyzed, we observed the same concave behavior of T which has been already elaborated in Section 4.1.2.

Note that this concave behavior can only be shown for some ranges of the mean processing time of lognormal distribution (Fig. 5(a and b)). This was mainly due to the

difficulties encountered in generating lognormal random variates for small mean values and very large variances.

4.1.6. Effect of $VT|$ -/-

As variability is transferred from the assembly station to the component stations (i.e., the standard deviations of the processing times of the component stations increase), T increases (Fig. 6 (a and b)). However, the change in T is relatively less significant in this case than the PV or CV -constant work transfer cases. This result is analogous to the previous findings on serial production systems: the effect of processing time variability on T is much less than the mean of processing time (Lau, 1992; Erel *et al.*, 1997). Again, the same behavior is observed in the very high CV case.

4.1.7. Interactions of $VT|$ -/- with parallelism and PV

When Fig. 6(a and b) is carefully analyzed, we observe a tendency that the increase in T is greater in large systems (i.e., more component stations), even though the improvement is not statistically significant. As discussed in Section 4.1.3. this result is due to the existence of more improvement potential in such systems.

4.2. Results on interdeparture time variability

We present the results on $IDTV$ in the same order as throughput.

4.2.1. Effect of parallelism (N)

As N increases, $IDTV$ increases at a decreasing rate (Fig. 7) due to more coupling in such systems. Similar to T , the effect is less significant in the high parallelism case. This suggests that systems with fewer component stations

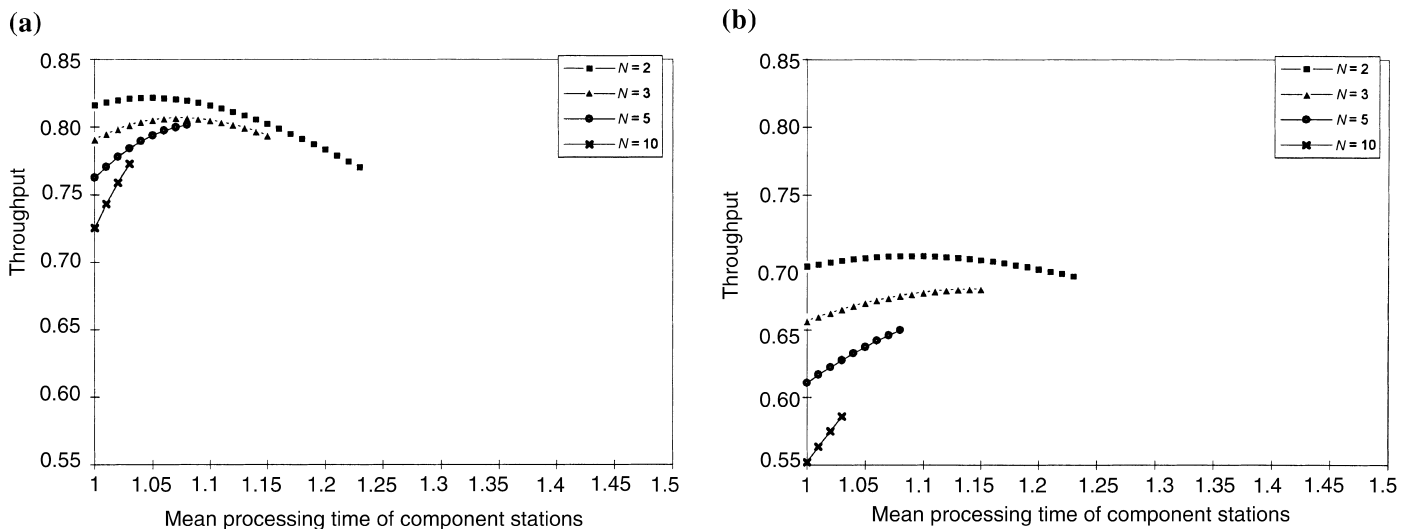


Fig. 5. The effect of $WT|PV|$ - on throughput for lognormal processing times when: (a) $PV = 0.3$; and (b) $PV = 0.6$.

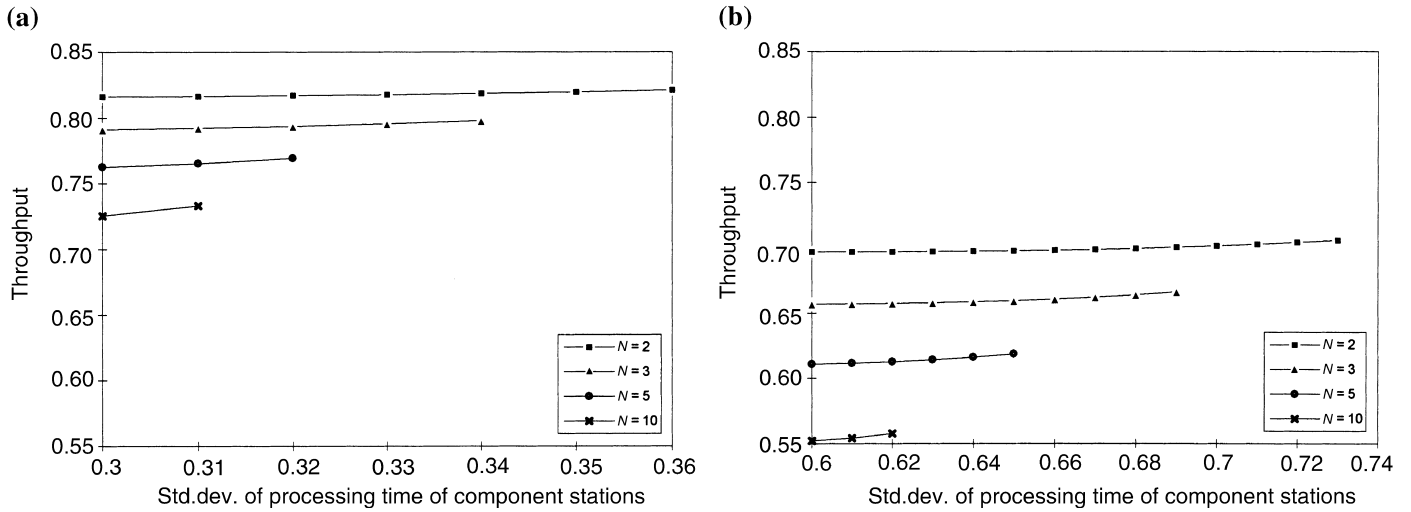


Fig. 6. The effect of $VT/-$ on throughput for lognormal processing times when: (a) $PV = 0.3$; and (b) $PV = 0.6$.

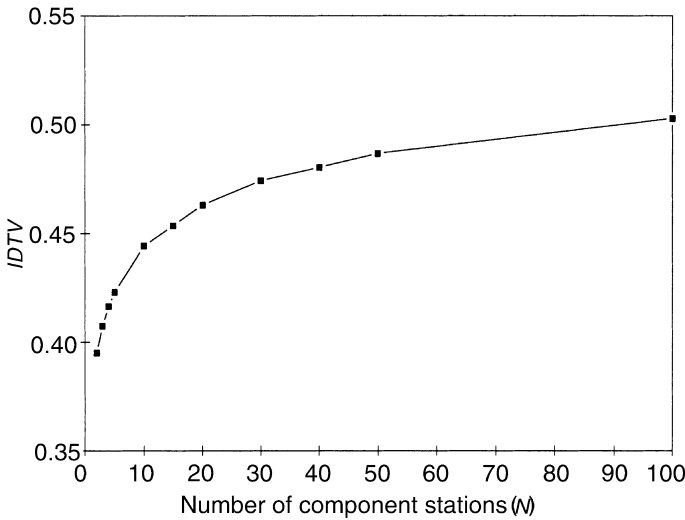


Fig. 7. The effect of parallelism (N) on interdeparture time variability in the balanced case for lognormal processing times with $CV = 0.289$.

are more advantageous for $IDTV$ from the system designer perspective.

4.2.2. Effect of $WT|CV/-$

We identified two cases each of which is explained below. The first represents the simple situation whereas the second is a relatively complicated case that requires further analysis.

4.2.2.1. *Exponential distribution case.* Similar to the throughput case (Section 4.1.2), the performance of the system improves by work transfer up to a certain level and then starts to deteriorate (Fig. 8a). However

differently to the previous case, the, $IDTV$ behavior is a convex function of the work transfer. Similar observations are made in the hyper-exponential case.

4.2.2.2. *Lognormal and uniform distributions cases.* Unlike the exponential case, the behavior of the $IDTV$ as a function of work transfer is a complex function. In this case, as seen in Fig. 8(b–f), the function is not unimodal with no apparent explanation. Hence, we perform additional experiments to understand this interesting behavior. Specifically, we measure the $IDTV$ in terms of its components: the variability of interarrival times to the assembly station (called the $S_{Component}$) and the variability of the assembly station (called the $S_{Assembly}$). Figure 9(a–d) illustrates the behavior of these two components. Since CV is kept constant during the work transfer, $S_{Assembly}$ decreases linearly with work transfer where the slope is higher in the high parallelism and high CV cases. On the contrary, the $S_{Component}$ increases with work transfer because the PV of component stations increases. As a result of the combined effect of these two components, $IDTV$ decreases as a function of work transfer in the high parallelism and high CV cases, but it increases in the low parallelism and low CV cases. Figure 8(e and f) also supports the above observations for uniform distribution. We also note that the results of a lognormal distribution with $CV = 1$ (Fig. 8d) are very similar to the exponential distribution case (Fig. 8a). This implies that the behavior of the $IDTV$ approaches the convex behavior of the simple case.

The above discussion suggests to practitioners that one should be cautious when implementing work transfer in the low CV and low parallelism cases since the complicated behavior of $IDTV$ can conflict with T .

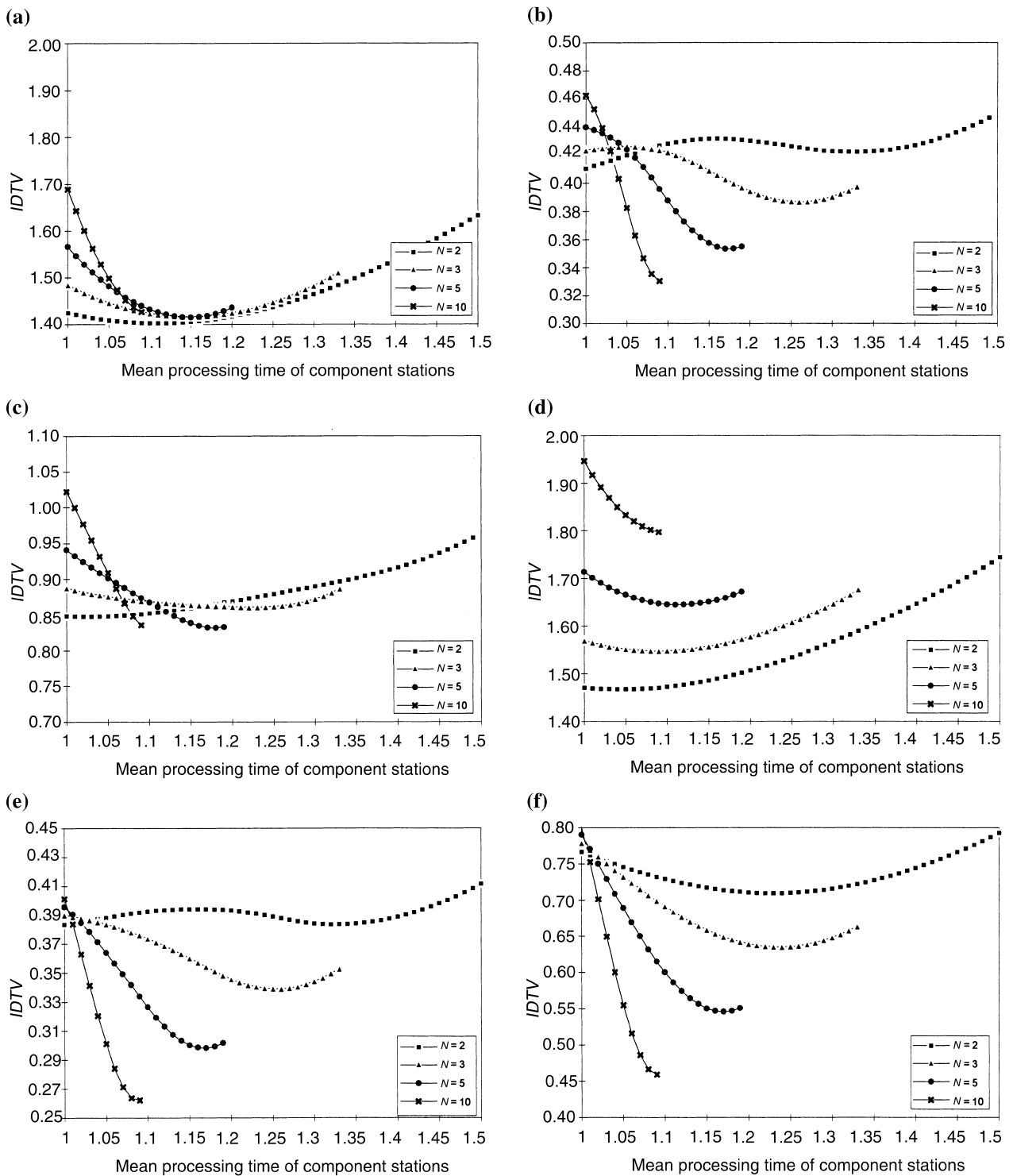


Fig. 8. The effect of $WT/CV/-$ on interdeparture time variability for: (a) exponential processing times; (b) lognormal processing times with $CV = 0.3$; (c) lognormal processing times with $CV = 0.6$; (d) lognormal processing times with $CV = 1.0$; (e) uniform processing times with $CV = 0.289$; and (f) uniform processing times with $CV = 0.5774$.

4.2.3. Effect of $WT/PV/-$

As seen in Fig 10(a and b), $IDTV$ first increases and then stabilizes at some level as work is transferred from the

assembly station to the component stations. This is probably due to the increase in mean processing times of the component stations. Note that the curves could not be

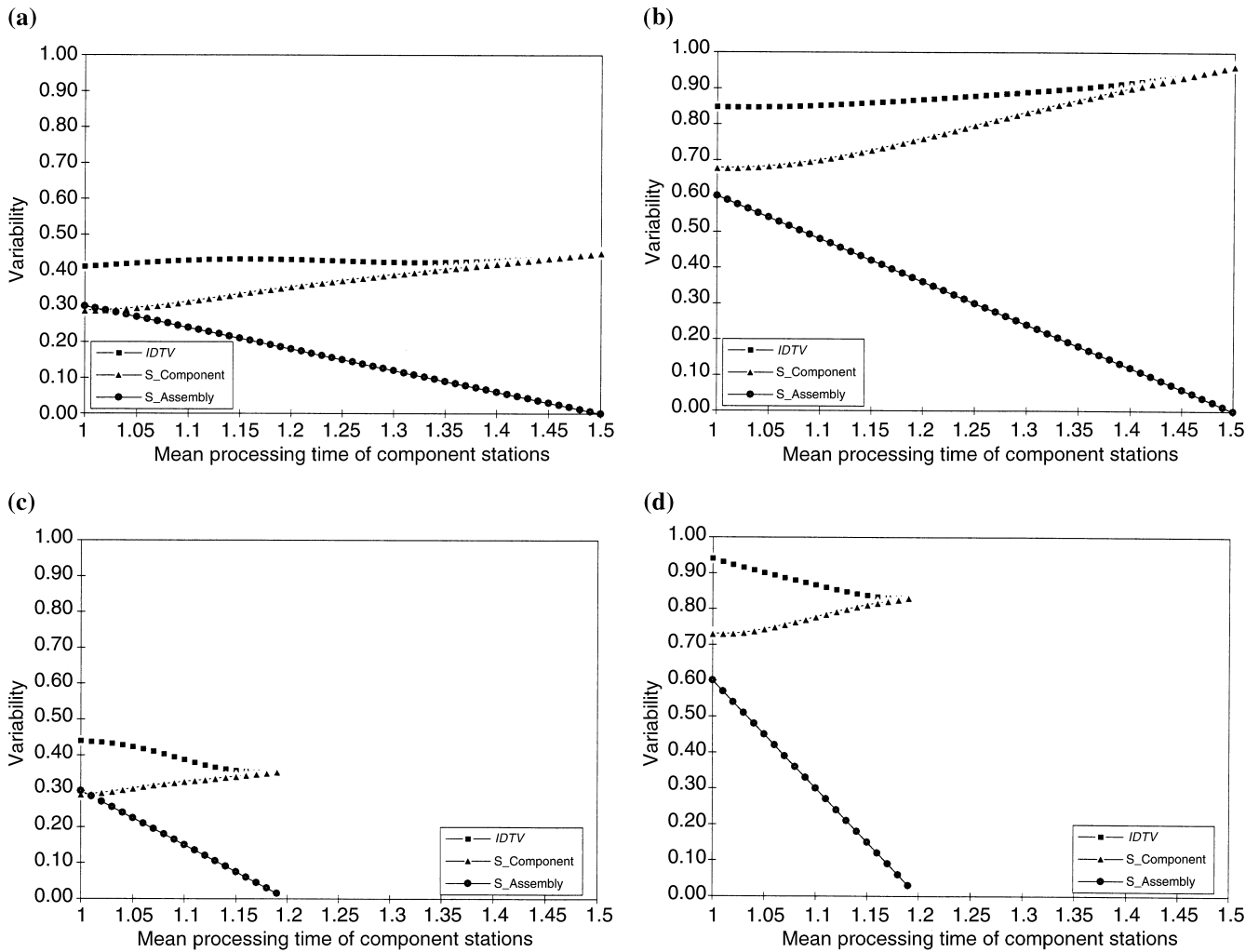


Fig. 9. The effect of WT/CV on interdeparture time variability and its components for lognormal processing times when: (a) $N = 2, CV = 0.3$; (b) $N = 2, CV = 0.6$; (c) $N = 5, CV = 0.3$; and (d) $N = 5, CV = 0.6$.

completed because of the difficulty encountered in generating lognormal random variates for small mean values and very large variances.

4.2.4. Effect of VT

$IDTV$ decreases as more variability is transferred to the component stations (Fig. 11(a and b)). This is because the assembly station, that acts as the bottleneck station, is favored by variability transfer and consequently the $IDTV$ decreases. In practice, this suggests that one should assign less variable operators, machines or other resources to the assembly station.

4.2.5. Interaction effect of VT and parallelism

As seen in Fig. 11(a and b), the rate of improvement in $IDTV$, as variability is transferred from the assembly to the component stations, is more significant in the high parallelism case than in the low parallelism case due

to more variability transfer possibilities in large systems.

4.2.6. Interaction effect of VT and PV

The percentage improvement in $IDTV$, as variability is transferred from assembly to component stations, is less in the high PV case (Fig. 11(a and b)). As a result, the potential benefits of variability transfer are not realized to their full extent due to the highly variable environment.

As a closing remark in this section, we can conclude that work and variability transfers are effective tools to improve system performance both in terms of T and $IDTV$, especially in the high parallelism and low CV cases.

4.3. Analysis of the correlation of the output process

We also analyze the correlation structure of the output process. Specifically, we consider the effects of correlation

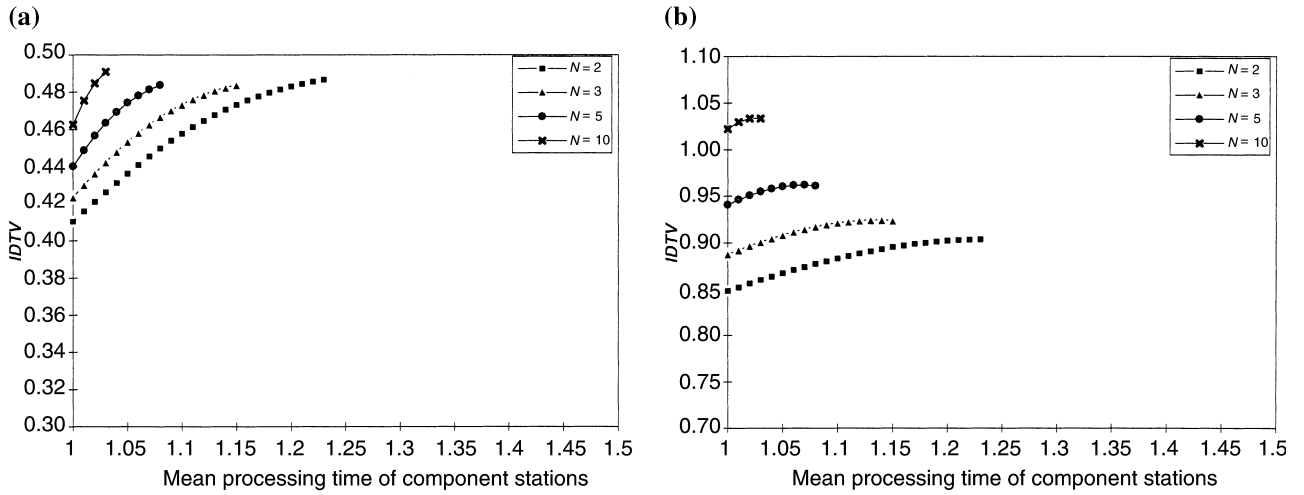


Fig. 10. The effect of WT/PV on interdeparture time variability for lognormal processing times when: (a) $PV = 0.3$; and (b) $PV = 0.6$.

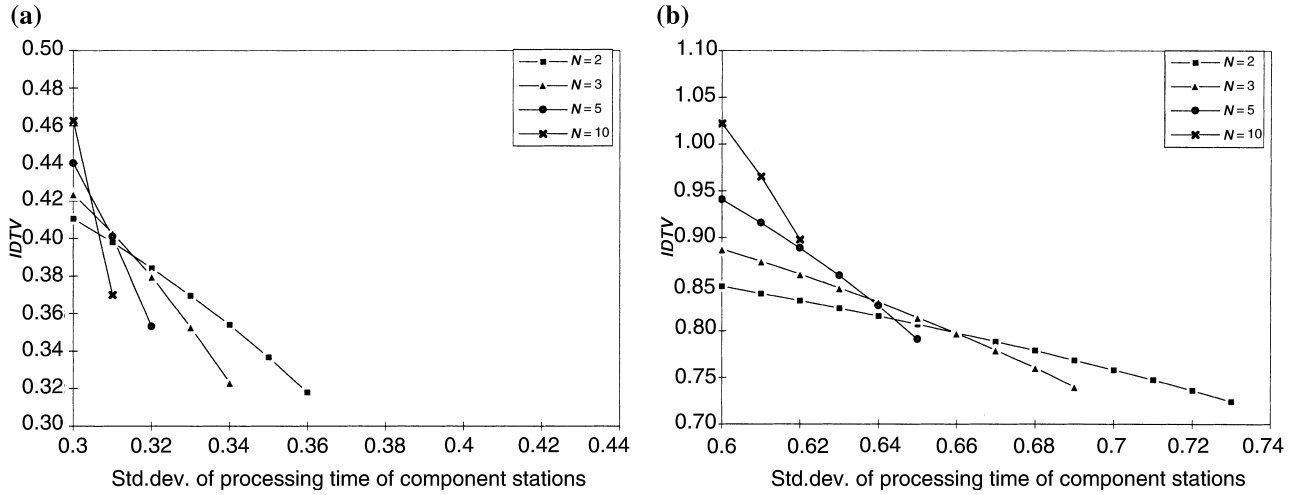


Fig. 11. The effect of VT on interdeparture time variability for lognormal processing times when: (a) $PV = 0.3$; and (b) $PV = 0.6$.

on $IDTV$ for system size (N), work transfer (WT) and variability transfer (VT). The objectives of these experimentations are: (i) to see if a non-zero correlation exists; and (ii) to test the validity of the previous conclusions in Section 4.2. We measure the variability of the output process as the asymptotic variance discussed in Hendricks (1992). In particular, we use the following variance estimate:

$$AVAR \approx \text{Var} \left(1 + 2 \sum_{j=1}^q \text{Corr}_j \right),$$

where the lag j correlations are zero for $j > q$ and q is finite.

First, we look at the effects of correlation on $IDTV$ for various system sizes. As seen in Fig. 12(a and b), $IDTV$ and $AVAR$ mostly display a similar behavior when CV is high ($CV = 0.6$). In the low CV case, however, $AVAR$

shows a steadily flat behavior as compared to $IDTV$. These observations can be explained as follows. The correlation is negative and decreases as N increases; this negative correlation results in $AVAR$ being lower than $IDTV$. In the high CV case (Fig. 12b), $AVAR$ shows an increasing behavior because the variance term is much larger than the sum of the negative correlations. But, in the low CV case, the variance and correlation terms balance each other and hence the flat behavior appears in Fig. 12a.

Next, we test if the previous conclusions on work and variability transfers hold in the correlation case. For this, we repeat the simulation experiments for $AVAR$ in both high and low CV cases for various N values. The results are summarized in Fig. 13(a and b).

When Fig. 13a is compared with Fig. 8(b and c), and Fig. 13b is compared with Fig. 11(a and b), we observe that both $AVAR$ and $IDTV$ display a similar behavior,

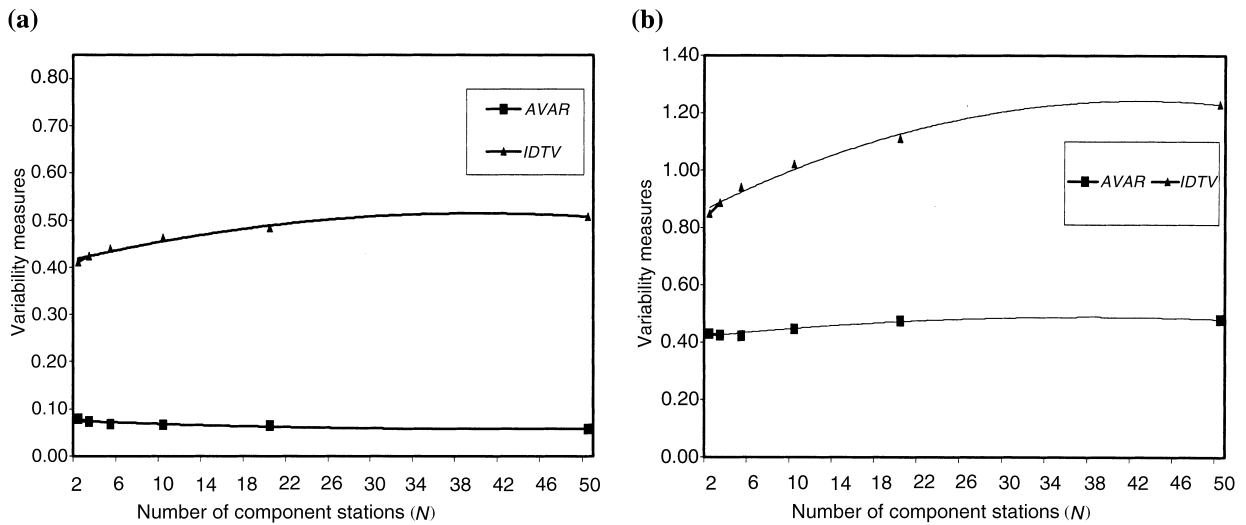


Fig. 12. The effect of N on $AVAR$ for the cases when: (a) $CV = 0.3$; and (b) $CV = 0.6$.

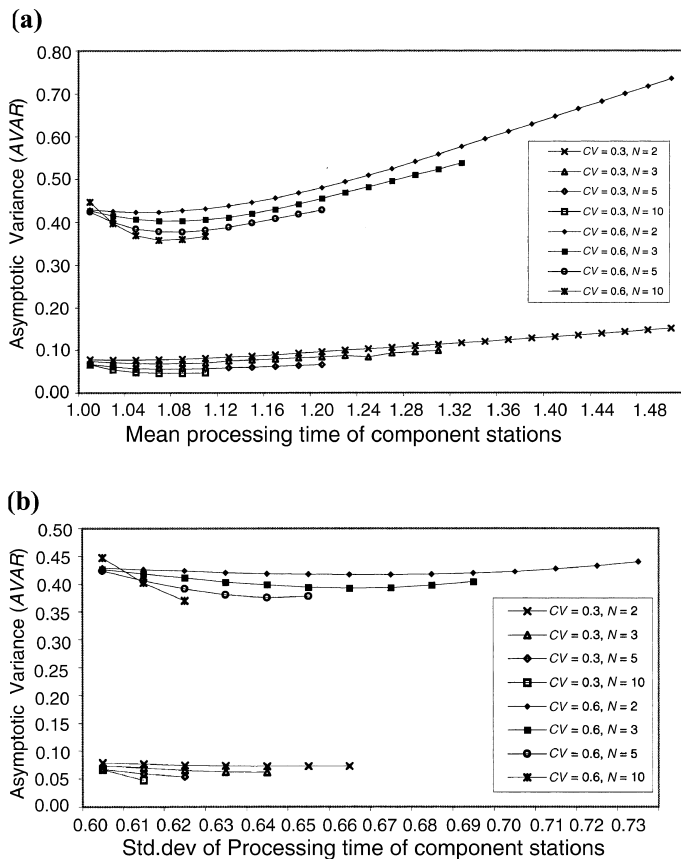


Fig. 13. (a) The effect of WT on $AVAR$; and (b) the effect of VT on $AVAR$.

but the curves are relatively flat in the $AVAR$ case. This is again due to the fact that the negative correlation term in $AVAR$ has a decreasing effect on variability. However this does not change our previous conclusions.

4.4. Optimal throughput and interdeparture time variability

Up to now, we have discussed the effects of several design factors on the system performance. In most cases, however, both practitioners and academics are more interested with the optimal ways to operate systems. Thus, in this section, we analyze the effects of optimal work and variability transfers on T and $IDTV$. To find the optima, we first generate the entire search space and select the maximum or minimum (depending on the objective function). The results are summarized below.

1. Optimal T decreases as the number of component stations increases for the exponential and lognormal distributions with a CV of 0.6. However, as discussed earlier, it exhibits a “hump behavior” for the lognormal distribution with a CV of 0.3 and the uniform distribution (see Fig. 14(a–e)). The negative effect of parallelism on optimal T is greater in the high CV case than in the low CV case. The hump behavior and its practical implications will be further discussed in Section 5.
2. Percentage improvement in T from the balanced case to optimal configuration (i.e., the optimal work transfer level) is greater in the high CV case and large systems (see the first two columns of Table 2).
3. Work transfer which optimizes T leads to a substantial improvement in $IDTV$ in large systems (i.e., $N = 5$ or 10). However, optimal configuration may cause $IDTV$ to deteriorate in small systems (see the first three columns of Table 2). As discussed in detail in Section 4.2.2.2 this is due to the fact that $IDTV$ sharply decreases for large N whereas it increases for small N (see Fig. 8(b and c)). This suggests that system designers should be more cautious when applying work transfer to small systems since the two objectives can conflict with each other.

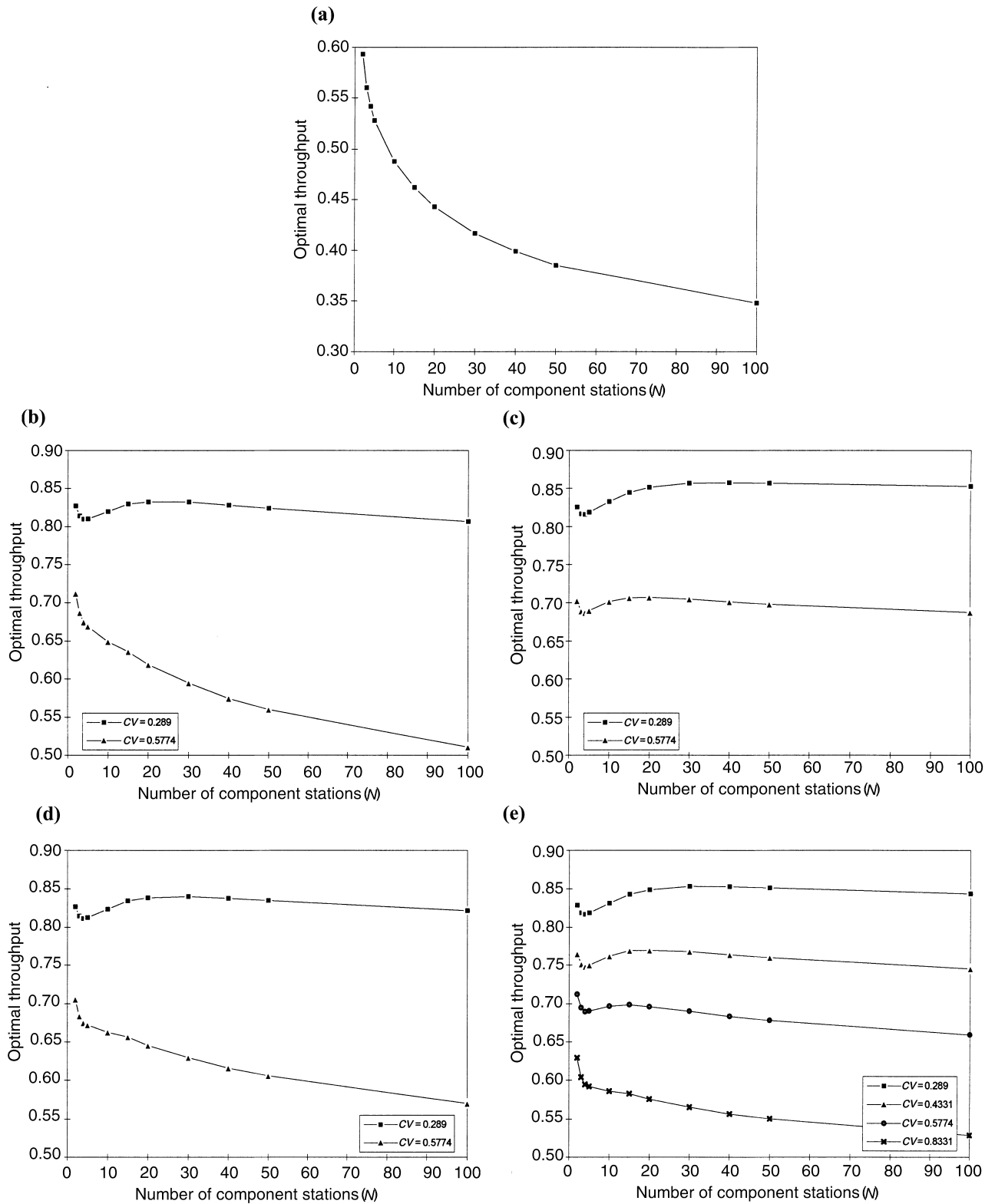


Fig. 14. The effect of $WT/CV/T$ on throughput for: (a) exponential processing times; (b) lognormal processing times; (c) uniform processing times; (d) gamma processing times; (e) normal processing times.

Table 2. Improvement from the balanced case in throughput and interdeparture time variability for lognormal processing times

| | $WT CV T$ | | $WT CV IDTV$ | | $VT - T \text{ or } IDTV$ | |
|--------------------------|----------------|----------|----------------|----------|---------------------------|----------|
| | Throughput (%) | IDTV (%) | Throughput (%) | IDTV (%) | Throughput (%) | IDTV (%) |
| Low CV ($CV = 0.3$) | | | | | | |
| $N = 2$ | 0.72 | -2.26 | 0.00 | 0.00 | 0.62 | 29.12 |
| $N = 3$ | 2.10 | -0.47 | -6.30 | 9.48 | 0.85 | 31.01 |
| $N = 5$ | 5.17 | 11.16 | 2.72 | 24.53 | 0.88 | 24.64 |
| $N = 10$ | 10.83 | 40.08 | 10.83 | 40.08 | 1.08 | 25.05 |
| High CV ($CV = 0.6$) | | | | | | |
| $N = 2$ | 1.13 | -0.34 | 0.61 | 0.07 | 1.25 | 17.08 |
| $N = 3$ | 3.25 | 2.53 | 1.65 | 3.11 | 1.38 | 19.93 |
| $N = 5$ | 7.42 | 11.65 | 7.02 | 13.17 | 1.24 | 18.95 |
| $N = 10$ | 13.30 | 22.37 | 13.30 | 22.37 | 0.97 | 13.87 |

- The above observation can be also made for $IDTV$; work transfer that optimizes $IDTV$ leads to an improvement in T in large systems. It is worthwhile to mention that in very large systems (i.e., $N = 10$ in our case), the optimal work transfer for T and $IDTV$ are the same.
- The results of experiments indicate that VT leads to same amount of improvement in both T and $IDTV$ (see Table 2 for the percentage improvements). It appears that the level of VT that optimizes T also optimizes $IDTV$ for any N (the converse is also true).
- We also observe that in small systems, variability transfer improves T and $IDTV$ more than work transfer, whereas the reverse is true for large systems. This suggests that work transfer for large systems and variability transfer for small systems should be the recommended policies to improve the overall performance of assembly systems.

5. Hump behavior

5.1. Throughput

As reported earlier in the paper, the optimal throughput as a function of the number of component stations displays a *hump behavior* for certain distribution functions (e.g., lognormal). This behavior was first observed by Baker *et al.* (1993) and discussed later by Rekhi *et al.* (1995) who relate this unexpected behavior to certain processing time distributions. In this section, we further examine this phenomenon and uncover the underlying details.

The hump behavior, which is depicted in Fig. 14(a–e) can be explained with Fig. 4(a–c). As can be seen in Fig. 4(b and c), the rate of improvement in T as a function of work transfer is greater in the low CV and high parallelism cases. Consequently, the curves associated with large N crossover the curves with small N in the low CV case (Fig. 4b). This means that the optimal T of large

N can be substantially greater than the one of small N , although T of large N is smaller than the one of small N in the balanced case (i.e., at the origin of Fig. 4b). Hence, the optimal throughput curve is not always a decreasing function of N , but rather displays a hump behavior by making ups and downs.

As seen in Fig. 14a, the hump behavior is not observed for the exponential distribution function. This was also reported in Baker *et al.* (1993) and Rekhi *et al.* (1995) without any detailed explanation: Baker *et al.* (1993) relate this phenomenon to an “intrinsic” property of the uniform distribution and leave it as an open research question. Rekhi *et al.* (1995) further examine this behavior in a larger experimental setting and conclude that this phenomenon is due to the long-tailed distribution functions. Our results depicted in Fig. 4a clearly show that the curves associated with different N do not cross-over each other as a function of work transfer. This is due to the fact that all work at the assembly station is depleted by the time the crossovers actually take place. Note that this also leads to unequal lengths of the curves in Fig. 4(a–c). Consequently, without crossovers, the optimal throughput curve is a decreasing function of N . As seen in Fig. 14(a–e) the above explanation is also valid for the other distributions with a high CV (note also that the hump does not disappear in Fig. 14c, since the CV of a uniform distribution could not be increased beyond a certain level; otherwise, negative values would have been sampled).

In summary, the hump behavior is a distribution-free phenomenon; that is, it can be observed for any distribution function as long as the CV can be varied. For the high CV case, it is not observed due to the reasons discussed above. However, as CV decreases, it starts to emerge as shown in Fig. 14(a–e).

5.2. Hump behavior of interdeparture time variability

Optimal $IDTV$ is an increasing function of the number of component stations for exponentially distributed

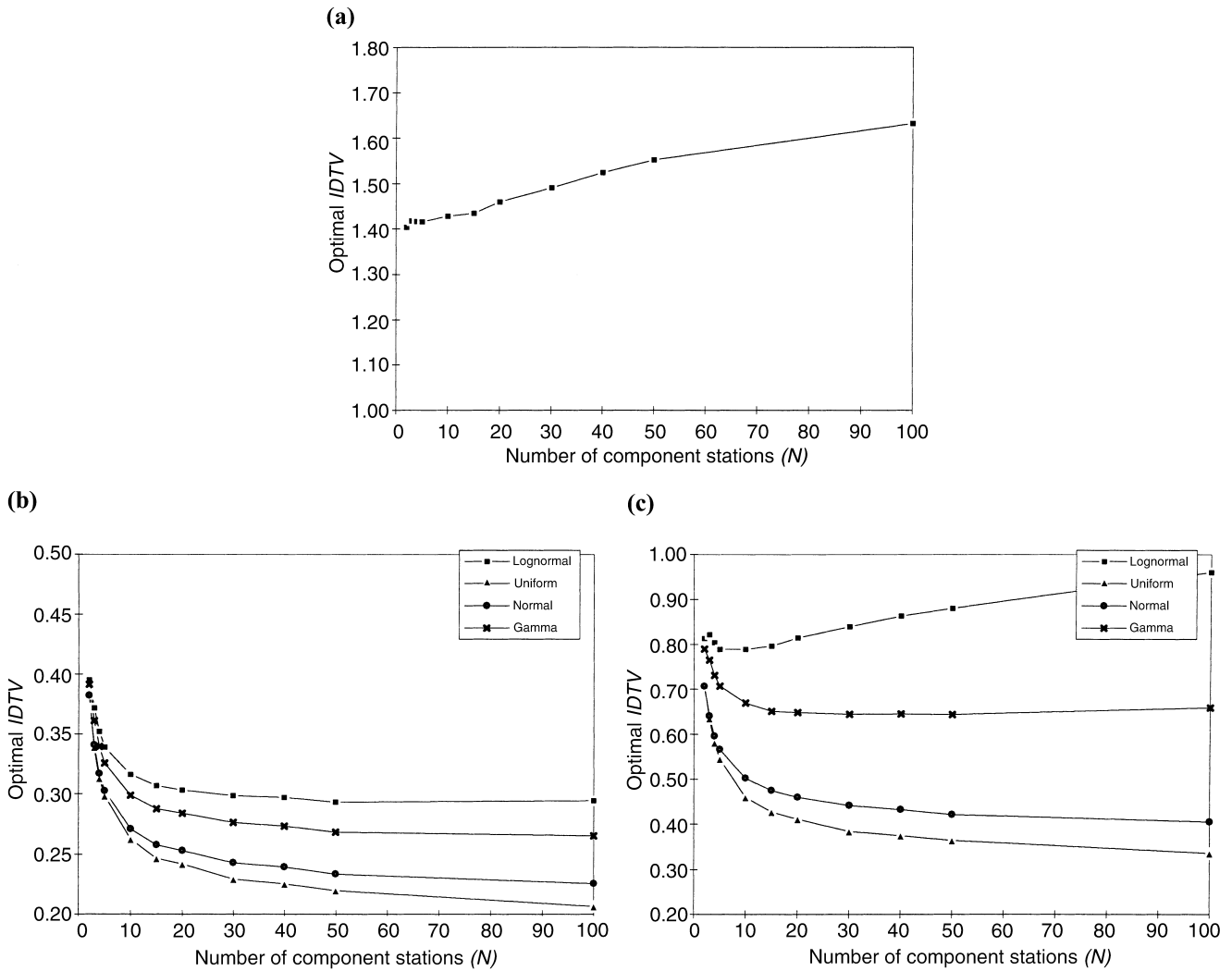


Fig. 15. The effect of $WT/CV/IDTV$ on interdeparture time variability for: (a) exponential processing times; (b) $CV = 0.289$; and (c) $CV = 0.6$.

processing times (Fig. 15a). However, for various other distributions it is a decreasing function of N at low CV (Fig. 15b). In the high CV case, the optimal $IDTV$ associated with lognormal distribution again displays a hump behavior (Fig. 15c). Similar to the throughput case discussed in Sections 4.1.2, 4.1.3 and 4.1.4, $IDTV$ decreases at a faster rate in the low CV and high parallelism cases. This leads to crossovers (see Fig. 8(b and f)) and the optimal $IDTV$ of a small system can be higher than that of a larger system. Note also that the hump behavior of $IDTV$ is just a mirror image of the hump behavior of the throughput.

6. Analysis of buffered systems

In this section, we extend our analysis to the buffered case. First, in Section 6.1. we compare two buffer allocation schemes, called the “pooled” and the “separated” buffer

configurations. Then we measure the sensitivity of our previous results to the buffered case in Section 6.2.

6.1. Buffer allocation schemes

Due to the structure of the system, two main types of buffer allocation are possible. In the first case (i.e., the separated type), each component station has its own dedicated buffer area whereas in the second case (i.e., the pooled type), all component stations have access to a common buffer area. Our initial expectation was that the pooled type would outperform the separated type as in the case of queuing systems. However, our results indicate that the pooled type is not always the best configuration for assembly systems. As can be seen in Fig. 16(a–d) the separated buffer type performs considerably better than the pooled type. This counter-intuitive result is explained as follows. Even though the pooled

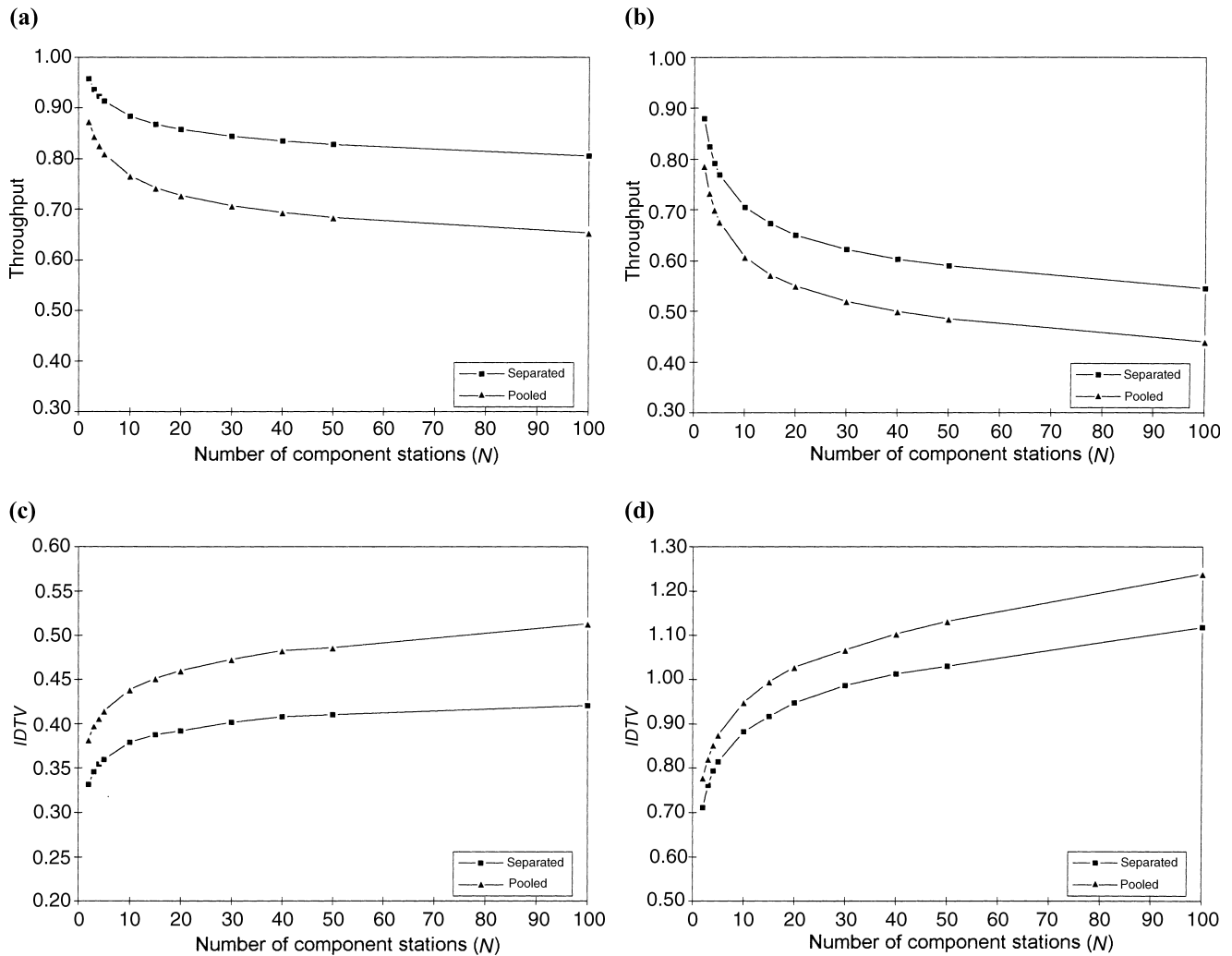


Fig. 16. The effect of parallelism (N) on throughput and interdeparture time variability in the balanced case for lognormal processing times when: (a) $CV = 0.3$; (b) $CV = 0.6$; (c) $CV = 0.3$; and (d) $CV = 0.6$.

type provides the component stations with more flexibility in using buffer spaces, it causes more blockage than the separated type due to some buffer spaces being occupied by the component stations. This affects the synchronization of the component stations and hence deteriorates the system performance. The above discussion holds for both T and $IDTV$. We also observed that as in the case of the unbuffered systems, the negative effect of N on T and $IDTV$ is pronounced as CV increases due to more coupling between the stations.

Note also that average WIP inventory is a linearly increasing function of N and the pooled type yields higher average WIP inventory than the separated type.

6.2. Work transfer and hump behavior in the buffered case

In contrast to the unbuffered case (Section 4), WT/CV does not improve T (in fact it even deteriorates the system performance) in small systems (e.g., $N = 2, 3, 5$) for the

separated buffer allocation type (Fig. 17a). However, it still has a positive effect on T for the pooled type (Fig. 17b). This observation can be explained by the fact discussed in Section 6.1. that the pooled type adversely affects the synchronization of the component stations and work transfer helps to alleviate this negative effect (Fig. 17b). In the separated case, there is not much coupling between the stations and work transfer can only lead to bottleneck component stations that eventually deteriorate T . In large systems, work transfer improves T for both buffer configurations.

The effect of WT/CV on $IDTV$ is similar to the unbuffered case for both the pooled and separated buffer configurations (see Fig. 17(c and d) in comparison to Fig. 8c). We also note that the average WIP inventory decreases with work transfer since it reduces the possibility of blockage of the component stations.

The hump behavior of the optimal T and $IDTV$ is smoothed out in the buffered case (Fig. 18(a–d)). Recall

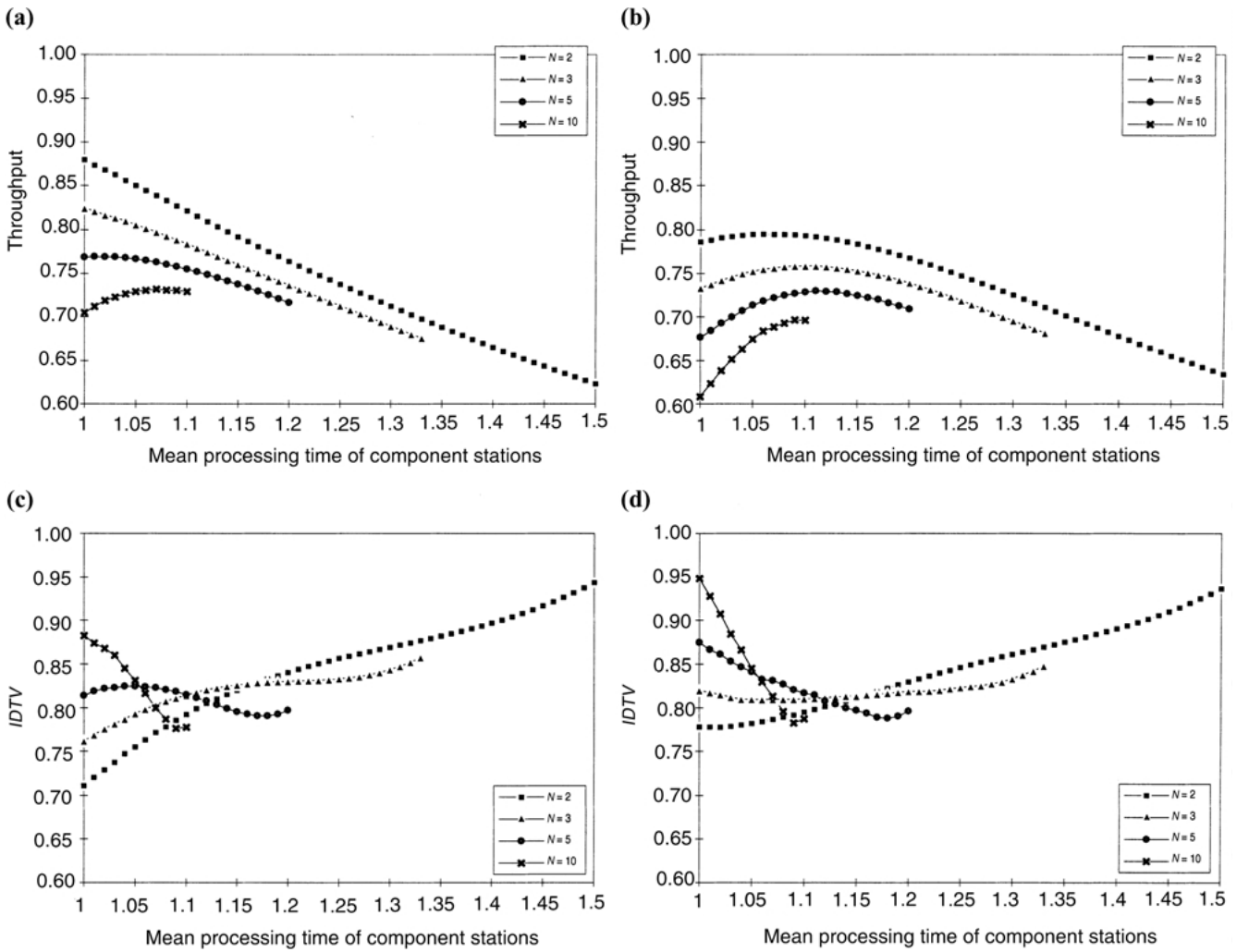


Fig. 17. The effect of WT/CV - on throughput and interdeparture time variability for a buffer size of one and lognormal processing times $CV = 0.6$ for: (a) a separated buffer allocation; (b) a pooled buffer allocation; (c) a separated buffer allocation; and (d) a pooled buffer allocation.

that WT/CV - creates the hump behavior as discussed in Section 5. Since buffers absorb the effect of work transfer, this intrinsic behavior of optimal T and $IDTV$ is flattened in the buffered systems (see Fig. 18a in comparison to Fig. 14b). We also note that the separated buffer configuration yields a better T and $IDTV$ than the pooled type.

7. Discussion

In this paper, we have examined the effects of various design factors (i.e., parallelism, processing time distributions, work and variability transfers, buffers and buffer allocation schemes) on throughput and interdeparture time variability of assembly systems. Based on extensive computational experiments, we have obtained several important findings that can guide practitioners

to design more effective systems and open new research avenues for academics. These new findings are summarized below:

1. The effect of work transfer on throughput is more pronounced in large systems with a low CV (which is the typical situation in practice).
2. Variability transfer also improves throughput. However the magnitude of this improvement is not as much as work transfer. The positive effect of variability transfer on throughput is again more significant in large systems.
3. In contrast to throughput, the positive effect of work transfer on interdeparture time variability is significant only in the systems with a high CV .
4. Variability transfer is also an effective tool to improve interdeparture time variability, especially in large systems.

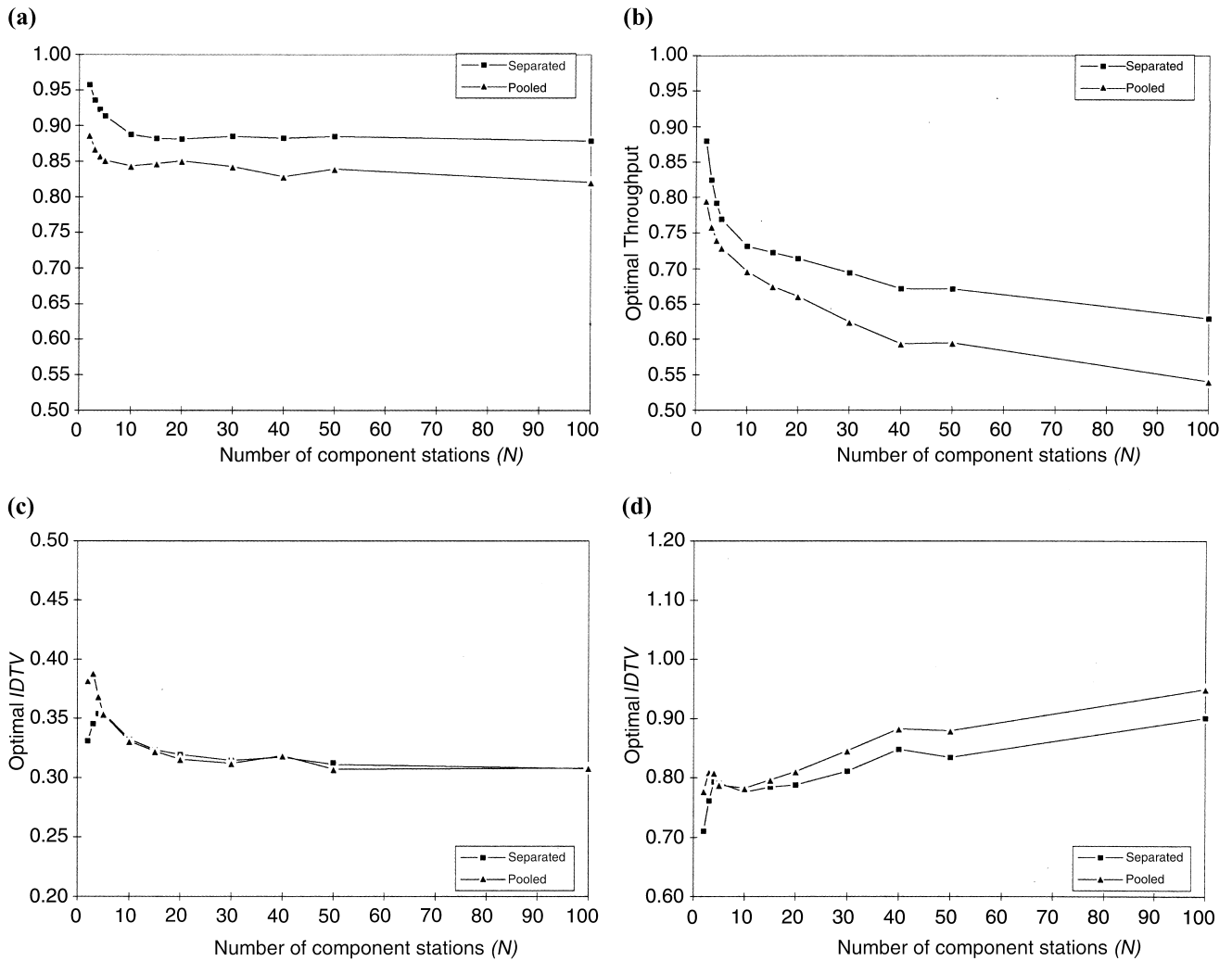


Fig. 18. The effect of $WT/CV/T$ on throughput and $WT/CV/IDTV$ on interdeparture time variability for a buffer size of one and lognormal processing times when: (a) $CV = 0.3$; (b) $CV = 0.6$; (c) $CV = 0.3$; and (d) $CV = 0.6$.

5. For large systems, work transfer that optimizes throughput also optimizes interdeparture time variability (the converse is also true). For small systems, however, variability transfer improves both throughput and interdeparture time variability more than work transfer.
6. The above results hold when the interdeparture time variability is measured considering correlation.
7. The hump behavior of optimal throughput is a distribution-free phenomenon and emerges in the systems with a low CV due to different improvement rates of throughput for different numbers of component stations.
8. The hump behavior is also observed for interdeparture time variability due to the same reasons explained above.
9. In the buffered systems, the separated configuration displays a better performance (throughput and interdeparture time variability) than the pooled con-

figuration. The same observation is also made for optimal throughput and interdeparture time variability. In general, buffers diminish the effects of work transfer and N on the performance and flatten the hump behavior.

Even though several features of assembly systems have been thoroughly analyzed in this paper, there still remain various research issues to be addressed. First, the work presented in this paper can be extended to systems in which the total work content is constant (i.e., total work content does not depend on N). Second, it would be interesting to compare pull and push loading mechanisms in assembly systems for interdeparture time variability and average WIP inventory. Third, it would be interesting to test the validity of the results when different distribution functions are used in the component stations and the assembly station. Along the same line, the fourth research item would be to test the results for higher moments

(third, fourth, and higher). Finally, two major lines of research in the areas of serial production and assembly systems can be combined in a single study to understand the interactions between these two systems and draw more general conclusions.

References

- Baker, K.R. (1992) Tightly-coupled production systems: models, analysis, and insights. *Journal of Manufacturing Systems*, **11**, 385–400.
- Baker, K.R. and Powell, S.G. (1995) A predictive model for the throughput of simple assembly systems. *European Journal of Operational Research*, **81**, 336–345.
- Baker, K.R., Powell, S.G. and Pyke, D.F. (1990) Buffered and unbuffered assembly systems with variable processing times. *Journal of Manufacturing and Operations Management*, **3**, 200–223.
- Baker, K.R., Powell, S.G. and Pyke, D.F. (1993) Optimal allocation of work in assembly systems. *Management Science*, **39**, 101–106.
- Bhatnagar, R. and Chandra, P. (1994) Variability in assembly and competing systems: effect on performance and recovery. *IIE Transactions*, **26**, 18–31.
- Erel, E., Sabuncuoglu, I. and Kok, A.G. (1997) Analyses of serial production line systems for interdeparture time variability and WIP inventory. Working paper, IE-OR 9618, Department of Industrial Engineering, Bilkent University.
- Hendricks, K.B. (1992) The output process of serial production lines of exponential machines with finite buffers. *Operations Research*, **40**, 1139–1147.
- Hendricks, K.B. and McClain, J.O. (1993) The output processes of serial production lines of general machines with finite buffers. *Management Science*, **39**, 1194–1201.
- Knott, K. and Sury, R.J. (1987) A study of work-time distributions on unpaced tasks. *IIE Transactions*, **19**, 50–55.
- Lau, H.S. (1992) On balancing variances of station processing times in unpaced lines. *European Journal of Operational Research*, **56**, 345–356.
- Law, A.M. and Kelton, W.D. (2000) *Simulation Modeling and Analysis*. 3rd edn., McGraw-Hill, Inc., Singapore.
- Powell, S.G. and Pyke, D.F. (1998) Buffering unbalanced assembly systems. *IIE Transactions*, **30**, 55–65.
- Rekhi, I., Chand, S. and Moskowitz, H. (1995) Optimal allocation of work in assembly systems revisited. Working paper, Krannert School of Management, Purdue University, West Lafayette, IN 47907-1287, USA.
- Simon, J.T. and Hopp, W.J. (1995) Throughput and average inventory in discrete balanced assembly systems. *IIE Transactions*, **27**, 368–373.

Biographies

Ihsan Sabuncuoglu is an Associate Professor of Industrial Engineering at Bilkent University. He received B.S. and M.S. degrees in Industrial Engineering from the Middle East Technical University and a Ph.D. degree in Industrial Engineering from the Wichita State University. Dr. Sabuncuoglu teaches and conducts research in the areas of simulation, scheduling, and manufacturing systems. He has published papers in *IIE Transactions*, *Decision Sciences*, *Simulation*, *International Journal of Production Research*, *International Journal of Flexible Manufacturing Systems*, *International Journal of Computer Integrated Manufacturing*, *Computers and Operations Research*, *European Journal of Operational Research*, *International Journal of Production Economics*, *Production Planning, Control*, *Journal of Operational Research Society*, *Journal of Intelligent Manufacturing* and *Computers and Industrial Engineering*. He is on the Editorial Board of the *International Journal of Operations and Quantitative Management*. He is an associate member of Institute of Industrial Engineering and Institute for Operations Research and the Management Science.

Erdal Erel is an Associate Professor in the Department of Business Administration at Bilkent University. He received his B.S. in Industrial Engineering from the Technical University of Istanbul, Turkey, and a M.S. from Stanford University. He received his Ph.D. in Industrial Engineering and Operations Research from the Virginia Polytechnic Institute and State University. His research interests are in the areas of manufacturing systems analysis, and production planning and control. He has published papers in *International Journal of Production Research*, *International Journal of Production Economics*, *International Journal of Operations and Production Management*, *Omega*, *Production Planning and Control*, *European Journal Operational Research*, and *Annals of Operations Research*. He is a member of the Institute of Industrial Engineering and Institute for Operations Research and Management Science.

A. Gurhan Kok is a Ph.D. student in the Operations and Information Management Department at the Wharton School, University of Pennsylvania. He received his M.S. and B.S. degrees from the Department of Industrial Engineering at Bilkent University. His research interests include design and analysis of production systems.

Contributed by the Manufacturing Systems Control Department