Analysis of gap times in a two-stage stochastic flowshop with overlapping operations

Inneke Van Nieuwenhuyse • Nico Vandaele

Department of Applied Economics, University of Antwerp, Prinsstraat 13, 2000 Antwerp, Belgium

inneke.vannieuwenhuyse@ua.ac.be • nico.vandaele@ua.ac.be

The impact of transfer batching (also referred to as *lot splitting*) on the performance of flowshops has received widespread attention in the literature. Most papers have emphasized the usefulness of lot splitting in cutting down average flow times, as it enables the overlapping of operations at different stages of the flowshop. However, as most analytical papers have studied deterministic flowshops, an important downside of lot splitting has until now been overlooked: i.e., the occurrence of idle times between the processing of consecutive sublots belonging to the same process batch (referred to as *gap times*). Clearly, gap times add no value to the product: they merely increase the process batch makespan at the different stages. In deterministic systems, these gap times may be avoided by synchronizing the processing rates of the different machines in the shop; in stochastic settings however, they may occur even when the system is perfectly synchronized, due to the inherent variability in the setup and processing times. Studying a two-stage flowshop with a single product type, this paper provides insight into the behavior of the gap times, and develops an approximation for the process batch makespan at the second stage in terms of the system characteristics and the lot splitting policy.

1. Introduction

In the operations management literature, it is widely acknowledged that batch sizing decisions influence performance measures such as cycle times and work-in-process levels. As these are decisive factors for the responsiveness of any production environment, setting batch sizes in a production system is an important control (see e.g. Hopp and Spearman 2000, Lambrecht et al. 1998, Benjaafar 1996).

When studying the impact of batching decisions in a production environment, one should make a distinction between two types of batches: *process batches* and *transfer batches*. A process batch (also referred to as a *production batch* or *production lot*) is defined as the quantity of a product processed on a machine without interruption by other items (Kropp

and Smunt 1990). In multiple-product environments, the use of process batches is often unavoidable due to capacity considerations: to switch from one product type to the next (e.g., to change fixtures or dies), a setup or changeover time is necessary, which consumes part of the capacity of the machine. After a setup has been performed, a certain quantity of the product (the process batch size) can be produced. Hence, a process batch can also be defined as the quantity of a product produced between two consecutive setups.
A transfer batch (or *transfer lot*) refers to the size of a sublot of the process batch, moved after production on one machine to another operation or machine (Kropp and Smunt 1990). The use of transfer batches is not linked to capacity considerations, but rather to flow considerations. Indeed, it is widely accepted that the use of transfer batch sizes smaller than the process batch size can reduce product flow times by smoothing workflow and minimizing congestion levels (e.g. Santos and Magazine 1985, Benjaafar 1996, Goldratt and Cox 1984, Hopp et al. 1990 and Umble and Srikanth 1995). This is due to the mechanism of *overlapping operations*: by allowing transportation of partial batches to a downstream station, this station can already start processing these partial batches while work proceeds at the upstream station, thereby accelerating the progress of work through the production facility (e.g. Graves and Kostreva 1986, Jacobs and Bragg 1988, Litchfield and Narasimhan 2000).

There exists a large body of research on the impact of lot splitting in deterministic flowshop environments. A large variation of lot streaming models has been proposed, depending on the constraints of the environment that one wishes to model: e.g., it may be assumed that all sublots have to be of equal size (as in Jacobs and Bragg 1988, Kalir and Sarin 2001), and/or it may be imposed that sublots have to be consistent throughout the system, implying that the same sublot sizes have to be used at each machine (as in Van Nieuwenhuyse and Vandaele 2004). The primary objective of these papers is to determine the optimal lot streaming procedure in order to minimize either the process batch makespan (e.g. Chen and Steiner 1996, Chen and Steiner 1998, Cheng et al. 2000), mean transfer batch flow time (Kalir and Sarin 2001, Sen et al. 1998, Bukchin et al. 2002, Van Nieuwenhuyse and Vandaele 2004), or a combination of both (Bukchin and Masin 2004) in such an environment.

While the favorable impact on flow times has received widespread attention, the use of lot splitting also has a downside: it may lead to idle times between the processing of consecutive sublots, belonging to the same process batch. These idle times are referred to as *gaps*, and are caused by the fact that the setup and processing times at the different stages in

the flowshop are not synchronized (see Van Nieuwenhuyse and Vandaele, 2004): hence, it may happen that a downstream stage finishes processing a sublot before the next sublot is available from the upstream stage.
In a deterministic environment, this downside can be avoided by explicitly imposing that a machine may not remain idle between consecutive sublots. This is referred to as the *no-idling assumption* (e.g. Baker and Jia 1993, Ramasesh et al. 2000). However, in a stochastic flowshop, the occurrence of gaps is a major issue: as setup and processing times at the different stages are variable, gaps may occur even when the average setup and processing times are synchronized. The occurrence of these gap times leads to an increase in the makespan of a process batch, without adding value to the product. In fact, gap times may even represent a cost for the system: during a gap time, the server has to be kept operational, i.e. ready for processing the next transfer batch when it arrives (Van Nieuwenhuyse and Vandaele 2004). Depending upon the type of server, this may entail labor and/or energy costs. Obviously, it is desirable from the point of view of a planner to have an estimate of the extent to which gap times will influence the average of the makespan of a process batch at a given stage, such that it can be taken into account in the planning system.

The objective of this paper is to develop insights and approximations for the average gap time in a two-stage stochastic flowshop with general setup and processing times. Obviously, in a stochastic system, the gap time will be a stochastic variable, influenced by the level of variability in the system and the lot splitting policy. Focusing on a flowshop with a single product type, we can derive a lower bound for the average process batch makespan at the second stage, as well as an approximation. The performance of the approximation is tested by means of an extensive simulation experiment, and turns out to be very satisfactory.
To the best of our knowledge, this paper is the first to take a closer look at the occurrence of gap times in a stochastic environment (analytical expressions for the gap times in a deterministic flowshop were derived previously in Van Nieuwenhuyse and Vandaele 2004).
The paper is organized as follows: in section 2, we first describe the assumptions of our setting. Next, section 3 presents some basic insights for the average gap time, along with a lower bound. As will be shown, the average gap time is analytically intractable. The remainder of the paper then focuses on estimating the average gap time, in order to yield a satisfactory approximation for the average process batch makespan at the second stage. This is done by means of simulation. Section 4 describes the setting of the simulation

experiment, and examines a suitable approximation for the average gap time based on the simulation results. Section 5 discusses the resulting approximations for the average process batch makespan, and tests the performance of the approximation versus simulation results, using the lower bound as a benchmark. In section 6, we apply the expressions to an example. Finally, section 7 summarizes the conclusions.

2. Notation and assumptions

As mentioned before, we will consider a two-stage stochastic flowshop with a single product type. It is assumed that products arrive in process batches of size N in front of the first stage. The product units in the process batch are processed one by one on each of the two servers. After processing on the first stage, products are collected to form a transfer batch of size L ($L \leq N$). As soon as a transfer batch is complete, it is moved to the next server (the transportation time between the two stages is not explicitly modelled). For simplicity as well as practical reasons, L is supposed to be a divisor of N, such that a single process batch is split into an integer number (T) of transfer batches ($T = N/L$). After stage two, the transfer batches belonging to the same process batch are regrouped for shipment.

Each stage m ($m = 1, 2$) requires a setup time $SU_m > 0$ to be performed at the start of every process batch. This setup time may be necessary, even in single-product type settings, e.g. when the product type is produced in different colors or sizes. In that case, a setup time is necessary at the start of every new process batch, in order to change the paint or the fixtures. While some papers in the literature consider setup times to be necessary at the start of every transfer batch (e.g. Bukchin et al. 2002), we consider setups only at the start of a process batch; it is assumed that setup times for the separate transfer batches (e.g. for mounting the parts on the machine) are negligible.

The processing time for a transfer batch i on stage m is denoted by $X_{m,i}$ with i ranging from 1 to T. In our two-stage setting, each of the stages is assumed to be a capacity server. This means that the transfer batch processing time on each of the stages is dependent on the transfer batch size, and hence, $X_{m,i}$ can be expressed as the sum of L unit processing times on server m (x_m), for any arbitrary i:

$$X_{m,i} = \sum_{j=(i-1)*L+1}^{i*L} x_{m,j}$$

Setup times as well as unit processing times are assumed to be random variables, with an arbitrary probability distribution. We will assume that the setup time on stage two can not start before the first transfer batch of the involved process batch is present in the input buffer of the second stage. This type of setup has been referred to in the literature as an *attached* setup (Potts and Kovalyev 2000, Chen and Steiner 1998). Hence, the first transfer batch of a process batch acts as a *flag* (Smunt et al. 1996): its arrival in front of stage two authorizes the start of the setup, thereby causing the operations on stage two to partly overlap with the operations on the first stage.

The buffers in front of the stages are assumed to have infinite capacity. When arriving in front of the second stage, it may happen that the flag has to wait in queue before the setup can be performed (e.g., when the server is still processing a transfer batch belonging to the previous process batch); this waiting time will be denoted by W_2.

For illustrative purposes, Figure 1 shows the progress of a process batch, consisting of 3 sublots, through a two-stage system.

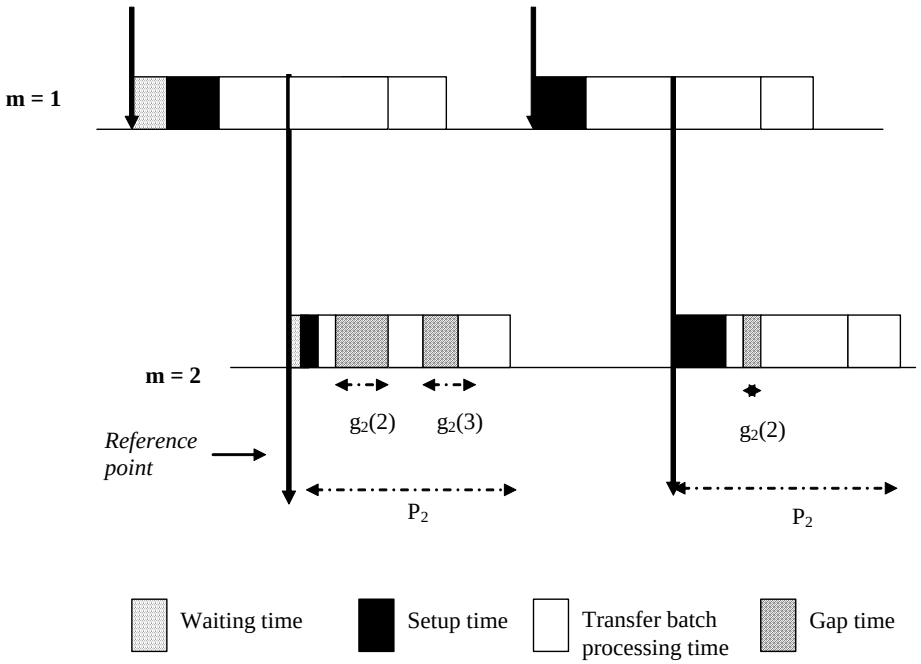

Figure 1: Flowchart of a process batch, consisting of 3 sublots, going through a 2-stage system

The makespan of a process batch at stage 2 is denoted by P_2; it is a stochastic variable, consisting of the setup time at stage 2 (SU_2), the individual transfer batch processing times ($X_{2,i}$), and the total gap time (G_2). This total gap time consists of a number of *partial* gaps

$g_2(i)$(where i denotes the number of the transfer batch following the gap):

$$G_2 = \sum_{i=2}^{T} g_2(i)$$

It is clear that a gap can never occur in front of the flag (hence, $g_2(1) = 0$), due to the attached setup at stage 2. It is also useful to note that in our definition, a gap time is an idle time occurring between the processing of two consecutive sublots belonging to the same process batch. Hence, the gap time does not include the idle time that may occur between the processing of two distinct process batches.

3. Basic insights and lower bound for the average gap time

For studying the total gap time G_2, it is clear that we only have to take into account how the system behaves after the so-called *reference point*, i.e. the point in time when the flag is finished on stage 1 and moved to the next machine (as depicted in Figure 1). As mentioned above, G_2 is given by:

$$G_2 = \sum_{i=2}^{T} g_2(i) \tag{1}$$

A partial gap $g_2(i)$ can be written as:

$$g_2(i) = \max[\underbrace{\sum_{k=2}^{i} X_{1,k} - (W_2 + SU_2 + \sum_{j=1}^{i-1}[X_{2,j} + g_2(j)])}_{V(i)}, 0] \qquad (i = 2, ..., T) \tag{2}$$

From this expression, we can derive that the probability distribution of a partial gap time $g_2(i)$ will have a zero-inflated shape: the probability that $g_2(i)$ equals zero is given by the probability that $V(i)$ is smaller than or equal to zero. For $g_2(i) > 0$, the density function of $g_2(i)$ is equal to the density function of $V(i)$.

Based upon expressions (1) and (2), the total gap G_2 can be written as:

$$G_2 = \sum_{i=2}^{T} \max[\underbrace{\sum_{k=2}^{i} X_{1,k} - (W_2 + SU_2 + \sum_{j=1}^{i-1}[X_{2,j} + g_2(j)])}_{V(i)}, 0] \tag{3}$$

which can be rewritten as follows (the proof is given in Appendix 1):

$$G_2 = \max[Z(2), ..., Z(T), 0] \tag{4}$$

with $Z(i) = \sum_{k=2}^{i} X_{1,k} - (W_2 + SU_2 + \sum_{j=1}^{i-1} X_{2,j})$ for $(i = 2, ..., T)$.
This expression is very useful, as it shows that the total gap is in fact a random variable consisting of the maximum of $T-1$ correlated random variables $Z(t)$ (with $t = 2, ..., T$) and a constant (zero). However, the exact distribution of each of the variables $Z(t)$ is unknown.

Based on expression (4), we can write the following for the average of the total gap:

$$E[G_2] = E[\max[Z(2), ..., Z(T), 0]]$$

As the exact distribution of each of the variables $Z(t)$ is unknown, $E[G_2]$ is analytically intractable. However, it is possible to derive a tractable lower bound. Indeed, based on Jensen's inequality (e.g., Ross 1996), we know that:

$$E[\max[Z(2), ..., Z(T), 0]] \geq \max[E[Z(2)], ..., E[Z(T)], 0]$$

Hence:

$$E[G_2]_{LB} = \max[E[Z(2)], ..., E[Z(T)], 0] \tag{5}$$

As $E[G_2]$ is analytically intractable, we need to develop an approximation. When testing the appropriateness of this approximation (see section 5), we can use the performance of this lower bound as a benchmark.

4. Approximation for the average gap time

4.1 Simulation experiment

As mentioned above, an analysis of the average gap time $E[G_2]$ can be made by considering the system behavior after the reference point (see e.g. Figure 1). In view of this analysis, a simulation experiment was designed using the following variables as factors:

- T: the number of transfer batches per process batch;
- L: the transfer batch size;
- $E(X_1)$: the average processing time per transfer batch on stage 1;

- $E(X_2)/E(X_1)$: the ratio of the average processing time per transfer batch on stage 2 versus the average processing time per transfer batch on stage 1;
- $E(SU_2)/E(X_1)$: the ratio of the average setup time per process batch on stage 2 versus the average processing time per transfer batch on stage 1;
- $c^2_{X_1}$: the squared coefficient of variation (SCV) of the processing time per transfer batch on stage 1;
- $c^2_{X_2}$: the SCV of the processing time per transfer batch on stage 2;
- $c^2_{SU_2}$: the SCV of the setup time on stage 2;
- the probability distribution used for SU_2, X_1, and X_2;
- ρ_1: the utilization rate of stage 1.

The average and variance of SU_1 were not used as factors in the simulation model, as SU_1 occurs before the reference point. Hence, we have reason to believe that the average and the SCV of SU_1 only have a minor impact on the outcome for $E[G_2]$ in the stochastic model: both parameters can at best have an indirect impact, as they may influence $E[W_2]$, which on its turn affects $E[G_2]$ (as evident from expression (4)). Given its minor importance, SU_1 was arbitrarily fixed at 15 time units (deterministic) in the experiment.

Table 1 gives an overview of the levels that were used in the simulation for the different factors. Combining all factors at all levels, this design resulted in 23,328 runs. The length of each run was fixed at 100,000 process batches, the first 20,000 of which were considered as warm-up period.
The interarrival times of process batches in front of stage 1 are deterministic, and controlled in order to yield the desired utilization rate ρ_1. Note that the combination of certain levels of ρ_1, $\frac{E(X_2)}{E(X_1)}$ and $\frac{E(SU_2)}{E(X_1)}$ yields a utilization rate higher than or equal to 1 on stage 2; obviously, these runs were skipped.

4.2 Approximation for the density function of the total gap time G_2

By plotting the observed frequency distribution of G_2 for different simulation runs of the experiment, we could draw important insights about the behavior of the gap time in terms

Factor	Levels	Number of levels
$T = \frac{N}{L}$	10,15,20	3
L	1	1
$E(X_1)$	10	1
$\frac{E(X_2)}{E(X_1)}$	$0.2, 0.5, 0.8, 1, 1.2, 1.5, 1.8, 2$	8
$\frac{E(SU_2)}{E(X_1)}$	$\frac{1}{3}, \frac{1}{2}, 1, \frac{3}{2}$	4
$c^2_{X_1}$	0.3,0.6,0.9	3
$c^2_{SU_2}$	0.3,0.6,0.9	3
$c^2_{X_2}$	0.3,0.6,0.9	3
probability distribution	gamma, beta, lognormal	3
ρ_1	0.4, 0.7, 0.95	3

Table 1: Input factors and levels for the simulation experiment

of the difference in average processing rates, and the difference in processing time variability on the two stages.

Figure 2 shows the histograms (gamma, beta and lognormal) for G_2 when $T = 20$, $\rho_1 = 0.4$, $\frac{E(SU_2)}{E(X_1)} = 0.5$, and $c^2_{X_1} = c^2_{SU_2} = c^2_{X_2} = 0.3$. The figure in the top pane refers to the setting where $\frac{E(X_2)}{E(X_1)} = 0.2$, so where stage 2 has a significantly higher processing rate than stage 1. As expected, the total gap time observed in this setting is always larger than zero. The probability distribution seems to be strikingly close to normal. Moreover, the histograms of the beta, gamma and lognormal distributions almost coincide.

The figure in the bottom pane refers to the setting where $\frac{E(X_2)}{E(X_1)} = 0.8$, so where the two processing rates are almost equal. This histogram is clearly *zero-inflated*: it shows a sharp peak at $G_2 = 0$. This is not surprising: it is intuitively clear that, as the ratio of $E(X_2)$ to $E(X_1)$ increases, the observed total gap time G_2 will more frequently equal zero. More importantly, the positive values of G_2 appear to be close to a normal distribution. Note that the variance of this distribution is considerably higher than the variance observed in the top pane.

The histograms in Figure 3 give further insight on the impact of variability: they show the observed frequency distribution of G_2 for the same settings as Figure 2, but with a higher SCV for setup and processing times ($c^2_X = c^2_{SU_2} = c^2_{X_2} = 0.9$).

Apparently, even in situations when the processing and setup times are highly variable, the density of G_2 is still close to normal. A comparison of the histograms in the top panes

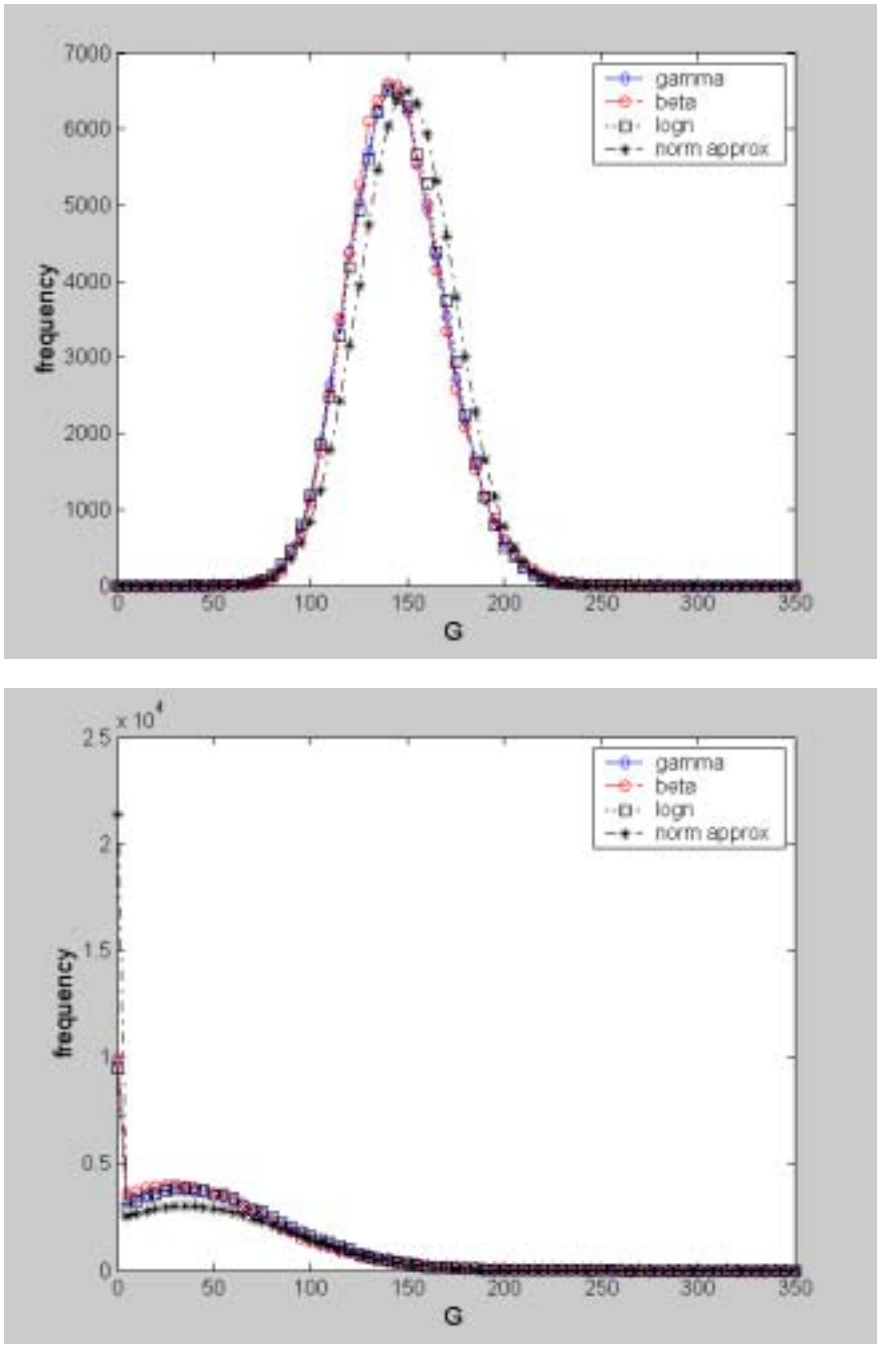

Figure 2: Sample histogram for G_2 when $\frac{E(X_2)}{E(X_1)} = 0.2$ (top pane), and $\frac{E(X_2)}{E(X_1)} = 0.8$ (bottom pane), when $c^2_{X_1} = c^2_{SU_2} = c^2_{X_2} = 0.3$.

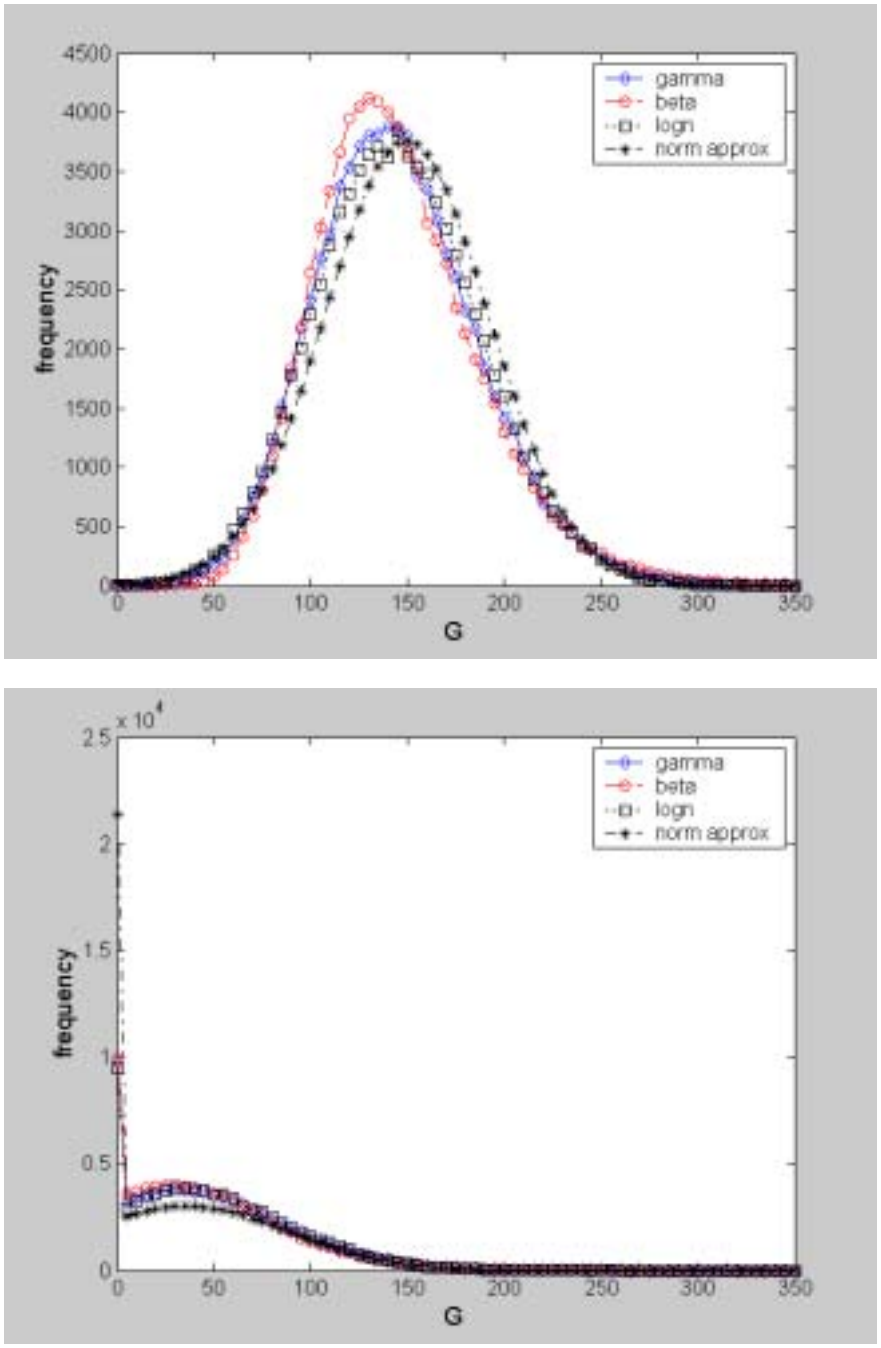

Figure 3: Sample histogram for G_2 when $\frac{E(X_2)}{E(X_1)} = 0.2$ (top pane), and $\frac{E(X_2)}{E(X_1)} = 0.8$ (bottom pane), when $c_X^2 = c_{SU_2}^2 = c_{X_2}^2 = 0.9$.

reveals that a higher variability in setup and processing times leads to a higher variability in G_2.

Histograms for other simulation settings yielded similar results. The histograms lead us to conjecture that the density function of G_2 can be reasonably approximated by the following:

$$G_2 = \max[V, 0], \quad \text{with } V \simeq N(m_V, \sigma_V^2) \tag{6}$$

i.e., G_2 can be written as the maximum of a normal variable V and 0. Following this argument, the probability mass observed at $G_2 = 0$ in settings with high values of $\frac{E(X_2)}{E(X_1)}$ can be approximated by the cumulative distribution function of V at 0:

$$prob(G_2 = 0) = \int_{-\infty}^{0} f_V(x) * dx = F_V(0)$$

where $f_V(x)$ and $F_V(x)$ denote respectively the (normal) density function and the (normal) cumulative distribution function of V.

The issue now is to determine m_V and σ_V^2, the mean and variance of this underlying normal distribution. Based upon further study of the histograms, the following candidates can be proposed (Van Nieuwenhuyse 2004):

$$\begin{aligned} m_V &= (T-1) * [E(X_1) - E(X_2)] - E(SU_2) \\ \sigma_V^2 &= (T-1) * [Var(X_1) + Var(X_2)] + Var(SU_2) \end{aligned} \tag{7}$$

For illustrative purposes, the density function of the corresponding (zero-inflated) normal distribution, with m_V and σ_V^2 given as in expression (7), is depicted in Figures 2 and 3, revealing that this approximation provides a close fit.

It is particularly useful to confront the proposed approximation for G_2 with the theoretically exact expression derived above (expression (4)). As mentioned before, this expression shows that the total gap time G_2 is a random variable given by the maximum of $T-1$ correlated random variables (where T refers to the number of sublots), and a constant (zero). Using the parameters for m_V and σ_V^2 proposed in (7), V is de facto approximated by:

$$\begin{aligned} V &= Z(T) + W_2 \\ &= \sum_{k=2}^{T} X_{1,k} - (SU_2 + \sum_{j=1}^{T-1} X_{2,j}) \end{aligned}$$

Hence, our approximation implicitly makes three simplifying assumptions:

1. **The impact of $\mathbf{W_2}$ on $\mathbf{G_2}$ is ignored:**

   Indeed, expression (4) (which is analytically exact) reveals that there is a negative relationship between the waiting time of the flag in front of stage 2 (W_2) and the gap time occurring on stage 2 (G_2). As we ignore this relationship, our approximation will tend to overestimate of the average gap time in cases where stage 2 is highly utilized. However, we do not expect this simplification to lead to significant errors on the estimation of the process batch makespan on stage 2: large values for W_2 are bound to occur when the processing rate of stage 2 is relatively low, implying that gap times will constitute a very small (or even negligible) part of the process batch makespan when compared to the actual processing time.

2. **The impact of $\mathbf{Z(2)}$ to $\mathbf{Z(T-1)}$ on $\mathbf{G_2}$ is ignored:**

   Expression (4) reveals that it suffices to have a positive value for any arbitrary $Z(i)$ $(i = 2, ..., T)$ in order to have a positive gap time on stage 2. In our approximation, we only take into account the impact of $Z(T)$. This simplification may negatively impact the performance of our approximation in systems where $E(X_1)$ is close to $E(X_2)$, or larger than $E(X_2)$. In such systems, there is a higher probability for $Z(T)$ to be negative while one of the other $Z(i)$ $(i = 2, ..., T-1)$ is positive. Hence, we can expect the approximation to underestimate the probability of a gap time occurring for this type of systems. Conversely, we can expect that the simplification will have little effect on the performance of our approximation, as long as $E(X_1)$ is significantly larger than $E(X_2)$.

3. **We assume that $\mathbf{Z(T) + W_2}$ can be reasonably approximated by a normal probability distribution:**

   For high values of T, we can indeed expect that the density function of $Z(T) + W_2$ will be close to normal by virtue of the central limit theorem. As the processing times $X_{1,k}$ on stage 1 are IID distributed, $\sum_{k=2}^{T} X_{1,k}$ will approach a normal distribution with mean $m_1 = (T-1) * E[X_1]$ and variance $\sigma_1^2 = (T-1) * Var[X_1]$ for high values of T. Similarly, $\sum_{k=2}^{T} X_{2,k}$ will also approach a normal distribution with mean $m_2 = (T-1) * E[X_2]$ and variance $\sigma_2^2 = (T-1) * Var[X_2]$. As the difference of two normal distributions is again a normal distribution (e.g. Blumenfeld 2001), we can expect that, for high values of T, $\sum_{k=2}^{T} X_{1,k} - \sum_{k=2}^{T} X_{2,k}$ will be normally distributed with mean $m = m_1 - m_2$ and variance $\sigma^2 = \sigma_1^2 + \sigma_2^2$. The only disturbing factor is the

presence of SU_2 in $Z(T)$; however, if SU_2 is relatively small compared to $\sum_{k=2}^{T} X_{1,k}$ and $\sum_{k=2}^{T} X_{2,k}$, we can expect its impact to be negligible.

4.3 Approximation for $E[G_2]$

Using m_V and σ_V^2 as in expression (7), and assuming a normal distribution, the approximation for $E[G_2]$ is given by:

$$\begin{aligned} E[G_2]_{app} &\approx \int_0^{+\infty} x * [\frac{1}{\sqrt{2\pi\sigma_V^2}} * Exp\{\frac{-(x-m_V)^2}{2\sigma_V^2}\}]dx \\ &= [\frac{-\exp\{\frac{-(x-m_V)^2}{2\sigma_V^2}\} * \sigma_V^2 + m_V * \sqrt{\frac{\pi}{2}} * \sigma_V * Erf[\frac{x-m_V}{\sqrt{2}\sigma_V}]}{\sqrt{2\pi}\sigma_V}]_0^{+\infty} \\ &= \frac{m_V}{2} + \frac{Exp\{\frac{-m_V^2}{2\sigma_V^2}\} * \sigma_V^2}{\sqrt{2\pi}\sigma_V} + \frac{m_V}{2} * Erf[\frac{m_V}{\sqrt{2}\sigma_V}] \end{aligned} \quad (8)$$

in which $Erf[z]$ refers to the error function:

$$Erf[z] = \frac{2}{\sqrt{\pi}} \int_0^z \exp\{-t^2\}dt$$

Note that expression (8) satisfies the lower bound defined in expression (5). Indeed, we can rewrite expression (8) as follows:

$$\begin{aligned} E[G_2]_{app} &\approx m_V - \underbrace{\int_{-\infty}^{0} x * [\frac{1}{\sqrt{2\pi\sigma_V^2}} * Exp\{\frac{-(x-m_V)^2}{2\sigma_V^2}\}]dx}_{<0} \\ &> m_V \end{aligned}$$

As $m_V = E[Z(T) + W_2] = E[Z(T)] + E[W_2]$, we then have:

$$E[Z(T)] < m_V < E[G_2]_{app}$$

which proves that the approximation satisfies the lower bound.

5. Performance of the approximation for the average process batch makespan

Using expression (8) for $E[G_2]_{app}$, we can now develop an approximation for $E[P_2]$:

$$\begin{aligned} E[P_2]_{app} &\approx E(SU_2) + T * E(X_2) \\ &\quad + \frac{m_V}{2} + \frac{Exp\{\frac{-m_V^2}{2\sigma_V^2}\} * \sigma_V^2}{\sqrt{2\pi}\sigma_V} + \frac{m_V}{2} * Erf[\frac{m_V}{\sqrt{2}\sigma_V}] \end{aligned} \quad (9)$$

Moreover, the lower bound on $E[G_2]$ in expression (5) gives us the following lower bound on $E[P_2]$:

$$E[P_2]_{LB} = E(SU_2) + T * E(X_2) + \max[E[Z(2)], ..., E[Z(T)], 0] \tag{10}$$

The performance of this approximation was tested versus the simulation results of the experiment, by determining the relative error ε of the approximation ($E[P_2]_{app}$) to the observed estimate of $E[P_2]$ in the simulations ($E[P_2]_{sim}$):

$$\varepsilon = \frac{(E[P_2]_{app} - E[P_2]_{sim})}{E[P_2]_{sim}}$$

and the absolute values of these relative errors:

$$\varepsilon_{abs} = \left| \frac{(E[P_2]_{app} - E[P_2]_{sim})}{E[P_2]_{sim}} \right|$$

As a benchmark, we also determined the performance of the lower bound.

Table 2 gives an overview of the resulting relative errors of the approximation, along with those for the lower bound.

	$E[P_2]_{app}$	$E[P_2]_{LB}$
average ε	-1.54	-4.36
stdev ε	1.62	4.13
median ε	-1.42	-2.93
5% percentile ε	-4.23	-12.47
95% percentile ε	0.76	-0.07
max ε	5.53	0.00004
min ε	-5.41	-17.88
average ε_{abs}	1.78	4.36
stdev ε_{abs}	1.36	4.13

Table 2: Summary statistics for the relative errors of $E[P_2]_{app}$ and $E[P_2]_{LB}$ compared to $E[P_2]_{sim}$ (in percent)

The table reveals that the range of relative errors on the approximation is very small, and that the overall performance of the approximation is satisfactory. In general, $E[P_2]_{app}$ tends to underestimate $E[P_2]_{sim}$: this is evident from the histogram of the relative errors, shown in the top pane of Figure 4. The histogram also shows that the observed frequency of the relative errors is largest at the bin containing $\varepsilon = 0$, which is certainly a desirable characteristic.

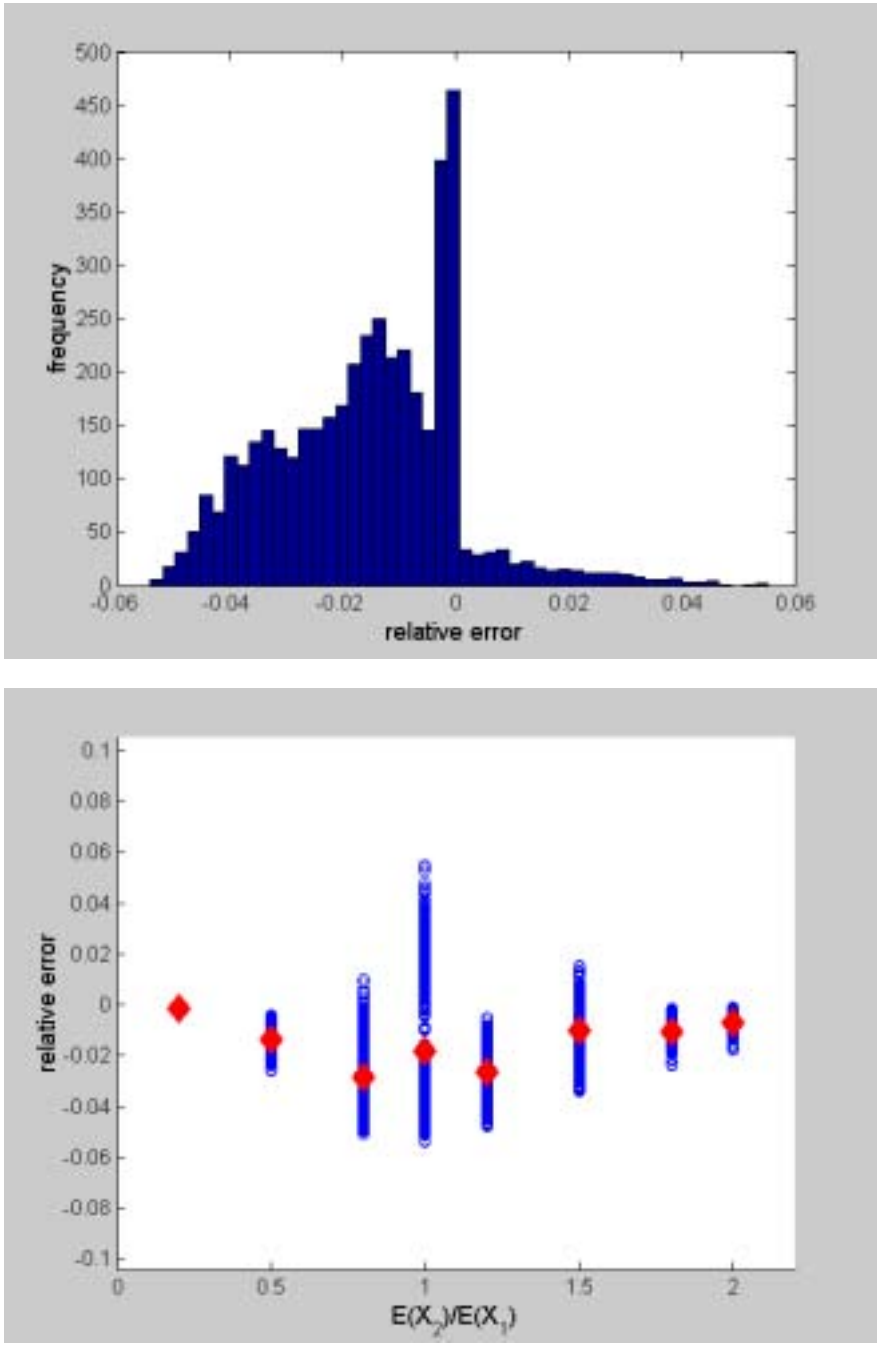

Figure 4: Histogram and scatterplot of the relative errors of $E[P_2]_{app}$ versus $E[P_2]_{sim}$

The tendency to underestimate is not surprising: as mentioned before, G_2 is theoretically given by the maximum of $(T-1)$ correlated random variables $(Z(2),...,Z(T))$, and zero (see expression (4)). Hence, G_2 will be positive whenever one of the $Z(i)$ $(i=2,...,T-1)$ is positive; our approximation however only reflects positive values for $Z(T)$, as we have neglected the impact of $Z(2)$ to $Z(T-1)$ (see assumption 2 above).
As expected, the precision of the approximation decreases slightly when $E(X_2)$ is close to $E(X_1)$. This is confirmed by the bottom pane of figure 4, which shows a scatterplot of the relative errors of the experiment in terms of $\frac{E(X_2)}{E(X_1)}$.

However, this observation does not undercut the power of the model; on the contrary, it is exactly in these settings that the model proves its usefulness to the fullest. As mentioned in the introduction, it is precisely in the case of synchronized stages (so, when $E(X_2)$ is close to $E(X_1)$) that gap times are hard to analyze, because they are primarily the consequence of the setup and processing time variability in the system. In synchronized settings, the lower bound $E[G_2]_{LB}$ will perform particularly poor, as it completely fails to take into account the impact of these variabilities. This is illustrated in the top pane of figure 5, which shows the average value of $E(G_2)_{sim}$ in terms of $\frac{E(X_2)}{E(X_1)}$ for the experiment, along with the average value of $E(G_2)_{LB}$.

When $E(X_2)$ is either very high or very low compared to $E(X_1)$, the performance of $E(G_2)_{LB}$ is rather good, as the occurrence of a gap time in these settings is largely determined by the difference in the mean processing times. However, $E(G_2)_{LB}$ obviously falls to zero as soon as $E(X_2)=E(X_1)$, while in reality, $E(G_2)_{sim}$ continues to gradually decrease as $\frac{E(X_2)}{E(X_1)}$ increases, approaching zero but never actually reaching zero. Hence, the lower bound for $E(G_2)$ seriously underestimates the simulated $E(G_2)$ for $\frac{E(X_2)}{E(X_1)}$ close to 1.
The structure of the approximation however ensures that $E(G_2)_{app}$ remains positive and only gradually drops to zero as $E(X_2)$ becomes significantly larger than $E(X_1)$. Indeed, the bottom pane of figure 5 shows that $E(G_2)_{app}$ follows the $E(G_2)_{sim}$ very closely.
Hence, while the approximation clearly outperforms the lower bound in terms of both average and standard deviation of ε (as revealed by the results in Table 2), the relative improvement of the approximation over the lower bound is most pronounced in settings where the average processing rates of both stages are of the same order of magnitude (so, when $E(X_2)$ is close to $E(X_1)$).

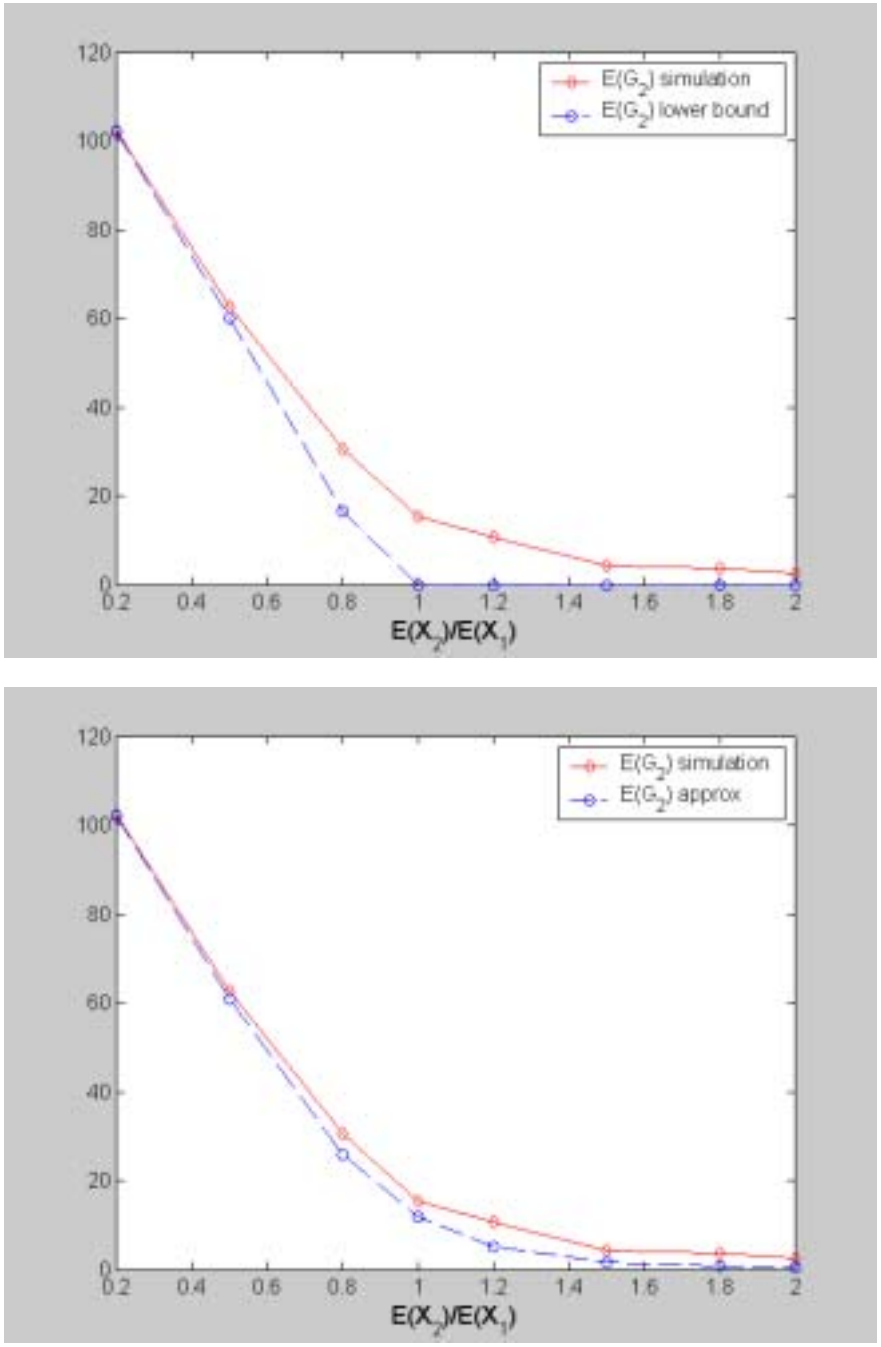

Figure 5: Comparison of the average value of $E[G_2]_{sim}$ versus the average value of $E[G_2]_{LB}$ (top pane), and versus the average value of $E[G_2]_{app}$

6. Example

Let's consider a two-stage flowshop where $N = 30$ and $SU_1 = 15$ (deterministic). The unit processing times on both stages are gamma distributed ($E(x_1) = 10$, $E(x_2) = 8$, $c^2_{x_1} = c^2_{x_2} = 0.9$), as is the setup time at stage 2 ($E(SU_2) = 0.5$, $c^2_{SU_2} = 0.9$). As the process batch size equals 30 product units, the number of transfer batches in this setting may be equal to $T = 1$ (in this case, no lot splitting is used), $T = 2$, $T = 3$, $T = 5$, $T = 6$, $T = 10$, $T = 15$ or $T = 30$ (in this case, the process batch is split in the maximum number of sublots).

Figure 6 shows the average process batch makespan on stage 2 for this setting, in terms of T. The approximation is clearly very close to the simulated value, at all values of T. An overview of the actual values and relative errors is given in Table 3.

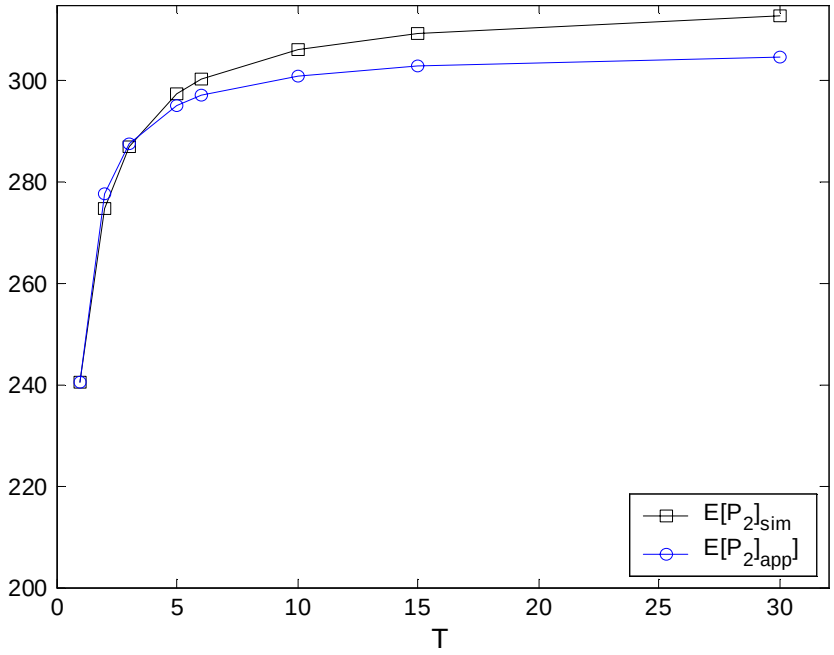

Figure 6: $E[P_2]_{app}$ versus $E[P_2]_{sim}$ for the example, in terms of T

Interestingly, the figure shows that the average process batch makespan on stage 2 increases as the process batch is split into a larger number of sublots. In fact, this will be the case whenever $E[x_1] > E[x_2]$: in that case, the average gap time on stage 2 will increase in T (for completeness, a proof is provided in Appendix 2). This does not mean, however, that lot splitting should be avoided. Rather, it points towards a trade-off between the well-known

T	$E[P_2]_{sim}$	$E[P_2]_{app}$	ε
30	312.7170	304.8295	-2.52%
15	309.3135	302.9026	-2.07%
10	306.1598	300.9744	-1.69%
6	300.2889	297.1130	-1.06%
5	297.4579	295.1793	-0.77%
3	286.8907	287.4147	0.18%
2	274.7345	277.5953	1.04%
1	240.5100	240.5000	0.00%

Table 3: Performance of $E[P_2]_{app}$ compared to $E[P_2]_{sim}$, for the example

improvement in flow time that can be obtained (which is due to the mechanism of overlapping operations, as described in section 1), and the gap time incurred on stage 2. As the occurrence of gap times may imply a cost for the system, this observation emphasizes the importance of taking them into account when assessing the desirability of lot splitting in a stochastic setting.

7. Conclusions

In this paper, we have analyzed the occurrence of gap times in a two-stage stochastic flowshop with lot splitting. We have been able to derive a number of basic insights into the behavior of gap times, in terms of the system's characteristics (e.g. the difference in processing rates, the variability present in setup and processing times at the different stages, and the lot splitting policy used). Based upon these insights, we have developed a suitable approximation for the average process batch makespan on the second stage. This approximation explicitly takes into account the impact of setup and processing time variability on the occurrence of gap times, and hence outperforms the lower bound.

Though the occurrence of gap times is a major issue in stochastic settings, this paper provides (to the best of our knowledge) the first analysis of their behavior. In the future, we would like to extend the current model towards settings with more than one product type. Our results have also pointed towards a trade-off between the improvement in flow time obtained by using a lot splitting policy, and the occurrence of gap times caused by this policy. Hence, our interest also goes towards the development of an analytical model for estimating flow times in terms of the lot splitting policy, as this would enable us to include this trade-off in

cost-benefit analyses.

Appendix 1

Without loss of generality, we prove this equality for $T = 5$. Note that V_i in expression (3) equals:

$$V_i = \sum_{k=2}^{i} X_{1,k} - (W_2 + SU_2 + \sum_{j=1}^{i-1}[X_{2,j} + g_2(j)])$$

such that we can write:

$$\begin{aligned} G_2 &= \sum_{i=2}^{T} \max[V_i, 0] \\ &= \underbrace{\max[Z(2), 0]}_{g_2(2)} + \underbrace{\max[Z(3) - g_2(2), 0]}_{g_2(3)} + \underbrace{\max[Z(4) - g_2(2) - g_2(3), 0]}_{g_2(4)} \\ &\quad \underbrace{+\max[Z(5) - \sum_{t=2}^{4} g_2(t), 0]}_{g_2(5)} \end{aligned}$$

Moreover, we know that, for all random variables X, A and B:

$$\max[X + A, X + B] = X + \max[A, B]$$

By repeatedly applying this property, we get:

$$\begin{aligned} g_2(2) + g_2(3) &= g_2(2) + [-g_2(2) + \max[Z(3), g_2(2)]] \\ &= \max[Z(3), \max[Z(2), 0]] \\ g_2(2) + g_2(3) + g_2(4) &= g_2(2) + g_2(3) + [-g_2(2) - g_2(3) + \max[Z(4), g_2(2) + g_2(3)]] \\ &= \max[Z(4), \max[Z(3), \max[Z(2), 0]]] \\ g_2(2) + g_2(3) + g_2(4) + g_2(5) &= \max[Z(5), \max[Z(4), \max[Z(3), \max[Z(2), 0]]]] \end{aligned}$$

Rewriting the last line, we get:

$$\begin{aligned} G_2 &= g_2(2) + g_2(3) + g_2(4) + g_2(5) \\ &= \max[Z(5), Z(4), Z(3), Z(2), 0] \end{aligned}$$

which is what we had to prove.

Appendix 2

From expression (8), we know that $E(G_2)$ can be approximated by:

$$E(G_2) = \frac{m_V}{2} + \frac{Exp\{\frac{-m_V^2}{2\sigma_V^2}\} * \sigma_V^2}{\sqrt{2\pi}\sigma_V} + \frac{m_V}{2} * Erf[\frac{m_V}{\sqrt{2}\sigma_V}]$$

with m_V and σ_V^2 given by:

$$\begin{aligned} m_V &= (T-1) * \frac{N}{T} * [E(x_1) - E(x_2)] - E(SU_2) \\ \sigma_V^2 &= (T-1) * \frac{N}{T} * [Var(x_1) + Var(x_2)] + Var(SU_2) \end{aligned}$$

Assuming that $E(G_2)$ is continuous in T, the derivative of $E(G_2)$ towards T can be calculated as (by means of the software package Mathematica):

$$\begin{aligned} \frac{\partial E(G_2)}{\partial T} = \frac{N}{4T^2}[&\underbrace{Exp\{\frac{-m_V^2}{2\sigma_V^2}\}\sqrt{\frac{2}{\pi}}\frac{Var(x_1)+Var(x_2)}{\sqrt{\frac{(T-1)N(Var(x_1)+Var(x_2))}{T} + Var(SU_2)}}}_{>0} \\ &+ 2(E(x_1) - E(x_2))\underbrace{(1 + Erf[\frac{m_V}{\sqrt{2}\sigma_V}])}_{\geq 0}] \end{aligned} \quad (11)$$

As the error function $Erf[z]$ can only vary between -1 and +1, $(1 + Erf[z])$ is always larger than or equal to 0, independent of z. Hence, we can conclude from expression (11) that:

$$E(x_1) - E(x_2) > 0 \Longrightarrow \frac{\partial E(G_2)}{\partial T} > 0$$

In other words, when the average unit processing time on stage 1 exceeds the average unit processing time on stage 2, the average gap time on stage 2 increases steadily when the process batch is split in a larger number of transfer batches. On the other hand, when $E(x_1) - E(x_2) < 0$, it may happen that $\frac{\partial E(G_2)}{\partial T}$ becomes negative. In that case, the average gap time on stage 2 will decrease in terms of T.

Acknowledgments

This research was supported by the Research Foundation-Flanders (FWO-Vlaanderen). Ms Van Nieuwenhuyse is a Postdoctoral Fellow of the Research Foundation-Flanders.

References

Baker, K.R., D. Jia. 1993. A comparative study of lot streaming techniques. *Omega* **21** 561–566.

Benjaafar, S. 1996. On production batches, transfer batches, and lead times. *IIE Transactions* **28** 357–362.

Blumenfeld, D. 2001. *Operations research calculations handbook*. CRC Press, Boca Raton, FL.

Bozer, Y.A., J. Kim. 1996. Determining transfer batch sizes in trip-based material handling systems. *The International Journal of Flexible Manufacturing Systems* **8** 313–356.

Bukchin, J., M. Tzur, M. Jaffe. 2002. Lot splitting to minimize average flow time in a two-machine flow shop. *IIE Transactions* **34** 953–970.

Bukchin, J., M. Masin. 2004. Multi-objective lot splitting for a single product m-machine flowshop line. *IIE Transactions* **36** 191–202.

Chen, J., G. Steiner. 1996. Lot streaming with detached setups in three-machine flowshops. *European Journal of Operational Research* **26** 591–611.

Chen, J., G. Steiner. 1998. Lot streaming with attached setups in three-machine flowshops. *IIE Transactions* **30** 1075–1084.

Cheng, T.C.E., B.M.T. Lin, A. Toker. 2000. Makespan minimization in the two-machine flowshop batch scheduling problem. *Naval Research Logistics* **47** 128–144.

Goldratt, E.M., J. Cox. 1984. *The Goal: a process of ongoing improvement*. North River Press, New York.

Graves, S.C., M.M. Kostreva. 1986. Overlapping operations in materials requirements planning. *Journal of Operations Management* **6** 283–294.

Hopp, W.J., M.L. Spearman, D.L. Woodruff. 1990. Practical strategies for lead time reduction. *Manufacturing Review* **3** 78–84.

Hopp, W. J., M.L. Spearman. 2000. *Factory Physics: Foundations of Manufacturing Management*. Irwin/McGraw-Hill, New York.

Jacobs, F.R., D.J. Bragg. 1988. Repetitive lots: flow-time reductions through sequencing and dynamic batch sizing. *Decision Sciences* **19** 281–294.

Kalir, A.A., S.C. Sarin. 2001. Optimal solutions for the single batch flow-shop lot streaming problem with equal sublots. *Decision Sciences* **32** 387–397.

Kropp, D.H., T.L. Smunt. 1990. Optimal and heuristic models for lot splitting in a flow shop. *Decision Sciences* **21** 691–709.

Litchfield, J.L., R. Narasimhan. 2000. Improving job shop performance through process queue management under transfer batching. *Production and Operations Management* **9** 336–348.

Potts, C.N., M.Y. Kovalyov. 2000. Scheduling with batching: a review. *European Journal of Operational Research* **120** 228–249.

Ramasesh, R.V., H. Fu, D. K. H. Fong, J. C. Hayya. 2000. Lot streaming in multistage production systems. *International Journal of Production Economics* **66** 199–211.

Ross, S.M. 1996. *Stochastic Processes*. John Wiley and Sons Inc., New York.

Ruben, R.A., F. Mahmoodi. 1998. Lot splitting in unbalanced production systems. *Decision Sciences* **29** 921–949.

Santos, C., M. Magazine. 1985. Batching in single-operations manufacturing systems. *Operations Research Letters* **4** 99–103.

Sen, A., E. Topaloglu, O.S. Benli. 1998. Optimal streaming of a single job in a two-stage flow shop. *European of Operational Research* **110** 42–62.

Smunt, T.L., A.H. Buss, D.H. Kropp. 1996. Lot Splitting in stochastic flow shop and job shop environments. *Decision Sciences* **27** 215–238.

Suri, R. 1998. *Quick Response Manufacturing: a companywide approach to reducing lead times*. Productivity Press, Portland, OR.

Umble, M., M.L. Srikanth. 1995. *Synchronous Manufacturing: principles for world-class excellence*. The Spectrum Publishing Company.

Van Nieuwenhuyse, I. 2004. Lot splitting in single-product flowshops: issues of delivery reliability, production disruptions and flow times. PhD thesis, Department of Applied Economics, University of Antwerp, Antwerp, Belgium.

Van Nieuwenhuyse, I., N. Vandaele. 2004. Determining the optimal number of sublots in a single-product, deterministic flow shop with overlapping operations. *International Journal of Production Economics* **92** 221–239.

Wagner, B.J., G.L. Ragatz. 1994. The impact of lot splitting on due date performance. *Journal of Operations Management* **12** 13–25.