

## Research Article

# Performance Analysis of a Kitting Process as a Paired Queue

**E. De Cuypere, K. De Turck, and D. Fiems**

*Department of Telecommunications and Information Processing, Ghent University, St. Pietersnieuwstraat 41, 9000 Gent, Belgium*

Correspondence should be addressed to E. De Cuypere; [eline.decuypere@telin.ugent.be](mailto:eline.decuypere@telin.ugent.be)

Received 19 February 2013; Revised 14 May 2013; Accepted 14 May 2013

Academic Editor: Zhengguang Wu

Copyright © 2013 E. De Cuypere et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nowadays, customers request more variation in a company's product assortment leading to an increased amount of parts moving around on the shop floor. To cope with this tendency, a kitting process can be implemented. Kitting is the operation of collecting the necessary parts for a given end product in a specific container, called a kit, prior to arriving at an assembly unit. As kitting performance is critical to the overall cost and performance of the manufacturing system, this paper analyses a two-part kitting process as a Markovian model. In particular, kitting is studied as a paired queue, thereby accounting for stochastic part arrivals, and kit assembly times. Using sparse matrix techniques, we assess the impact of kitting interruptions, bursty part arrivals and phase-type distributed kit assembly times on the behaviour of the part buffers. Finally, a cost-profit analysis of kitting processes is conducted and an approximation for a two-part kitting process is established.

## 1. Introduction

Nowadays manufacturing systems are often composed of multiple in-house fabrication units [1]. The semifinished products stemming from these units are the input materials for other fabrication units or for assembly lines. Hence, efficient transport of materials between the different stages of the production process is key for overall production cost minimisation. Kitting is a particular strategy for supplying materials to an assembly line. Instead of delivering in containers, each holding a single type part and all holding the same number of parts, kitting collects the necessary set of parts for an individual end product in a specific container, referred to as kit, prior to arriving at an assembly unit [1–6].

Kitting mitigates storage space requirements at the assembly station since no part inventories need to be kept there. Moreover, parts are placed in proper positions in the container such that assembly time reductions can be realised. Additional benefits include reduced learning time of the workers at the assembly stations and increased quality of the product. Although kitting is a nonvalue adding activity, its application can reduce the overall materials handling time [6]. Indeed activities such as selecting and gripping parts are performed more efficiently. Furthermore, the whole operator walking time is drastically reduced or even eliminated since

kits, each containing a complete set of components, are brought to the assembly station [7]. The advantages mentioned above do not come for free since the kitting operation itself incurs additional costs such as the time and effort for planning the allocation of the parts into kits and the kit preparation itself. Moreover, the introduction of a kitting operation in a production process involves a major investment and the effect on efficiency is uncertain. Therefore, it is important to analyse the performance of kitting in a production environment prior to its actual introduction. This is the subject of the present paper.

In the literature, most authors consider a kitting process as a queueing system with stochastic part arrivals and kit assembly times. Hopp and Simon [8] developed a model for a kitting process with exponentially distributed processing times for kits and Poisson arrivals. They found accurate bounds for the required buffer capacity of kitting processes with two parts. Explicitly accounting for finite buffer capacities, Som et al. [3] further refined the results of Hopp and Simon.

Of course real buffers always have a finite capacity, the capacity being constrained by the storage room. However, if the capacity is large enough, we can have a good approximation of a process with a finite capacity on the basis of a model with unlimited capacity. This means that there is

always enough space for upcoming parts, which simplifies the analysis. Unfortunately, the assumption of an infinite buffer is not valid for kitting processes. If the capacity is assumed to be infinite, then the model will degrade to an unstable stochastic system. Harrison [9] showed for a multiple input generalisation of the GI/G/1 queue that it is necessary to impose a restriction on the size of the buffer to ensure stability in the operations of a kitting process. Under this assumption, the probability to have a certain long-term stock position is equal and independent of the current stock position. This was also demonstrated by Latouche [10] who studied waiting lines with paired customers. We can consider this analysis as an abstraction of a kitting process with two types of parts.

In this work, we focus on a kitting process modulated by a Markovian environment. The introduction of this environment allows us to study kitting under more realistic stochastic assumptions: kitting interruptions, bursty part arrivals, phase-type distributed kit assembly times, and so forth. Our paper extends the results on kitting in a Markovian environment [11].

The remainder of this paper is organised as follows. Section 2 describes the kitting process at hand. In Section 3, Chapman-Kolmogorov equations are derived and their numerical solution is discussed. To illustrate our approach, Section 4 considers a number of numerical examples. In particular, we assess the impact of kitting interruptions, bursty part arrivals, and phase-type distributed kit assembly times on the behaviour of the part buffers. Then, a cost-profit analysis of kitting processes is conducted and an approximation for a two-part kitting process is established. Finally, conclusions are drawn in Section 5.

## 2. Model Description

In this paper, we study a two-queue kitting process, as depicted in Figure 1. Each queue has a finite capacity—let  $C_\ell$  denote the capacity of buffer  $\ell$ ,  $\ell = 1, 2$ —and models the inventory of parts of a single type. New parts arrive at the buffers and if both buffers are nonempty, a kit is assembled by collecting a part from each buffer. Hence, departures from the buffers are synchronised, and the buffers are paired. Operation of part buffers therefore considerably differs from other queueing systems.

Arrivals at both buffers are modelled by a Markovian arrival process and kit assembly is not instantaneous. For ease of modelling, it is assumed that there is a modulating Markov process, arrival and service rates depending on the state of this process. To be more precise, the kitting process is modelled as a continuous-time Markov process with state space  $\mathcal{C}_1 \times \mathcal{C}_2 \times \mathcal{K}$ , whereby  $\mathcal{C}_\ell = \{0, \dots, C_\ell\}$  for  $\ell = 1, 2$  and with  $\mathcal{K} = \{1, 2, \dots, K\}$  being the state space of the modulating process. At any time, the state of the kitting process is described by the triplet  $[m, n, i]$ , with  $m$  and  $n$  being the number of parts in the first and second buffer, respectively, and  $i$  being the state of the modulating process. We now describe the state transitions.

- (i) The state of the modulating process can change when there are neither arrivals nor departures. Let

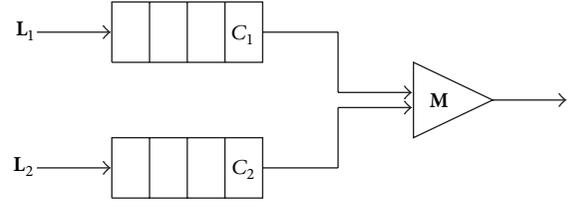


FIGURE 1: Kitting process: the buffers are on the left, and the triangularly shaped kitting process is on the right.

$\alpha_{ij}$  denote the transition rate from state  $i$  to state  $j$  ( $i, j \in \mathcal{K}, i \neq j$ ) and let  $\mathbf{A}$  denote the corresponding generator matrix.

- (ii) The state of the modulating process may remain the same or may change when there is an arrival. Let  $\lambda_{ij}^{(\ell)}$  denote the (marked) transition rate from state  $i$  to state  $j$  when there is an arrival at buffer  $\ell$ ,  $\ell = 1, 2$ . Moreover, let  $\mathbf{L}_\ell$  denote the corresponding generator matrix. Note that such marked transitions from state  $i$  to state  $i$  are allowed.
- (iii) Analogously, the state of the modulating process may remain the same or may change when there is a departure (in each buffer). Let  $\mu_{ij}$  and  $\mathbf{M}$  denote the corresponding transition rate and generator matrix, respectively.

Summarising, arrivals at and departures from the buffers are described by the generator matrices  $\mathbf{A}$ ,  $\mathbf{L}_1$ ,  $\mathbf{L}_2$ , and  $\mathbf{M}$ . So far, no diagonal elements of  $\mathbf{A}$  have been defined. To simplify notation, it will be further assumed that the diagonal elements are chosen such that the row sums of  $\mathbf{A} + \mathbf{L}_1 + \mathbf{L}_2 + \mathbf{M}$  are zero.

The computational method employed here does not require any homogeneity of the generator matrices. When required by the applications at hand, intensities may depend on the buffer content. In this case, we introduce superscripts to make this dependence explicit. For example,  $\mathbf{M}^{(m,n)}$  denotes the generator matrix of state transitions with departure when there are  $m$  parts in buffer 1 and  $n$  parts in buffer 2. In addition, we use arguments for rates as we already used superscripts and subscripts to distinguish the arrival rates at the different queues. For example,  $\lambda_{ij}^{(\ell)}(m, n)$  denotes the arrival rate at buffer  $\ell = 1, 2$  from state  $i$  to state  $j$  when there are  $m$  parts in buffer 1 and  $n$  parts in buffer 2.

*Example 1.* In the most basic setting, parts arrive at the buffers in accordance to an independent Poisson processes with rate  $\lambda_1$  and  $\lambda_2$  and kit assembly times are exponentially distributed with parameter  $\mu$ . In this case, there is no need to have a modulating Markov process, and the state is completely described by the number of parts in each buffer,  $(m, n)$ . We have

$$\begin{aligned} \mathbf{M} &= [\mu], & \mathbf{L}_1 &= [\lambda_1], \\ \mathbf{L}_2 &= [\lambda_2], & \mathbf{A} &= [-\lambda_1 - \lambda_2 - \mu]. \end{aligned} \tag{1}$$

*Example 2.* To account for burstiness in the arrival process of the parts at the different buffers, the modulating process

allows the mitigation of the Poissonian arrival assumptions. We can replace the Poisson processes by a two-class Markovian arrival processes. Multiclass Markovian arrival processes allow for intricate correlation and can be efficiently characterised from trace data [12, 13]. As we have two types of arrivals, the Markovian arrival process is described by the generator matrix  $\Lambda_1$  of transitions with arrivals at buffer 1, the generator matrix  $\Lambda_2$  with arrivals at buffer 2, and the generator matrix  $\Lambda_0$  without arrivals. As usual, the diagonal elements of  $\Lambda_0$  are negative and ensure that the row sums of  $\Lambda_0 + \Lambda_1 + \Lambda_2$  are zero. Retaining exponentially distributed kit assembly times, we have

$$\begin{aligned} \mathbf{L}_1 &= \Lambda_1, & \mathbf{L}_2 &= \Lambda_2, \\ \mathbf{A} &= \Lambda_0 - \mu\mathbf{I}, & \mathbf{M} &= \mu\mathbf{I}. \end{aligned} \quad (2)$$

Here  $\mathbf{I}$  denotes the identity matrix.

*Example 3.* As for the arrival processes, the model at hand is sufficiently flexible to include phase-type distributed kit assembly times. The phase-type distribution is completely characterised by an initial probability vector  $\boldsymbol{\tau}$  and the matrix  $\mathbf{T}$  which corresponds to non-absorbing transitions [14]. Let  $\mathbf{t}' = -\mathbf{T}\mathbf{e}'$  be the column vector with the rates to the absorbing state and let  $\mathbf{f}$  be a row vector with zero elements except the first one. Assuming Poisson arrivals in both buffers (with rates  $\lambda_1$  and  $\lambda_2$ , resp.), we get the following matrices:

$$\begin{aligned} \mathbf{L}_1^{(m,n)} &= \lambda_1 \mathbf{I} (1 - \mathbb{1}_{\{m=0, n>0\}}) + \lambda_1 \mathbf{e}' \boldsymbol{\tau} \mathbb{1}_{\{m=0, n>0\}}, \\ \mathbf{L}_2^{(m,n)} &= \lambda_2 \mathbf{I} (1 - \mathbb{1}_{\{m>0, n=0\}}) + \lambda_2 \mathbf{e}' \boldsymbol{\tau} \mathbb{1}_{\{m>0, n=0\}}, \\ \mathbf{A}^{(m,n)} &= \mathbf{T} \mathbb{1}_{\{m>0, n>0\}} - \lambda_1 \mathbf{I} - \lambda_2 \mathbf{I}, \\ \mathbf{M}^{(m,n)} &= \mathbf{t}' \boldsymbol{\tau} \mathbb{1}_{\{m>1, n>1\}} + \mathbf{t}' \mathbf{f} (1 - \mathbb{1}_{\{m>1, n>1\}}). \end{aligned} \quad (3)$$

Here, it is assumed that the background state equals 1 if one of the buffers is empty. When service starts again, the background state is chosen in accordance with the probability vector  $\boldsymbol{\tau}$ .

### 3. Analysis

Having established the modelling assumptions and settled our notation, we now focus on the analysis of the kitting process.

*3.1. Balance Equations.* We aim to define a set of equations for the steady-state probability vector for the Markov process  $[Q_1(t), Q_2(t), S(t)]$ ,  $Q_\ell(t)$  being the number of parts in buffer  $\ell$  at time  $t$  and  $S(t)$  being the state of the background process at time  $t$ .

Let  $\pi_i(m, n) = \lim_{t \rightarrow \infty} \Pr[Q_1(t) = m, Q_2(t) = n, S(t) = i]$  be the steady-state probability to be in state  $[m, n, i]$  and let  $\boldsymbol{\pi}(m, n)$  be the vector with elements  $\pi_i(m, n)$ , for  $i \in \mathcal{X}$ . Figure 2 shows a fragment of the transition rate diagram of the kitting model in state  $[m, n, i]$ . As mentioned above, two independent input streams arrive at the buffers with intensity  $\lambda_{ij}^{(\ell)}$  and are processed into kits with intensity  $\mu_{ij}$ . Upon

completion of a kit, the content of both buffers is decreased by 1. Note that we only show the transitions whereby the modulating Markov process remains in state  $i$ . Moreover, possible dependence of the transition rates on the buffer sizes is not indicated.

Based on the transition rate diagram, we now derive the balance equations of the kitting process at hand.

- (i) First, consider the case where both buffers are neither empty nor full,  $0 < n < C_1$  and  $0 < m < C_2$ . We have

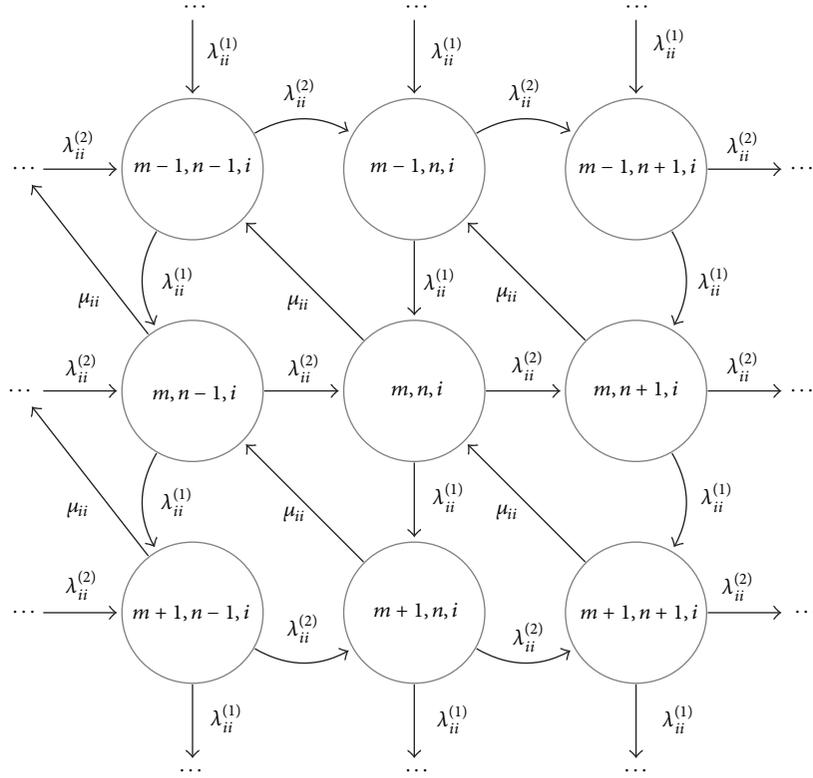
$$\begin{aligned} \pi_i(m, n) & \left( \sum_{j=1}^K \lambda_{ij}^{(1)}(m, n) + \sum_{j=1}^K \lambda_{ij}^{(2)}(m, n) \right. \\ & \left. + \mu_{ij}(m, n) + \sum_{j=1, j \neq i}^K \alpha_{ij}(m, n) \right) \\ & = \sum_{j=1}^K \pi_j(m-1, n) \lambda_{ji}^{(1)}(m-1, n) \\ & \quad + \sum_{j=1}^K \pi_j(m, n-1) \lambda_{ji}^{(2)}(m, n-1) \\ & \quad + \sum_{j=1}^K \pi_j(m+1, n+1) \mu_{ji}(m+1, n+1) \\ & \quad + \sum_{j=1, j \neq i}^K \pi_j(\mathbf{m}, \mathbf{n}) \alpha_{ij}(m, n), \end{aligned} \quad (4)$$

for  $i, j \in \mathcal{X}$  or equivalently,

$$\begin{aligned} \boldsymbol{\pi}(m-1, n) \mathbf{L}_1^{(m-1, n)} + \boldsymbol{\pi}(m, n-1) \mathbf{L}_2^{(m, n-1)} \\ + \boldsymbol{\pi}(m+1, n+1) \mathbf{M}^{(m+1, n+1)} + \boldsymbol{\pi}(m, n) \mathbf{A}^{(m, n)} = \mathbf{0}. \end{aligned} \quad (5)$$

- (ii) If buffer 1 is empty and buffer 2 is neither empty nor full ( $m = 0$  and  $0 < n < C_2$ ), we have

$$\begin{aligned} \pi_i(0, n) & \left( \sum_{j=1}^K \lambda_{ij}^{(1)}(0, n) + \sum_{j=1}^K \lambda_{ij}^{(2)}(0, n) \right. \\ & \left. + \sum_{j=1, j \neq i}^K \alpha_{ij}(0, n) \right) \\ & = \sum_{j=1}^K \pi_j(0, n-1) \lambda_{ji}^{(2)}(0, n-1) \\ & \quad + \sum_{j=1}^K \pi_j(1, n+1) \mu_{ji}(1, n+1) \\ & \quad + \sum_{j=1, j \neq i}^K \pi_j(0, n) \alpha_{ij}(0, n), \end{aligned} \quad (6)$$

FIGURE 2: Fragment of the transition rate diagram for state  $(m, n, i)$ .

for  $i, j \in \mathcal{N}$  or equivalently,

$$\begin{aligned} \boldsymbol{\pi}(0, n-1) \mathbf{L}_2^{(0, n-1)} + \boldsymbol{\pi}(1, n+1) \mathbf{M}^{(1, n+1)} \\ + \boldsymbol{\pi}(0, n) \left( \mathbf{A}^{(0, n)} + \text{diag}(\mathbf{M}^{(0, n)} \mathbf{e}') \right) = \mathbf{0}. \end{aligned} \quad (7)$$

(iii) Similarly, if buffer 2 is empty and buffer 1 is neither empty nor full ( $n = 0$  and  $0 < m < C_1$ ), we have

$$\begin{aligned} \boldsymbol{\pi}(m-1, 0) \mathbf{L}_1^{(m-1, 0)} + \boldsymbol{\pi}(m+1, 1) \mathbf{M}^{(m+1, 1)} \\ + \boldsymbol{\pi}(m, 0) \left( \mathbf{A}^{(m, 0)} + \text{diag}(\mathbf{M}^{(m, 0)} \mathbf{e}') \right) = \mathbf{0}. \end{aligned} \quad (8)$$

(iv) If both buffers are empty ( $m = 0$  and  $n = 0$ ), we have

$$\begin{aligned} \boldsymbol{\pi}(1, 1) \mathbf{M}^{(1, 1)} \\ + \boldsymbol{\pi}(0, 0) \left( \mathbf{A}^{(0, 0)} + \text{diag}(\mathbf{M}^{(0, 0)} \mathbf{e}') \right) = \mathbf{0}. \end{aligned} \quad (9)$$

(v) If buffer 1 is empty and buffer 2 is full ( $m = 0$  and  $n = C_2$ ), we get

$$\begin{aligned} \boldsymbol{\pi}(0, C_2-1) \mathbf{L}_2^{(0, C_2-1)} \\ + \boldsymbol{\pi}(0, C_2) \left( \mathbf{A}^{(0, C_2)} + \text{diag}(\mathbf{M}^{(0, C_2)} \mathbf{e}' + \mathbf{L}_2^{(0, C_2)} \mathbf{e}') \right) = \mathbf{0}. \end{aligned} \quad (10)$$

(vi) Similarly, if buffer 1 is full and buffer 2 is empty ( $m = C_1$  and  $n = 0$ ), we have

$$\begin{aligned} \boldsymbol{\pi}(C_1-1, 0) \mathbf{L}_1^{(C_1-1, 0)} \\ + \boldsymbol{\pi}(C_1, 0) \left( \mathbf{A}^{(C_1, 0)} + \text{diag}(\mathbf{M}^{(C_1, 0)} \mathbf{e}' + \mathbf{L}_1^{(C_1, 0)} \mathbf{e}') \right) = \mathbf{0}. \end{aligned} \quad (11)$$

(vii) Finally, if both buffers are full ( $m = C_1$  and  $n = C_2$ ), we find

$$\begin{aligned} \boldsymbol{\pi}(C_1-1, C_2) \mathbf{L}_1^{(C_1-1, C_2)} + \boldsymbol{\pi}(C_1, C_2-1) \mathbf{L}_2^{(C_1, C_2-1)} \\ + \boldsymbol{\pi}(C_1, C_2) \left( \mathbf{A}^{(C_1, C_2)} + \text{diag}(\mathbf{L}_1^{(C_1, C_2)} \mathbf{e}' + \mathbf{L}_2^{(C_1, C_2)} \mathbf{e}') \right) = \mathbf{0}. \end{aligned} \quad (12)$$

**3.2. Performance Measures.** Given the steady-state vectors  $\boldsymbol{\pi}(m, n)$ , we can now obtain a number of interesting performance measures for the kitting system. For ease of notation, let  $\pi(m, n) = \boldsymbol{\pi}(m, n) \mathbf{e}'$  denote the probability to have  $m$  parts in buffer 1 and  $n$  parts in buffer 2. Moreover, let  $\boldsymbol{\pi}^{(1)}(m) = \sum_n \boldsymbol{\pi}(m, n)$  and  $\boldsymbol{\pi}^{(2)}(n) = \sum_m \boldsymbol{\pi}(m, n)$  denote the marginal probability vectors of the buffer 1 and 2 content, respectively, for the different background states. Finally, the probability mass functions of the buffer contents equal  $\pi^{(1)}(m) = \boldsymbol{\pi}^{(1)}(m) \mathbf{e}'$  and  $\pi^{(2)}(n) = \boldsymbol{\pi}^{(2)}(n) \mathbf{e}'$ .

The following performance measures are of interest.

- (i) The mean buffer 1 and 2 content is  $EQ_1$  and  $EQ_2$ , respectively,

$$EQ_1 = \sum_m^{C_1} \pi^{(1)}(m) m, \quad EQ_2 = \sum_n^{C_2} \pi^{(2)}(n) n. \quad (13)$$

- (ii) The variance of the buffer 1 and 2 is  $\text{Var } Q_1$  and  $\text{Var } Q_2$ , respectively,

$$\begin{aligned} \text{Var } Q_1 &= \sum_m^{C_1} \pi^{(1)}(m) m^2 - (EQ_1)^2, \\ \text{Var } Q_2 &= \sum_n^{C_2} \pi^{(2)}(n) n^2 - (EQ_2)^2. \end{aligned} \quad (14)$$

- (iii) The effective load of the system  $\rho_{\text{eff}}$  is the amount of time that kitting is ongoing. As kitting is only ongoing when none of the buffers is empty, we have

$$\rho_{\text{eff}} = 1 - \pi^{(1)}(0) - \pi^{(2)}(0) + \pi(0, 0). \quad (15)$$

- (iv) Let the throughput  $\eta$  be defined as the number of kits departing from the system per time unit. Taking into account all possible states from which we can have a departure, we find

$$\eta = \sum_{m=1}^{C_1} \sum_{n=1}^{C_2} \pi(m, n) \mathbf{M}^{(m,n)} \mathbf{e}^j. \quad (16)$$

- (v) The shortage probability  $K$  is the probability that one of the buffers is empty as

$$K = \pi^{(1)}(0) + \pi^{(2)}(0) - \pi(0, 0). \quad (17)$$

- (vi) The loss probability  $b_i$  in buffer  $i$  is the probability that an arriving part cannot be stored in buffer  $i$ ,  $i = 1, 2$ . By noting that the accepted arrival load equals the departure rate, we find

$$b_i = \frac{\rho_i - \eta}{\rho_i}, \quad (18)$$

where  $\rho_i$  is the arrival load of part  $i = \{1, 2\}$ . If the arrival load in both buffers is the same, then the loss probability is also the same:  $b_1 = b_2$ . If the arrival load is not the same, then the excess load in the most loaded buffer is lost as well.

**3.3. Methodology: Sparse Matrix Techniques.** Queueing models for kitting processes are rather complicated. Indeed, the modelled kitting process has a multidimensional state space. Even for relative moderate buffer capacity, the multidimensionality leads to huge state spaces; this is the so-called state-space explosion problem. For many queueing systems, infinite-buffer assumptions may mitigate this problem.

Given some buffer system with finite capacity, more efficient numerical routines can be constructed for the corresponding queueing system with infinite capacity. Unfortunately, as mentioned above, the infinite-buffer capacity assumption is not applicable for kitting processes and therefore cannot simplify the analysis. Recall that the infinite-capacity model is always unstable. For all input parameters except trivial ones (no arrivals), some or all of the queues grow unbounded with positive probability.

Consequently, the multidimensionality of the state space and the inapplicability of the infinite-buffer assumption yield Markov processes with a finite but very large state space. However, the number of possible state transitions from any specific state is limited. This means that most of the entries in the generator matrix are zero; the matrix is sparse. As illustrated by the numerical examples, using sparse matrices and their associated specialised algorithms result in manageable memory consumption and processing times, compared to standard algorithms.

The method used here to solve the sparse matrix equation is the projection method GMRES (generalized minimum residual). Details can be found in Stewart [15] (p. 194–197); Philippe et al. [16], Buchholz [17], and Saad and Schultz [18]. To solve the linear equation  $Ax = b$ , the GMRES algorithm approximates  $x$  by the vector  $x_n \in K_n$  in a Krylov subspace  $K_n = \text{span}\{b, Ab, \dots, A^{n-1}b\}$  which minimises the norm of the residual  $\|Ax_n - b\|$ . Since the residual norm is minimised at every step of the method, it is clear that it is nonincreasing. However, the work and storage requirement per iteration increases linearly with the iteration count. Hence, the cost of  $n$  iterations grows by  $O(n^2)$  which is a major drawback of GMRES. This limitation is usually overcome by restarting the algorithm. After a chosen number of iterations  $m$ , the accumulated data is cleared and the intermediate results are used as the initial data for the next  $m$  iterations [18].

To ensure fast convergence, it is also key to properly choose the initial vector that is passed on to the algorithm. We rely on MATLAB's build in GMRES algorithm and assume a uniform initial vector if no additional information about the solution is known. However, this is often the case. Calculation speed can be further improved as, in practice, performance measures are not calculated for an isolated set of parameters. For example, when a plot is created, a parameter is varied over a range of values. In this case, a previously calculated steady-state vector for some set of parameters can be used as a first estimate of the steady-state vector for a new "perturbed" set of parameters. Using previously calculated steady-state vectors is trivial if the state spaces corresponding to the parameter sets are equal. In this case, the previously calculated steady-state vector can be passed on unmodified. If the state space changes, the steady-state vector must be rescaled to the new state space. In general, adding zero-probability states if the state space increases or removing states if the state space decreases turns out to be ineffective. This is easily explained by a simple example. Assume that we increase the queue capacity of one of the part buffers. Typically, even for moderate load, a considerable amount of probability mass can be found for queue size equal to capacity. Increasing the

queue size and assigning zero probability to the new states is not a good estimate for the new steady-state vector. Also for the system with higher capacity, a considerable amount of probability mass can be found when the queue size equals the capacity (while zero probabilities were assigned).

#### 4. Numerical Results

With the balance equations at hand, we now illustrate our numerical approach by means of some examples.

*4.1. Bursty Part Arrivals.* As a first example, we quantify the impact of production inefficiency on the performance of a kitting process. To this end, we compare part buffers with Poisson arrivals to corresponding kitting systems with interrupted Poisson arrivals. The arrival interruptions account for inefficiency in the production process. Kit assembly times are assumed to be exponentially distributed with service rate equal to one, this value being independent of the number of parts in the different buffers. This is a kitting process with Markovian arrivals as described in Example 2 in Section 2.

The interrupted Poisson process considered here is a two-state Markovian process. In the active state, new parts arrive in accordance with a Poisson process with rate  $\lambda$  whereas no new parts arrive in the inactive state. Let  $\alpha$  and  $\beta$  denote the rate from the active to the inactive state and vice versa, respectively. We then use the following parameters to characterise the interrupted Poisson process:

$$\sigma = \frac{\beta}{\alpha + \beta}, \quad \kappa = \frac{1}{\alpha} + \frac{1}{\beta}, \quad \rho = \lambda\sigma. \quad (19)$$

Note that  $\sigma$  is the fraction of time that the interrupted Poisson process is active, the absolute time parameter  $\kappa$  is the average duration of an active and an inactive period, and  $\rho$  is the arrival load of the parts.

Figure 3 shows the mean number of parts in buffer 1 versus the arrival load, for various values of the buffer capacities  $C_1$  and  $C_2$  for Poisson arrivals (for both buffers) as well as for interrupted Poisson arrivals (again for both buffers). The arrival load is set to  $\rho = 0.8$  for all curves. In addition, we set  $\sigma = 0.4$  and  $\kappa = 10$  for the interrupted Poisson processes. Clearly, the mean buffer content increases as the arrival load increases as expected. Moreover, if more buffer capacity is available, it will also be used: the mean buffer content increases for increasing values of  $C_1 = C_2$ . Comparing interrupted Poisson and Poisson processes, burstiness in the production process has a negative impact on performance—more buffering is required—if the queues are not fully loaded ( $\rho < 1$ ). As for ordinary queues, the opposite can be observed for overloaded buffers. In this case, the effect of a large burst is mainly reflected in loss and not in additional queue content. Burstiness also yields larger periods without arrivals during which the buffer size decreases.

By numerical examples, we can quantify expected buffer behaviour—for example, more production yields higher buffer content, higher buffer capacity mitigates the loss probability, and so forth. However, less trivial behaviour can be observed as well. Figure 4 depicts the probability that

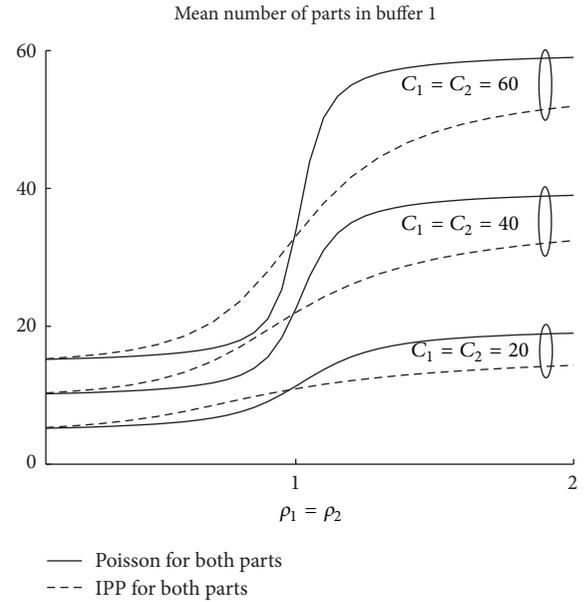


FIGURE 3: When the queues are not fully loaded, burstiness in the production process has a negative impact on the performance.

the buffer is full versus the buffer capacity  $C_1 = C_2$ . We compare performance of kitting with Poisson arrivals to kitting with interrupted Poisson arrivals at one buffer and at both buffers. As in the preceding figure, the interrupted Poisson processes are characterised by  $\sigma = 0.4$  and  $\kappa = 10$ . As expected, the probability that the buffer is full decreases for increasing values of the buffer capacities. Moreover, to reduce this probability, more buffer capacity is required for the case of two interrupted Poisson processes than for the case of two Poisson processes. For the kitting process with one Poisson and one interrupted Poisson process, non-trivial performance results can be observed. Namely, interruptions in the production of a part more negatively affect buffer performance of the other part. Indeed, buffer 1 is full with higher probability if the arrivals at buffer 2 are interrupted than if the arrivals at buffer 1 are interrupted. First note that the loss probabilities in both buffers are the same. For the Poisson buffer, parts are rejected at the arrival rate  $\rho_1$  whenever the buffer is full. For the IPP buffer, parts are rejected when the buffer is full and the arrival process is in the on-state at rate  $\rho_2/\sigma$ . As the loss probability in both buffers is the same, we have  $\Pr[Q_1 = C_1]\rho_1 = \Pr[Q_2 = C_2, S = 1]\rho_2/\sigma$ , or equivalently  $\Pr[Q_1 = C_1] = \Pr[Q_2 = C_2 | S = 1]$ . As the second queue is more likely to be full when there are arrivals we have  $\Pr[Q_2 = C_2 | S = 1] \geq \Pr[Q_2 = C_2]$  which shows  $\Pr[Q_1 = C_1] \geq \Pr[Q_2 = C_2]$ .

*4.2. Phase-Type Distributed Kit Assembly Times.* The second numerical example quantifies the impact of the distribution of the kit assembly times on kitting performance. In particular, we here study Erlang-distributed kit assembly times. Limiting ourselves to Poisson arrivals to both buffers, this numerical example fits Example 3 of Section 2.

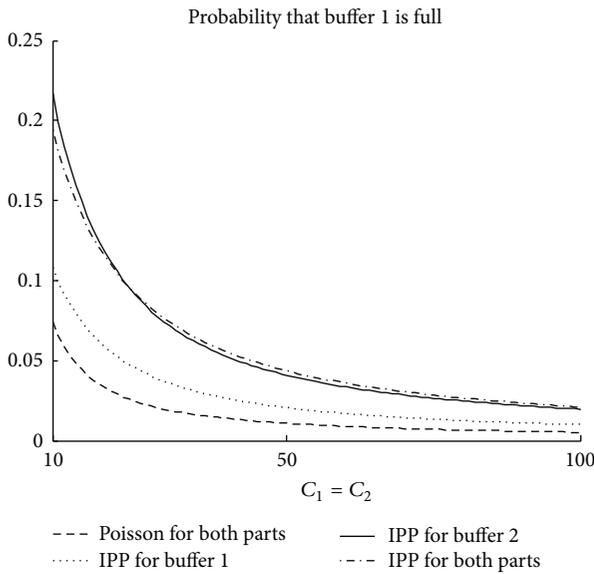


FIGURE 4: Irregularity in the production of a part leads to a higher probability to have a full buffer for the other part.

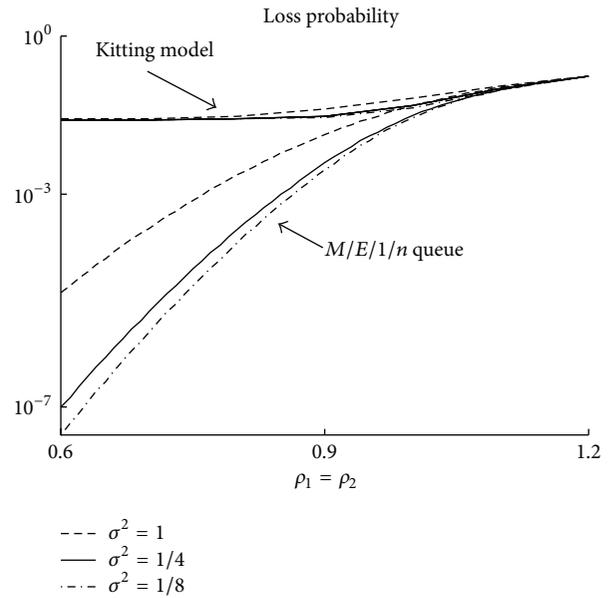


FIGURE 6: Variation of the kitting time distribution does not have a significant impact on the loss probability.

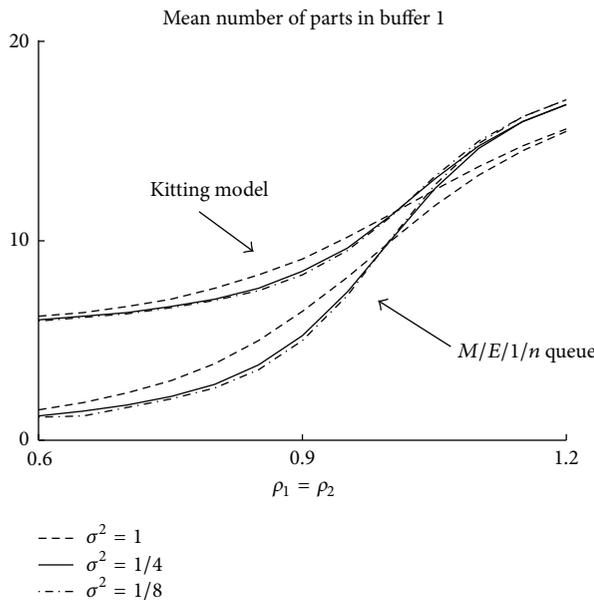


FIGURE 5: Variation of the kitting time distribution does not have a significant impact on the mean number of parts.

Figures 5 and 6 depict the mean number of parts in buffer 1 and the loss probability in buffer 1 for the kitting process and, as a reference point, for the  $M/E/1/N$  queue as well. In both figures, the arrival load is varied and different values of the variance of the kitting time distribution are assumed as indicated. The mean kitting time is equal to 1 for all curves and the capacity of both buffers is equal to 20. In underload ( $\rho < 1$ ), kitting performs worse than the  $M/E/1/N$  queue: the mean buffer content and the loss probability have a higher value. This follows from the fact that kitting stops when one of the buffers is empty. By increasing the load, it is obvious

that the buffer content converges to the capacity and the loss probability to one. It is most interesting to observe that the shape of the service time distribution only has a small effect on these performance measures. Indeed, there is no significant performance difference when  $\sigma^2$  equals 1/4 and when it equals 1/8.

4.3. Cost and Profit Analysis. We now add a cost structure to the kitting process under study. In particular, cost and profit for the kitting systems of Sections 4.1 and 4.2 are analysed.

The proposed cost function is

$$c_1 (EQ_1 + EQ_2) + c_2 K + c_3 (b_1 + b_2), \quad (20)$$

where  $c_1$  is the holding cost of a part in the buffer,  $c_2$  is the shortage cost in one or in both of the buffers, and  $c_3$  is the loss cost. Note that for all figures, the input parameters are symmetric for both parts such that  $EQ_1 = EQ_2$  and  $b_1 = b_2$ .

4.3.1. Poisson Arrivals. In Figure 7, the total cost of applying kitting systems versus the buffer capacity (varying from 1 to 30) is depicted. Limiting ourselves to Poisson arrivals for both buffers,  $\rho_i = \lambda_i = 0.8$  for part  $i = \{1, 2\}$ . We compare kitting performance for different costs as depicted. As expected, higher cost values lead to higher total cost. The cost structure also affects the optimal buffer capacity. When looking at the different costs separately, we observe that a higher holding cost decreases the optimal buffer capacity (from 15 to 12) and a higher loss cost increases the optimal capacity (from 12 to 14), whereas the optimal buffer capacity remains the same (12) for a higher shortage cost. Obviously, buffering is interesting when the storage cost is low and when the cost of rejecting parts (because of a full buffer) is high. It is most interesting to observe that, as the capacity increases, the

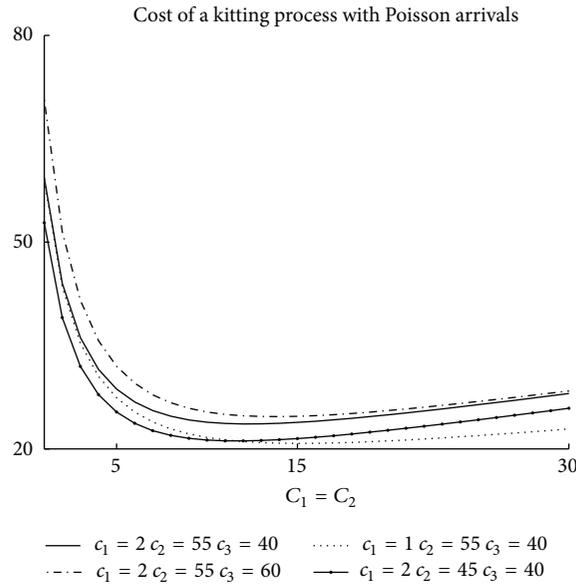


FIGURE 7: Each type of cost has a different impact on the value of the optimal buffer capacity.

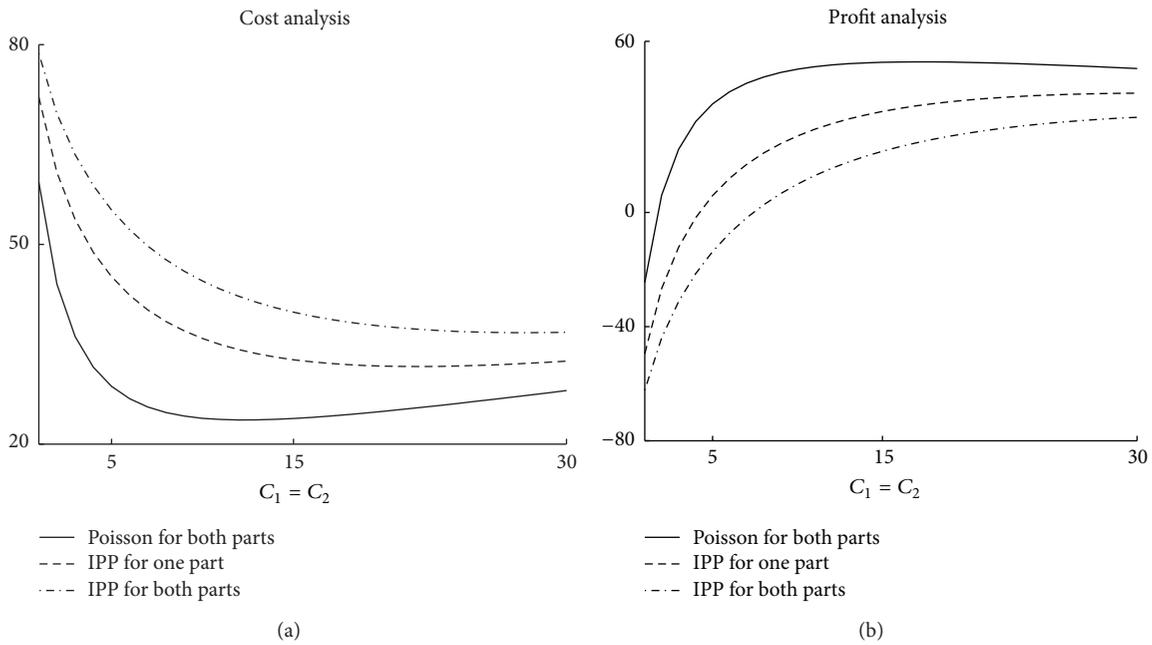


FIGURE 8: Production inefficiency results in higher required storage space.

cost models converge when the sum of the holding and the shortage cost  $(c_1 + c_2)$  is equal. Indeed, as the loss tends to zero, the loss cost tends to zero as well. Furthermore, when the capacity equals one, the state space of the kitting model has size 4 and hence the cost is easily calculated explicitly as

$$\text{total cost} = (2(c_1 + c_3)(\mu + 2\lambda) + 3\mu c_2)(3\mu + 2\lambda)^{-1}. \tag{21}$$

4.3.2. *Bursty Part Arrivals.* Next, we analyse the cost and the profit of kitting systems with different Markovian arrivals. We

use the same values to model the arrivals as in Figure 4. In Figures 8(a) and 8(b) the total cost (left) and the profit (right) versus the buffer capacity are depicted. We consider a holding cost  $c_1$  equal to 2, a shortage cost  $c_2$  equal to 55, and a loss cost  $c_3$  equal to 40. On the left figure, the optimal buffer capacity for Poisson arrivals is 12, for an interrupted Poisson arrival at one buffer it equals 22, and for an interrupted Poisson process at both buffers the optimal capacity is 28. As expected, the optimal buffer capacity increases as the burstiness in the production process increases. Concerning the profit analysis, we assume that the yield equals the throughput multiplied

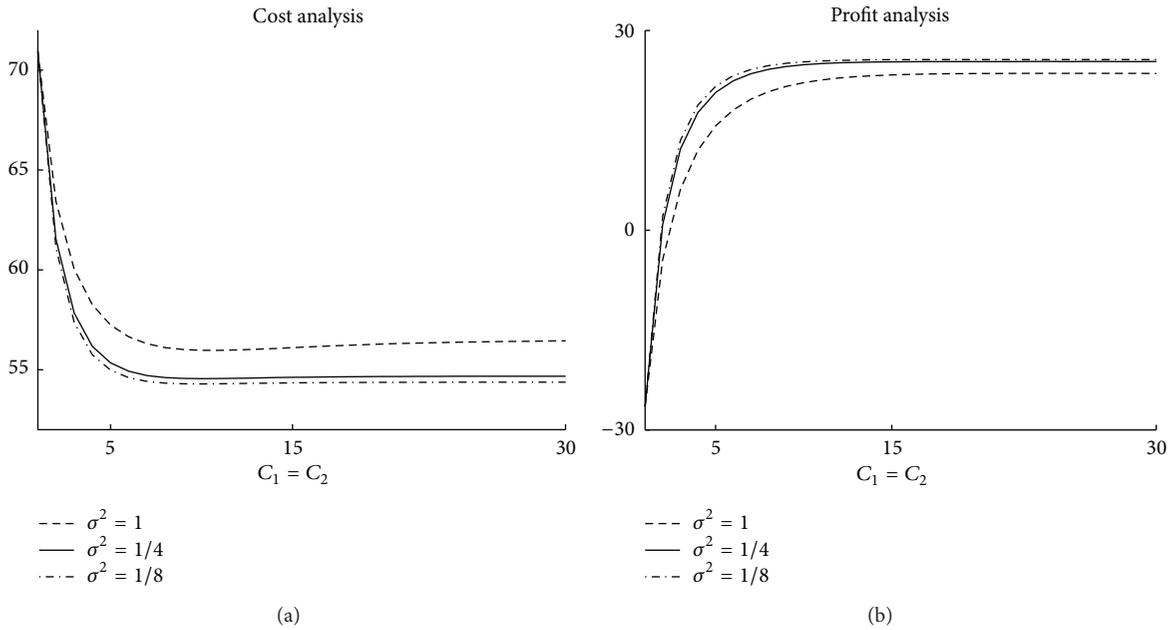


FIGURE 9: Shape of the service time distribution does not impact significantly the value of the optimal buffer capacity.

by a sale unit equal to 100. Assuming a maximum storage room of 30 parts, we observe that the optimal buffer capacity is 5, 11, and 12 for the depicted arrival processes, respectively. Consequently, kitting systems under production inefficiency require much higher storage space, especially when profit is applied as the parameter to determine the optimal buffer capacity. Moreover, the optimal capacity is very sensitive to the burstiness parameters  $\sigma$  and  $\kappa$ . In the example at hand, the plots suggest that the cost function is a convex function of the buffer capacity. However, one can also easily choose the cost parameters to obtain nonconvex cost functions.

4.3.3. *Phase-Type Kit Assembly Times.* Finally, Figures 9(a) and 9(b) depict the cost (left) and the profit (right) for a kitting system with Erlang-distributed kit assembly times versus the buffer capacity. As in the preceding figure, the profit equals the difference between the yield (equal to the throughput multiplied by 100) and the cost (defined by the parameters  $c_1 = 2, c_2 = 55$  and  $c_3 = 40$ ). Out of the numerical examples, we observe that the shape of the service time distribution has no significant impact on profit and cost. These results confirm those found in Section 4.2.

4.4. *Performance Analysis of Solution Methods.* We compare the performance of (sparse) GMRES and solving the Markov chain by standard LU decomposition Strang [19] on a kitting process with Poisson arrivals with rate  $\lambda_i = 0.8$  for part  $i = \{1, \dots, 2\}$  and with exponentially distributed kit assembly times with rate  $\mu$ . Figure 10 depicts both methods in terms of speed versus the state space for a kitting process starting from a symmetric buffer capacity  $C_1 = C_2 = 1$  to  $C_1 = C_2 = 60$ . While LU performs better than GMRES when the capacity (and state space) is small, the figure clearly

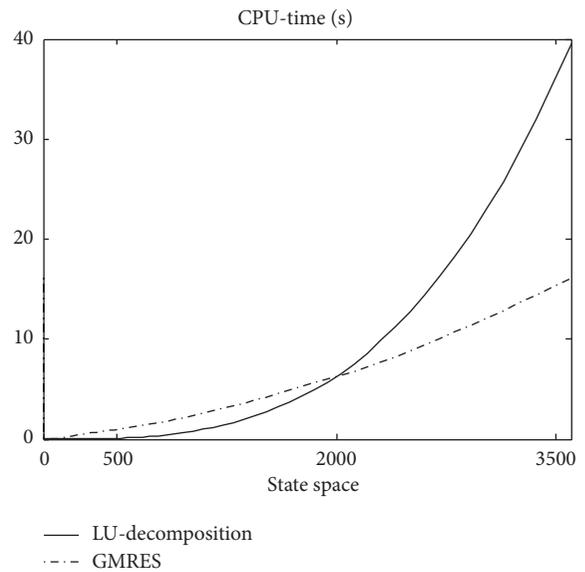


FIGURE 10: The GMRES method performs well in terms of speed.

shows that GMRES performs considerably better than LU decomposition for capacity.

## 5. Conclusion

In this paper, we evaluate the performance of two-part kitting processes in a Markovian setting. Furthermore, a cost-profit analysis is conducted and an approximation for our model is given. Note that the particularity of the studied kitting systems is that the part buffers are paired. This means that each demand requires both parts such that a

kit can only be assembled if both inventories are nonempty. Methodologically, we have applied sparse matrix techniques (e.g., GMRES) as most of the entries in the generator matrix have a value equal to zero. The solution is not exact but performs well in terms of solution speed and accuracy.

As our numerical results show, the interplay between the different queues leads to complex performance behaviour. For example, interruptions in the production of a part more negatively affect buffer performance of the other part. Overall, we observe extreme sensitivity of kitting performance on arrival process parameters while performance is reasonable insensitive to variation of the kitting time distribution. Furthermore, we determine the optimal buffer capacity based on a cost-profit analysis and establish a quite good approximation for our two-part kitting process.

## References

- [1] L. Medbo, "Assembly work execution and materials kit functionality in parallel flow assembly systems," *International Journal of Industrial Ergonomics*, vol. 31, no. 4, pp. 263–281, 2003.
- [2] Y. A. Bozer and L. F. McGinnis, "Kitting versus line stocking: a conceptual framework and a descriptive model," *International Journal of Production Economics*, vol. 28, no. 1, pp. 1–19, 1992.
- [3] P. Som, W. Wilhelm, and R. Disney, "Kitting process in a stochastic assembly system," *Queueing Systems*, vol. 17, no. 3–4, pp. 471–490, 1994.
- [4] H. Bryzner and M. I. Johansson, "Design and performance of kitting and order picking systems," *International Journal of Production Economics*, vol. 41, no. 1–3, pp. 115–125, 1995.
- [5] S. Ramachandran and D. Delen, "Performance analysis of a Kitting process in stochastic assembly systems," *Computers & Operations Research*, vol. 32, no. 3, pp. 449–463, 2005.
- [6] R. Ramakrishnan and A. Krishnamurthy, "Analytical approximations for kitting systems with multiple inputs," *Asia-Pacific Journal of Operational Research*, vol. 25, no. 2, pp. 187–216, 2008.
- [7] B. Johansson and M. I. Johansson, "High automated Kitting system for small parts: a case study from the Volvo uddevalla plant," in *Proceedings of the 23rd International Symposium on Automotive Technology and Automation*, pp. 75–82, Vienna, Austria, 1990.
- [8] W. J. Hopp and J. T. Simon, "Bounds and heuristics for assembly-like queues," *Queueing Systems*, vol. 4, no. 2, pp. 137–155, 1989.
- [9] J. M. Harrison, "Assembly-like queues," *Journal of Applied Probability*, vol. 10, pp. 354–367, 1973.
- [10] G. Latouche, "Queues with paired customers," *Journal of Applied Probability*, vol. 18, no. 3, pp. 684–696, 1981.
- [11] E. De Cuypere and D. Fiems, "Performance evaluation of a kitting process," in *Proceedings of the 17th International Conference on analytical and stochastic modelling techniques and applications*, vol. 6751, Lecture Notes in Computer Science, Venice, Italy, 2011.
- [12] P. Buchholz, P. Kemper, and J. Kriege, "Multi-class Markovian arrival processes and their parameter fitting," *Performance Evaluation*, vol. 67, no. 11, pp. 1092–1106, 2010.
- [13] D. Fiems, B. Steyaert, and H. Bruneel, *A Genetic Approach To Markovian Characterisation of H. 264 Scalable Videopages*, Multimedia Tools and Applications, 2011.
- [14] G. Latouche and V. Ramaswami, *Introduction To Matrix Analytic Methods in Stochastic Modeling*, ASA-SIAM Series on Statistics and Applied Probability, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1999.
- [15] W. J. Stewart, *Introduction To the Numerical Solution of Markov Chains*, Princeton University Press, Princeton, NJ, USA, 1994.
- [16] B. Philippe, Y. Saad, and W. Stewart, "Numerical methods in markov chain modeling," *Operations Research*, vol. 40, no. 6, pp. 1156–1179, 1992.
- [17] P. Buchholz, "Structured analysis approaches for large Markov chains," *Applied Numerical Mathematics*, vol. 31, no. 4, pp. 375–404, 1999.
- [18] Y. Saad and M. H. Schultz, "GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems," *Journal on Scientific and Statistical Computing*, vol. 7, no. 3, pp. 856–869, 1986.
- [19] G. Strang, *Linear Algebra and Its Applications*, Academic Press, Orlando, FL., 2nd edition, 1980.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

