Available online at www.sciencedirect.com
SCIENCE DIRECT
Computers & Operations Research 32 (2005) 449–463
computers & operations research
www.elsevier.com/locate/dsw

# Performance analysis of a kitting process in stochastic assembly systems

Satheesh Ramachandran[a], Dursun Delen[b,*]

[a]*Knowledge Based Systems, Inc., 1408 University Drive East, College Station, TX 77845, USA*
[b]*Department of MSIS, Oklahoma State University, 700 N. Greenwood Ave., Tulsa, OK 74106, USA*

**Abstract**

Kitting (accumulating components required for an assembly) plays a crucial role in determining the performance of a small-lot, multi-product, multi-level manufacturing system. In this paper, we analyze the kitting process as of a stochastic assembly system by treating it as an assembly-like queue. Specifically, we investigate the dynamics involved in a simple kitting process where two independent input streams feed into an assembly process. Unlike previous studies in this domain, we relax the assumption of finite buffer capacity constraint on the input buffers, and still show that the system remains stable under fairly mild conditions. It is expected that the findings of this study will provide manufacturing system designers with wider variety of control parameters to choose from in evaluating the system performance under a much broader set of control policies, which would lead to minimizing the associated costs.
© 2003 Elsevier Ltd. All rights reserved.

*Keywords:* Kitting; Assembly operations; Double-ended queue; Performance analysis

## 1. Introduction

Analysis of assembly operations plays a crucial role in improving the overall system performance in small-lot, multi-product, multi-level manufacturing operations, especially when the system operates under a stochastic environment [1,2]. According to Chen and Wilhelm [3] assembly operations form a significant portion of the overall product cycle time (hence the total manufacturing cost) in many industries including semiconductor manufacturing. Funk [4] reports that assembly operations comprise of up to 40% of total manufacturing cost in the electronics industry. Therefore, efficient control and management of assembly operations is crucial in reducing the cycle time of the final product.

* Corresponding author. Tel.: +1-918-594-8283; fax: +1-918-594-8281.
*E-mail address:* delen@okstate.edu (D. Delen).

0305-0548/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0305-0548(03)00247-8

Conventionally, analysis of assembly operations has been based on the assumption that the system operates deterministically. A more realistic analysis hinges on the recognition of the stochastic elements (i.e., random arrival and random service times) that influence the system. Component availability at the various buffers (and consequently, the delivery schedules) is significantly affected by these stochastic elements. The goal of this paper is to understand and evaluate the implications of kitting operations on the performance measures of assembly systems that operate under stochastic conditions.

Successful management of kitting operations increases the productivity of any assembly process [5]. In the electronics manufacturing industry, efficient kitting mechanisms simplify material flow and provide for better shop floor control [6]. Kitting operations are also studied at the level of production strategies such as MRP and JIT systems; where production is managed by either a push or a pull mechanism [7]. In such cases, efficient control of the kitting operations found to play an important role in lowering work in process (WIP) inventory and hence decreasing the operational cost. Researchers at the center for quick response manufacturing at the University of Wisconsin-Madison have been working on examining and comparing the analytical performance of push and pull production control strategies [8,9]. Their approximation models favors JIT (pull/Kanban) over MRP (push)-type production strategies.

Another domain that witnesses widespread use of kitting operations is the subcontracting practice in supply chain management [10], where subcontractors supply the individual components and services for the various products to the prime manufacturer and the manufacturer assembles the kits. One such application environment is found in the various US department-of-defense (DoD) aircraft repair depots (such as the Oklahoma City Air Logistics Center OC-ALC). In the shop floor lingo at the depots, a kit is an actual collection of parts needed to assemble an asset (such as a helicopter engine) to completion. Typically, these parts, which could either be manufactured internally or supplied by external contractors, are gathered in an assembly methodology to aid production. Given that kitting-type operations are commonly found in these environments, a central problem here is efficient control of the kit assembly process that optimizes the delivery of these kits based upon the actual upstream demand for these kits. One such recent initiative titled "lean value chain (LVC) for critical parts procurement" sponsored by the Air Force Wright Laboratory's Manufacturing Technology Directorate involved developing solutions that enable coordinated response to anticipated and known critical part problems [11,12]. A critical part is defined as any part whose anticipated or actual lack of availability will prevent on-time completion of the weapon system. Critical parts are often the result of ill-defined (or lack thereof) of control policies that dictate their delivery to the kitting process (historically, within the OC-ALC facilities such as the GE Rotor shop, there was little or no control policies for the parts ordering and procurement processes). The focus of the LVC project was to reengineer the processes of linking production with material/component procurement with the current effort being focused on developing and incorporating analytically driven control policies.

In a related ongoing research effort, Leung and Kamath [13] analyze a single-stage assembly system where two components are assembled into a single product via a kitting operation. Each component has its own finite buffer for temporary storage while waiting for its counterpart. When a pair of components is available, the components move into the assembly station, which has its own input buffer. They develop a model to approximately calculate performance measures, such as mean system time and mean queue length, when the component arrival

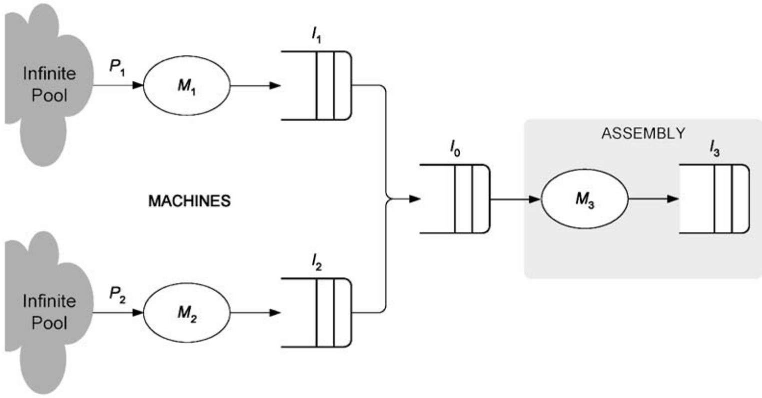


Fig. 1. The structure of the assembly system under investigation.

processes are Poisson and the assembly time follows a general distribution. In Kamath and Leung [14], this single-stage building block is used in the context of a production network to test the usefulness of the approximation developed. Chang and Chen [15] also looked at tandem queues as assembly-like queues in order to develop control policies that would increase the system performance measures.

In this paper, we investigate a simple kitting process with two input streams for an assembly system with the aim of understanding the dynamics involved. We assume that the arrival streams feeding the kitting process have state-dependent arrival intensities. The assembly system has a similar structure as modeled in [7,16], and is shown in Fig. 1.

In Fig. 1, $M_1$ and $M_2$ are machines processing parts $P_1$ and $P_2$, and $M_3$ assembles these machined components. $I_1$ and $I_2$ are the buffers for the machined components, $I_0$ is the buffer for the kitted component and $I_3$ is the buffer for the assembled component. Machines $M_1$ and $M_2$ are assumed to operate independently. They withdraw raw materials from their respective pools of infinite capacity and supply machined components to the buffers $I_1$ and $I_2$, respectively. A component arriving at buffer $I_1$ ($I_2$) is immediately kitted with a part from buffer $I_2$ ($I_1$), if one is available, and a kit is supposed to be ready for assembly operation at machine $M_3$. If the kit cannot be composed, the machined part is held in the buffer $I_1$ ($I_2$), and awaits the arrival of a "matching" part from $I_2$ ($I_1$). Once composed, the kit of matching components from $I_1$ and $I_2$ is sent to $I_0$ and the kit is considered to have arrived at $I_0$ (Fig. 2).

For exponential service times at $M_1$ and $M_2$ and finite buffer capacities at $I_1$ and $I_2$, Som et al. [16] characterize the occupancy distribution at $I_1$ and $I_2$ at kit departure epochs. Completed kits are shown to arrive at $I_0$ according to a Markov-renewal process. Also, when machines $M_1$ and $M_2$ have identical processing rates, and buffer capacities at $I_1$ and $I_2$ are large enough, they show that the arrival of completed kits to $I_0$ is well approximated by a Poisson process. This leads to the decoupling of the kitting operations from assembly, and hence to an easy analysis of the downstream assembly operations.

Stochastic assembly systems are often studied as assembly like queues [17,18]. Many followed the same approach in developing approximations for computing the performance measures of complex assembly operations [19,20]. Harrison [18] in a primarily theoretical study, established

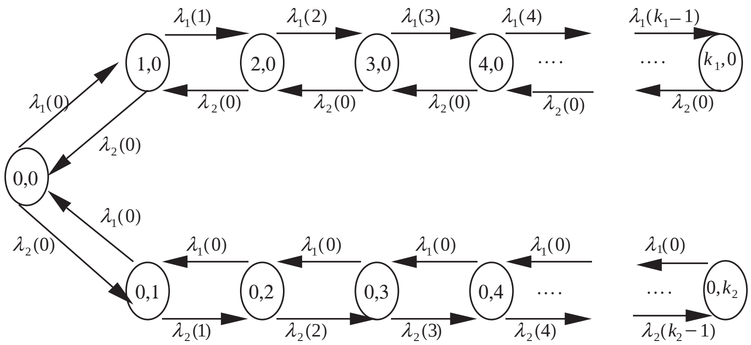


Fig. 2. Rate diagram for the kitting process.

stability conditions for an assembly queue with renewal and mutually independent arrival streams. He showed that a sufficient and necessary condition for the queue to be stable is for the component buffers at $I_1$ and $I_2$ to be finite. Thus, when there are no limitations on the inventory levels, the buffer sizes at $I_1$ and $I_2$ blow up, even when the arrival intensities (processing rates for $M_1$ and $M_2$) are the same. This can be intuitively explained by visualizing the queueing phenomena in the context of a double-ended queue [21]. A double-ended queue can best be described as the classical taxicab problem where taxies and passengers form two mutually separate queues [22]. A customer waits in the customer queue until a taxi is available, and taxies wait in the taxi queue until a customer is available. The two queues are interdependent and their combination is known as a double-ended queue. The underlying queueing process maps into a random walk on $\{\dots,-2,-1,0,1,2,\dots\}$, which is transient or recurrent null except when the queues are bounded. Hence, most of the analysis of assembly queues and kitting operations incorporates the finite buffer size assumption. A more realistic approach is to view the machine processing rates as control parameters, which dictate the performance measures of the system. The finite buffer capacities case is a specific policy for setting the control parameters which guarantees stability, but it need not be the optimal policy. The assumption of finite buffer capacity to ensure stability could be fairly restrictive. This is particularly true since in this case the system remains stable under fairly moderate conditions, allowing the system to be evaluated under a much broader set of control policies. This approach offers system designers with a wider variety of control parameters to choose from to minimize the associated costs. In this paper, we evaluate the system when the arrival rates (or the machine processing rates) to $I_1$ and $I_2$ are controlled as a function of the buffer sizes at $I_1$ and $I_2$, respectively. The service times of machines $M_1$ and $M_2$ are assumed to be exponential, and dependent on the buffer sizes at $I_1$ and $I_2$, respectively. Under these conditions in the next section, we characterize the probability laws for buffer sizes at $I_1$ and $I_2$ and establish conditions for system stability. Then we derive the waiting time distributions for kits arriving at $I_0$. We also show that waiting times degenerate to exponential waiting times under the conditions assumed by Som et al. [16].

The remainder of the paper is organized as follows. Section 2 presents the definitions, theorems and proofs of our approach. A simple numerical cost structure example is provided in Section 3. Section 4 concludes the paper by presenting the results and implications of our study.

## 2. Definitions and main results

Let $\lambda_1(n_1){:}n_1 \geqslant 0$ and $\lambda_2(n_2){:}n_2 \geqslant 0$ be the processing rate of machines $M_1$ and $M_2$, and respectively, where $n_1$ and $n_2$ are the number of machined components waiting at corresponding buffers $I_1$ and $I_2$. This is a generic characterization; for example, the special case for finite buffers of $k_1$ and $k_2$ at $I_1$ and $I_2$, respectively, and constant and identical machine processing times at $M_1$ and $M_2$ can be defined by the following conditions:

$$\begin{aligned}\lambda_1(n) &= \lambda, \quad 0 \leqslant n \leqslant k_1\\ &= 0, \quad n > k_1\end{aligned}$$

$$\begin{aligned}\lambda_2(n) &= \lambda, \quad 0 \leqslant n \leqslant k_2\\ &= 0, \quad n > k_2.\end{aligned}$$

In order to establish conditions on the control functions $\lambda_1(n_1){:}n_1 \geqslant 0$ and $\lambda_2(n_2){:}n_2 \geqslant 0$ which enable system stability, we characterize the probability laws of the number in system and establish the waiting time for the completed kits arriving at $I_0$ as a function of the processing rate at the machines $M_1$ and $M_2$. We show that when the inventory capacities at $I_1$ and $I_2$ are unlimited, the system is stable for very mild conditions on the control functions. Studying the behavior of the inventory levels at $I_1$ and $I_2$, and the departure rate as a function of the control parameters is useful in the selection of these parameters. We arrive at these stability conditions by first developing characterizations for a finite capacity system, and then developing the unlimited buffer size case as a limiting case of the finite capacity system.

**Theorem 1.** *Following the previous research* [2,6,16,17], *we let the service times of machines $M_1$ and $M_2$ be exponentially distributed with parameters $\lambda_1(n_1)$, and $\lambda_2(n_2)$, where $n_1$ and $n_2$ are the number of machined components at buffers $I_1$* and *$I_2$, respectively. The permissible queue sizes in both stations are $k_1$ and $k_2$, respectively. Let $\pi_{0,0}, \pi_{0,1}, \ldots, \pi_{0,k_2}, \pi_{1,0}, \ldots, \pi_{k_1,0}$ be the steady-state probabilities for the system states and let*

$$L = \left\{ \begin{array}{l} 1 + \left[\dfrac{\lambda_1(0)}{\lambda_2(0)} + \dfrac{\lambda_1(0)\lambda_1(1)}{\lambda_2^2(0)} + \dfrac{\lambda_1(0)\lambda_1(1)\lambda_1(2)}{\lambda_2^3(0)} + \cdots + \dfrac{\lambda_1(0)\ldots\lambda_1(k_1-1)}{\lambda_2^{k_1}(0)}\right] \\ \quad + \left[\dfrac{\lambda_2(0)}{\lambda_1(0)} + \dfrac{\lambda_2(0)\lambda_2(1)}{\lambda_1^2(0)} + \dfrac{\lambda_2(0)\lambda_2(1)\lambda_2(2)}{\lambda_1^3(0)} + \cdots + \dfrac{\lambda_2(0)\ldots\lambda_2(k_2-1)}{\lambda_2^{k_2}(0)}\right] \end{array} \right\}.$$

*Then,*

$$\pi_{0,0} = \frac{1}{L},$$

$$\pi_{0,k_2} = \frac{\lambda_2(k_2-1)\ldots\lambda_2(0)}{\lambda_1^{k_2}(0)L},$$

$$\pi_{k_1,0} = \frac{\lambda_1(k_1-1)\ldots\lambda_1(0)}{\lambda_2^{k_1}(0)L},$$

$$\pi_{0,n} = \frac{\lambda_1^{k_2-n}(0)}{\lambda_2(k_2-1)\ldots\lambda_2(n)}\,\pi_{0,k_2} \quad \textit{for } 0 < n < k_2,$$

$$\pi_{n,0} = \frac{\lambda_2^{k_1-n}(0)}{\lambda_1(k_1-1)\ldots\lambda_1(n)}\,\pi_{k_1,0} \quad \textit{for } 0 < n < k_1.$$

**Proof.** Let the state space for this assembly process be described as a two-tuple $(n_1, n_2)$, where $n_1$ and $n_2$ correspond to the number of parts in the buffers $I_1$ and $I_2$, respectively. The kitting process is such that if $n_1 > 0 (n_2 > 0)$ then $n_2 = 0 (n_1 = 0)$. Assuming an infinitesimal kitting time, (i.e., a part arriving at either buffer is immediately kitted with a part from the complementary buffer), we have if $n_1 > 0$, it follows that $n_2 = 0$ and vice versa.

The balance equation for state $(k_1, 0)$ gives

$$\pi_{k_1-1,0} = \frac{\lambda_2(0)}{\lambda_1(k_1-1)}\,\pi_{k_1,0}.$$

Similarly,

$$\pi_{k_1-2,0} = \frac{\lambda_2^2(0)}{\lambda_1(k_1-1)\lambda_1(k_1-2)}\,\pi_{k_1,0}$$

and in general for $0 < n < k_1$ it follows that

$$\pi_{n,0} = \frac{\lambda_2^{k_1-n}(0)}{\lambda_1(k_1-1)\ldots\lambda_1(n)}\,\pi_{k_1,0} \tag{1}$$

and

$$\pi_{0,0} = \frac{\lambda_2^{k_1}(0)}{\lambda_1(k_1-1)\ldots\lambda_1(0)}\,\pi_{k_1,0}. \tag{2}$$

By symmetry, we also have for $0 < n < k_2$

$$\pi_{0,n} = \frac{\lambda_1^{k_2-n}(0)}{\lambda_2(k_2-1)\ldots\lambda_2(n)}\,\pi_{0,k_2} \tag{3}$$

and

$$\pi_{0,0} = \frac{\lambda_1^{k_2}(0)}{\lambda_2(k_2-1)\ldots\lambda_2(0)}\,\pi_{0,k_2}.$$

Equating the two expressions for $\pi_{0,0}$, we get

$$\pi_{0,k_2} = \frac{\lambda_2^{k_1}(0)}{\lambda_1^{k_2}(0)}\,\frac{\lambda_2(k_1-1)\ldots\lambda_2(0)}{\lambda_1(k_1-1)\ldots\lambda_1(0)}\,\pi_{k_1,0}. \tag{4}$$

Substituting (3) in (2) we get

$$\pi_{0,n} = \frac{\lambda_2^{k_1}(0)}{\lambda_1^{n}(0)}\,\frac{\lambda_2(n-1)\ldots\lambda_2(0)}{\lambda_1(k_1-1)\ldots\lambda_1(0)}\,\pi_{k_1,0}. \tag{5}$$

The normalizing equation for this system is

$$(\pi_{k_1,0} + \pi_{k_1-1,0} + \pi_{k_1-2,0} + \pi_{k_1-3,0} + \cdots + \pi_{1,0}) + (\pi_{0,k_2} + \pi_{0,k_2-1} + \pi_{0,k_2-2} + \pi_{0,k_2-3} + \cdots + \pi_{0,1}) + \pi_{0,0} = 1.$$

Substituting for all the probabilities in terms of $\pi_{k_1,0}$, using Eqs. (1) and (5) in the above equation, we have

$$\pi_{k_1,0}\left[\begin{array}{l} 1 + \left\{\dfrac{\lambda_2(0)}{\lambda_1(k_1-1)} + \cdots + \dfrac{\lambda_2^{k_1-1}(0)}{\lambda_1(k_1-1)\ldots\lambda_1(1)} + \dfrac{\lambda_2^{k_1-1}(0)}{\lambda_1(k_1-1)\ldots\lambda_1(1)}\right\} \\ \quad + \left\{\dfrac{\lambda_2^{k_1}(0)}{\lambda_1^{k_2}(0)}\dfrac{\lambda_2(k_2-1)\ldots\lambda_2(0)}{\lambda_1(k_1-1)\ldots\lambda_1(0)} + \dfrac{\lambda_2^{k_1}(0)}{\lambda_1^{k_2-1}(0)}\dfrac{\lambda_2(k_2-2)\ldots\lambda_2(0)}{\lambda_1(k_1-1)\ldots\lambda_1(0)}\right. \\ \quad \left. + \cdots + \dfrac{\lambda_2^{k_1}(0)}{\lambda_1(0)}\dfrac{\lambda_2(0)}{\lambda_1(k_1-1)\ldots\lambda_1(0)}\right\} \end{array}\right] = 1.$$

After employing Eq. (2), the expression for $\pi_{0,0}$ follows as $\pi_{0,0} = 1/L$, where $L$ is as defined above. □

Next, we extend the finite capacity case to infinite buffers and derive some sufficient conditions for stability. For stability, we check conditions under which $\pi_{0,0} > 0$. This is equivalent to checking the condition that the series in the denominator of the expression for $\pi_{0,0}$ converges. The following theorem states the stability conditions for the control function of the kitting process.

**Theorem 2.** *If* $\lim_{k\to\infty}\lambda_1(k) < \lambda_2(0)$ *and* $\lim_{k\to\infty}\lambda_2(k) < \lambda_1(0)$ *then the system is stable.*

**Proof.** The queue is stable *iff* the series

$$\begin{array}{l} 1 + \left[\dfrac{\lambda_1(0)}{\lambda_2(0)} + \dfrac{\lambda_1(0)\lambda_1(1)}{\lambda_2^2(0)} + \dfrac{\lambda_1(0)\lambda_1(1)\lambda_1(2)}{\lambda_2^3(0)} + \cdots + \dfrac{\lambda_1(0)\ldots\lambda_1(k_1-1)}{\lambda_2^{k_1}(0)} + \cdots\right] \\ \quad + \left[\dfrac{\lambda_2(0)}{\lambda_1(0)} + \dfrac{\lambda_2(0)\lambda_2(1)}{\lambda_1^2(0)} + \dfrac{\lambda_2(0)\lambda_2(1)\lambda_2(2)}{\lambda_1^3(0)} + \cdots + \dfrac{\lambda_2(0)\ldots\lambda_2(k_2-1)}{\lambda_1^{k_2}(0)} + \cdots\right] \end{array} \quad \text{converges.}$$

The series above has all positive terms. A sufficient condition for the above series to converge is that both of the following series converge.

$$1 + \frac{\lambda_1(0)}{\lambda_2(0)} + \frac{\lambda_1(0)\lambda_1(1)}{\lambda_2^2(0)} + \frac{\lambda_1(0)\lambda_1(1)\lambda_1(2)}{\lambda_2^3(0)} + \cdots,$$

and

$$1 + \frac{\lambda_2(0)}{\lambda_1(0)} + \frac{\lambda_2(0)\lambda_2(1)}{\lambda_1^2(0)} + \frac{\lambda_2(0)\lambda_2(1)\lambda_2(2)}{\lambda_1^3(0)} + \cdots.$$

Let the $k$th term in the first series be $a_k$ and the $k$th term in the second series be $b_k$. Series 1 converges if, $\lim_{k\to\infty} a_{k+1}/a_k < 1$. Similarly, Series 2 converges if $\lim_{k\to\infty} b_{k+1}/b_k < 1$. We have,

$\lim_{k\to\infty} a_{k+1}/a_k = \lim_{k\to\infty} \lambda_1(k)/\lambda_2(0) < 1 \Rightarrow \lim_{k\to\infty} \lambda_1(k) < \lambda_2(0)$ and

$$\lim_{k\to\infty} \frac{b_{k+1}}{b_k} = \lim_{k\to\infty} \frac{\lambda_2(k)}{\lambda_1(0)} < 1 \quad \Rightarrow \quad \lim_{k\to\infty} \lambda_2(k) < \lambda_1(0),$$

thus proving the theorem. □

Based upon this result, it is evident that system stability is guaranteed under mild conditions on the control functions. Intuitively, the above result states that the system is guaranteed stability as long as the control functions $\lambda_2(k)$ and $\lambda_1(k)$ which represents the tendency to drift towards $(0,\infty)$ and $(\infty,0)$, respectively, are finally dominated by $\lambda_1(0)$ and $\lambda_2(0)$ (which represent the tendency of the system to pull back to the state $(0,0)$), respectively.

Next we establish the waiting time distribution for kits arriving at buffer $I_0$. Let $T_1, T_2, T_3, \ldots$ be the times of completion of successive kits. Let $X_1, X_2, X_3, \ldots$ be the queue sizes. Then Som et al. [16] show that when the maximum permissible buffer sizes $k_1$ and $k_2$ are finite $(X_n, T_n)$ form a Markov renewal process. They develop expressions for $P\{T_{n+1} - T_n \leqslant t\}$ and show that it approximates an exponential distribution as $k_1$ and $k_2$ become infinitely large. We use the result in Theorem 1 under the more general assumptions of infinite buffer capacity and controlled arrival rates to characterize the waiting time distributions for kits arriving at $I_0$.

**Theorem 3.** *Let $N_t$ be the number of kits completed up until time $t$. Let $E$ be the state space for the process i.e., $E = \{(k_1,0),\ldots,(0,0),\ldots,(0,k_2)\}$. Let $T_{N_t+1} - t = W_{N_t+1}$. Then the distribution of the waiting time $W_{N_t+1}$ at steady state is given by*

$$\lim_{t\to\infty} P\{W_{N_t+1} \leqslant y\} = A(1 - \mathrm{e}^{-\lambda_2(0)y}) + B(1 - \mathrm{e}^{-\lambda_1(0)y}) + (1 - A - B)(1 - \mathrm{e}^{-\lambda_1(0)y})(1 - \mathrm{e}^{-\lambda_2(0)y})$$

*where*

$$A = \sum_{n=1}^{\infty} \pi_{n,0} \quad \text{and} \quad B = \sum_{n=1}^{\infty} \pi_{0,n}.$$

**Proof.**

*Case* (a):

$$\lim_{t\to\infty} P\{W_{N_t+1} \leqslant y/\text{state of system} = (k,0),\ k > 0\}$$
$$= P\{\text{arrival at station } I_2 \text{ before time } y \text{ units/state of system} = (k,0),\ k > 0\}$$
$$= 1 - \mathrm{e}^{-\lambda_2(0)y}.$$

*Case* (b):

$$\lim_{t\to\infty} P\{W_{N_t+1} \leqslant y/\text{state of system} = (0,k),\ k > 0\}$$
$$= P\{\text{arrival at station } I_1 \text{ before time } y \text{ units/state of system} = (0,k),\ k > 0\}$$
$$= 1 - \mathrm{e}^{-\lambda_1(0)y}.$$

*Case* (c):

$$\lim_{t\to\infty} P\{W_{N_t+1} \leqslant y/\text{state of system} = (0,0)\}$$
$$= P\{\text{arrival at station } I_1 \,\&\, I_2 \text{ before time } y \text{ units/state of system} = (0,0)\}$$
$$= (1 - \mathrm{e}^{-\lambda_1(0)y})(1 - \mathrm{e}^{-\lambda_2(0)y}).$$

Let

$$A = \sum_{n=1}^{\infty} \pi_{n,0}, \quad \text{and} \quad B = \sum_{n=1}^{\infty} \pi_{0,n}.$$

Then, $\pi_{0,0} = 1 - A - B$, and we have,

$$\lim_{t\to\infty} P\{W_{N_t+1} \leqslant y\} = A(1 - \mathrm{e}^{-\lambda_2(0)y}) + B(1 - \mathrm{e}^{-\lambda_1(0)y}) + (1 - A - B)(1 - \mathrm{e}^{-\lambda_1(0)y})(1 - \mathrm{e}^{-\lambda_2(0)y}). \tag{6}$$

Next, we derive the joint distribution of two successive kit completions times from an arbitrary time $t$. □

**Theorem 4.** *Let $T_1, T_2, T_3, \ldots$ be the times of completion of successive kits. Let $N_t$ be the number of kits completed up until time $t$. Let the buffer sizes be finite at $k_1$ and $k_2$, respectively. Let $E$ be the state space for the process i.e., $E = \{(k_1,0),\ldots,(0,0),\ldots,(0,k_2)\}$. Then the joint distribution of the waiting time and the next inter-departure time is given by*

$$\begin{aligned}
&\lim_{t\to\infty} P\{T_{N_t+1} - t \leqslant y_1, T_{N_t+2} - T_{N_t+1} \leqslant y_2\} \\
&= A'(1 - \mathrm{e}^{-\lambda_2(0)y_1})(1 - \mathrm{e}^{-\lambda_2(0)y_2}) + B'(1 - \mathrm{e}^{-\lambda_1(0)y_1})(1 - \mathrm{e}^{-\lambda_1(0)y_2}) \\
&\quad + \pi_{1,0}\left[\frac{\lambda_1(0)}{\lambda_1(0) + \lambda_2(0)}(1 - \mathrm{e}^{-\lambda_2(0)y_1})(1 - \mathrm{e}^{-\lambda_2(0)y_2})\right. \\
&\quad \left. + \frac{\lambda_2(0)}{\lambda_1(0) + \lambda_2(0)}(1 - \mathrm{e}^{-\lambda_2(0)y_1})(1 - \mathrm{e}^{-\lambda_2(0)y_2})(1 - \mathrm{e}^{-\lambda_1(0)y_2})\right] \\
&\quad + \pi_{0,1}\left[\frac{\lambda_1(0)}{\lambda_1(0) + \lambda_2(0)}(1 - \mathrm{e}^{-\lambda_1(0)y_1})(1 - \mathrm{e}^{-\lambda_1(0)y_2})(1 - \mathrm{e}^{-\lambda_2(0)y_2})\right. \\
&\quad \left. + \frac{\lambda_2(0)}{\lambda_1(0) + \lambda_2(0)}(1 - \mathrm{e}^{-\lambda_1(0)y_1})(1 - \mathrm{e}^{-\lambda_1(0)y_2})\right] \\
&\quad + \pi_{0,0}\left[\begin{array}{l} \dfrac{\lambda_1(0)}{\lambda_1(0) + \lambda_2(0)}(1 - \mathrm{e}^{-\lambda_1(0)y_1})(1 - \mathrm{e}^{-\lambda_2(0)y_1})(1 - \mathrm{e}^{-\lambda_2(0)y_2}) \\ \quad + \dfrac{\lambda_2(0)}{\lambda_1(0) + \lambda_2(0)}(1 - \mathrm{e}^{-\lambda_1(0)y_1})(1 - \mathrm{e}^{-\lambda_2(0)y_1})(1 - \mathrm{e}^{-\lambda_1(0)y_2}) \end{array}\right],
\end{aligned} \tag{7}$$

*where*

$$A' = \sum_{n=1}^{k_1} \pi_{n,0} \quad \textit{and} \quad B' = \sum_{n=1}^{k_2} \pi_{0,n}.$$

**Proof.** We can write the above expression as

$$\begin{aligned} &P\{T_{N_t+1} - t \leqslant y_1, T_{N_t+2} - T_{N_t+1} \leqslant y_2\} \\ &\quad = \sum_{i \in E} \pi_i P\{T_{N_t+1} - t \leqslant y_1, T_{N_t+2} - T_{N_t+1} \leqslant y_2 / \text{state at } t = i\} \\ &\quad = \sum_{i \in E} \pi_i P\{T_{N_t+2} - T_{N_t+1} \leqslant y_2 / T_{N_t+1} - t \leqslant y_1\} P\{T_{N_t+1} - t \leqslant y_1 / \text{state at } t = i\}. \end{aligned} \tag{8}$$

*Case* (a) ($X_{N_t} = (k, 0),\ k > 1$):

$$P\{T_{N_t+2} - T_{N_t+1} \leqslant y_2 / T_{N_t+1} - t \leqslant y_1,\ \text{state at } t = i\} = (1 - \mathrm{e}^{-\lambda_2(0) y_2}). \tag{9}$$

*Case* (b) ($X_{N_t} = (0, k),\ k > 1$):

$$P\{T_{N_t+2} - T_{N_t+1} \leqslant y_2 / T_{N_t+1} - t \leqslant y_1,\ \text{state at } t = i\} = (1 - \mathrm{e}^{-\lambda_1(0) y_2}). \tag{10}$$

*Case* (c) ($X_{N_t} = (1, 0)$):

$$\begin{aligned} &P\{T_{N_t+2} - T_{N_t+1} \leqslant y_2 / T_{N_t+1} - t \leqslant y_1,\ \text{state at } t = i\} \\ &\quad = \frac{\lambda_1(0)}{\lambda_1(0) + \lambda_2(0)} (1 - \mathrm{e}^{-\lambda_2(0) y_2}) + \frac{\lambda_2(0)}{\lambda_1(0) + \lambda_2(0)} (1 - \mathrm{e}^{-\lambda_1(0) y_2})(1 - \mathrm{e}^{-\lambda_2(0) y_2}). \end{aligned} \tag{11}$$

*Case* (d) ($X_{N_t} = (0, 1)$):

$$\begin{aligned} &P\{T_{N_t+2} - T_{N_t+1} \leqslant y_2 / T_{N_t+1} - t \leqslant y_1,\ \text{state at } t = i\} \\ &\quad = \frac{\lambda_1(0)}{\lambda_1(0) + \lambda_2(0)} (1 - \mathrm{e}^{-\lambda_1(0) y_2})(1 - \mathrm{e}^{-\lambda_2(0) y_2}) + \frac{\lambda_2(0)}{\lambda_1(0) + \lambda_2(0)} (1 - \mathrm{e}^{-\lambda_1(0) y_2}). \end{aligned} \tag{12}$$

*Case* (e) ($X_{N_t} = (0, 0)$):

$$\begin{aligned} &P\{T_{N_t+2} - T_{N_t+1} \leqslant y_2 / T_{N_t+1} - t \leqslant y_1,\ \text{state at } t = i\} \\ &\quad = \frac{\lambda_1(0)}{\lambda_1(0) + \lambda_2(0)} (1 - \mathrm{e}^{-\lambda_2(0) y_2}) + \frac{\lambda_2(0)}{\lambda_1(0) + \lambda_2(0)} (1 - \mathrm{e}^{-\lambda_1(0) y_2}). \end{aligned} \tag{13}$$

Substituting Eqs. (9) and (13) in Eq. (8) and using Theorem 3, we obtain the desired result.

Theorem 3 can be used to estimate expected remaining waiting times. Theorem 4 can be used to study the correlations structure of the kit completion process. Based upon Theorem 4, the following corollary states that the waiting times are independent under the assumption of $\lambda_1(n)=\lambda_2(n)=\lambda$. □

**Corollary.** *When both buffer capacities are infinite and $\lambda_1(n)=\lambda_2(n)=\lambda$ for all n, then the above joint distribution reduces to a product of two exponentially distributed intervals.*

**Proof.** Let $k_1 = k_2 = k$. Then, $\pi_i = 1/(2k+1)\ \forall i \in E$. The right-hand side in Eq. (7) as given in Theorem 4, reduces to

$$\left(\frac{2k-2}{2k+1}+\frac{1}{2k+1}\right)(1-\mathrm{e}^{-\lambda y_1})(1-\mathrm{e}^{-\lambda y_2})$$
$$+\frac{1}{2k+1}[(1-\mathrm{e}^{-\lambda y_1})(1-\mathrm{e}^{-\lambda y_2})^2+(1-\mathrm{e}^{-\lambda y_1})^2(1-\mathrm{e}^{-\lambda y_2})]$$
$$+\frac{1}{2k+1}(1-\mathrm{e}^{-\lambda y_1})(1-\mathrm{e}^{-\lambda y_2})(2-\mathrm{e}^{-\lambda y_1}-\mathrm{e}^{-\lambda y_2}).$$

If we let $k \to \infty$, we have $\lim_{t\to\infty} P\{T_{N_t+2}-T_{N_t+1} \leqslant y_2, T_{N_t+1}-t \leqslant y_1\}=(1-\mathrm{e}^{-\lambda y_1})(1-\mathrm{e}^{-\lambda y_2})$. □

In the next section, we use simple numerical examples to gain further insights and select control parameters for minimizing overall system costs.

## 3. Numerical example

Based upon the results of the previous sections it is evident that system stability is guaranteed under reasonably mild conditions on the control functions. This poses the system designer with the following question: given a particular cost structure (such as the inventory holding cost, delivery rate requirements), from the class of control functions satisfying the stability criterion, what is the 'optimal' function for the processing rates? Consider the following set that defines sequence-tuples that satisfy the stability criterion:

$$S=\left\{(a_j,\, a_k){:}a_j=[a_{j,n}]_{n=0}^{\infty},\ a_k=[a_{k,n}]_{n=0}^{\infty} \text{ and } \lim_{n\to\infty} a_{j,n}<a_{k,0};\ \lim_{n\to\infty} a_{k,n}<a_{j,0}\right\}.$$

Any element of the set $(a_j, a_k)\in S$ is an admissible control policy. Then the overall cost function $f$ can be generically defined in terms of the control policy and the cost parameters as

$$\textit{cost function} = f((a_j,\, a_k),\ \textit{cost parameters}),$$

and the optimal control policy $(a_j^*, a_k^*)$ satisfies

$$f(a_j^*,\, a_k^*)=\min_{All(a_j,a_k)\in S} f((a_j,\, a_k),\ \textit{cost parameters}).$$

Intuitively it can be reasoned that there exists no particular control function that minimizes total cost over all cost structures. In other words, the nature of the control function would be dependent on particular cost structure that is present in the application domain. Also, the set of control policies that are admissible in a domain would be dictated by the capacities of the machines producing the individual components (machines $M_1$ and $M_2$). Then the set of admissible control policies are restricted to

$$S = \left\{ \begin{array}{ll} (a_j, a_k){:}a_j = [a_{j,n}]_{n=0}^{\infty},\ a_k = [a_{k,n}]_{n=0}^{\infty} \text{ and } & \lim\limits_{n\to\infty} a_{j,n} < a_{k,0};\ \lim\limits_{n\to\infty} a_{k,n} < a_{j,0}; \\ & \sup[a_{j,n}]_{n=0}^{\infty} \leqslant \lambda_{\max}^1;\ \sup[a_{k,n}]_{n=0}^{\infty} \leqslant \lambda_{\max}^2 \end{array} \right\},$$

where $\lambda_{\max}^1$ and $\lambda_{\max}^2$ are the upper limits for production capacities at machines $M_1$ and $M_2$, respectively. A closed-form analytical solution to the problem defined above is difficult, and is the focus of our ongoing research investigation. Nevertheless, we can leverage the research results from the previous sections to develop simple, yet practical, control strategies (which although not rigorous, they offer some level of control on the operational costs). We show a sample numerical exercise that illustrates the application of the theoretical results. Consider three specific classes of control functions for processing rates at the machines that satisfy the stability criterion.

$$(1)\ a_{j,n} = a_{k,n} = \lambda_1(n) = \lambda_2(n) = C_0\left(\frac{1}{r^n}\right); \quad r > 1,\ n \geqslant 0,$$

$$(2)\ a_{j,n} = a_{k,n} = \lambda_1(n) = \lambda_2(n) = \begin{cases} C_1\left(\dfrac{1}{n}\right), & n > 0, \\ C_2, & n = 0. \end{cases}$$

$$(3)\ a_{j,n} = a_{k,n} = \lambda_1(n) = \lambda_2(n) = \begin{cases} C_3\left(1 - \dfrac{n}{M}\right), & 0 \leqslant n \leqslant M, \\ 0 & \text{otherwise.} \end{cases}$$

All of these control functions belong to a specific control function type that can be tuned in terms of parameters (we can call this the *control function parameter*). For example, the first control function type is a geometric control function, in which by modifying the parameter $r$, different production intensities (and different cost performances) can be achieved. A simple formulation for the cost function would incorporates a tradeoff between the *lateness cost* of assembled kits (cost that is proportional to the waiting time for completed kits) and the *holding cost* of components at the individual buffers. The lateness cost reduces when the inter-departure times of successive assemblies have a lower mean (kits are generated at a higher rate). If we tend to have a high number of parts in the buffers (higher *holding costs*), then we tend to move away from the situation in which both buffers are starved, which reduces the mean of the inter-departure times. Inherent in the notion of a *lateness cost* is the assumption of an infinite demand of assembled kits at the downstream buffer $I_0$. Let

$C_h$ = Holding cost of a part in buffer $I_1$ or $I_2$.

$C_l$ = Lateness cost of an assembled part.

$I$ = Total number of parts in buffers $I_1$ and $I_2$.

$W$ = Remaining waiting time of an assembled kit.

$TC$ = Total cost.

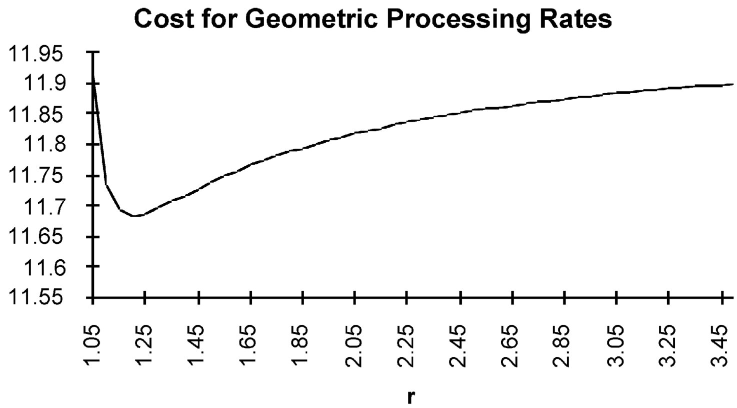


Fig. 3. Cost values for geometric processing rates with $C_h = 0.5$ and $C_l = 5$.

Table 1
Total cost at different cost combinations for three machine processing rates

| Control policies | Policy 1, $C_h = 0.5, C_l = 5,$ $C_0 = 1$, optimal $r$ | Policy 2, $C_h = 0.5, C_l = 5,$ $C_1 = 1, C_2 = 1$ | Policy 3, $C_h = 0.5, C_l = 5,$ $C_3 = 1, M = 10,000$ |
|---|---|---|---|
| Cost | 11.683 | 11.621 | 12.863 |

Then, define the total system costs as

$$TC = C_h E[I] + C_l E[W],$$

$E[I]$ is the expected number of components at buffers $I_1$ and $I_2$, and is defined as

$$E[I] = \sum_{i=0}^{\infty} iP(I = i).$$

The results from Theorem 1 can be used to compute $P(I = i)$. Similarly, if $f(y)$ is the density corresponding to the distribution in Theorem 3, then the expected remaining waiting time $E[W]$ of an assembled kit is defined as follows:

$$E[W] = \int_0^{\infty} y f(y)\, \mathrm{d}y.$$

For a particular combination of $C_h\,(=0.5)$ and $C_l\,(=5)$, and parameter $C_0 = 1$, Fig. 3 shows the values for $TC$ as a function of $r$ (achieving an optimal cost value of 11.683). The entries in Table 1 compare this optimal $TC$ value with the second (for $C_1 = 1, C_2 = 1$) and the third (for $C_3 = 1, M = 10,000$) control policies. We can see that the second processing rate (where the processing rate equals the reciprocal of the number of parts in the buffer) performs the best.

Although the above analysis makes simplifying assumptions and employs a primitive cost structure, it serves to illustrate the relevance of using machine processing rate functions as control parameters in improving the performance of assembly operations.

## 4. Conclusions

In this paper, we characterize probability laws for queue sizes at buffers for a kitting process. We derive the distribution of remaining waiting time for kits feeding the downstream process. We show that queue sizes at the component buffers are stable under very mild conditions for the control functions. This is in contrast to previous research, which analyses kitting systems of finite component buffer capacities, where the finite buffer sizes are imposed to ensure stability. This offers system designers a wide variety of control functions to choose from so as to have the flexibility to minimize cost given the cost structure at hand.

## Acknowledgements

The authors would like to express their gratitude to Professor W.E. Wilhelm and to Professor Manjunath Kamath for their invaluable comments and directions for improving the content and presentation of this paper.

## References

[1] Takahashi M, Osawa H, Fujisawa T. A stochastic assembly system with resume levels. Asia-Pacific Journal of Operational Research 1998;15:127–46.
[2] Chen JF, Wilhelm WE. An evaluation of heuristics for allocating components to kits in small-lot, multi-echelon assembly systems. International Journal of Production Research 1993;31(12):2835–56.
[3] Chen JF, Wilhelm WE. Kitting in multi-echelon, multi-product assembly systems with part substitutable. International Journal of Production Research 1997;35(10):2871–97.
[4] Funk JL. The potential market for robot assembly. International Journal of Production Research 1986;24(3):663–86.
[5] Ding FY. Kitting in JIT production: a kitting project at a tractor plant. IE Solutions, September 1992. p. 42–4.
[6] Wilhelm WE, Som P, Carroll B. A model for implementing a paradigm of time managed material flow control in certain assembly systems. International Journal of Production Research 1992;30(9):2063–86.
[7] Wilhelm WE, Som P. Analysis of single-stage, single-product, stochastic, MRP controlled assembly system. European Journal of Operations Research 1998;108:74–93.
[8] Ananth K, Suri R, Vernon M. Re-examining the performance of push, pull and hybrid material control strategies for multi-product flexible manufacturing systems. Technical Report. December 2000. p. 1–28.
[9] Krishnamurthy A, Suri R, Vernon M. A new approach for analyzing queueing models of material control strategies in manufacturing systems. In: Proceedings of the Fourth International Workshop on Queueing Networks with Finite Capacity (QNETs2000), West Yorkshire, U.K., July 2000.
[10] Simchi-Levi D, Kaminsky P, Simchi-Levi E. Designing and managing the supply chain, 2nd ed. New York: McGraw-Hill/Irwin; 2003.
[11] Painter M, Mayer R. Lean value chain for critical parts procurement. Final report prepared for air force wright laboratory's materials and manufacturing technology directorate—contract number F33615-98-C-5168, 2002.
[12] Benjamin P, Delen D, Lo A, Painter M, Graul M. A critical parts dashboard (CPD). Proceedings of the International Conference on Artificial Intelligence (IC-AI'2001), Las Vegas, Nevada: CSREA Press, June 25–28, 2001. p. 1562–7.
[13] Leung YT, Kamath M. Performance analysis of single-stage assembly system. ORSA/TIMS Joint National Meeting, 1994.
[14] Kamath M, Leung YT. Performance analysis of production networks involving assembly operations. INFORMS Los Angeles National Meeting, Spring 1995.
[15] Chang KH, Chen WF. Admission control policies for two-stage tandem queues with no waiting spaces. Computers and Operations Research 2003;30:589–601.

[16] Som P, Wilhelm WE, Disney RL. Kitting process in a stochastic assembly system. Queueing Systems 1994;17: 471–90.
[17] Hopp WJ, Simon JT. Bounds and Heuristics for assembly-like queues. Queueing Systems: Theory and Applications 1989;4:137–56.
[18] Harrison JM. Assembly-like queues. Journal of Applied Probability 1971;10:354–67.
[19] Hemachandra N, Eedupuganti SK. Performance analysis and buffer allocations in some open assembly systems. Computers and Operations Research 2003;30(5):695–704.
[20] Zhuang L, Wong YS, Fuh JYH, Yee CY. On the role of a queueing network model in the design of a complex assembly system. Robotics and Computer-Integrated Manufacturing 1998;14:153–61.
[21] Dobbie JM. A double-ended queueing problem of Kendall. Operations Research 1961;9:755–7.
[22] Kashyap BRK. A double ended queueing system with limited waiting space. Proceedings of the National Institute of Science of India 1965;31:559–70.