Optimal Control of Pull Manufacturing Systems* †

J. R. Perkins‡ and P. R. Kumar§

Abstract

We consider the problem of optimal control of pull manufacturing systems. We study a fluid model of a flow shop, with buffer holding costs non-decreasing along the route. The system is subject to a constant exogenous demand, thus incurring additional shortfall/inventory costs. The objective is to determine the optimal control for the production rate at each machine in the system.

We exhibit a decomposition of the flow shop into "sections" of contiguous machines, where, in each section, the head machine is the bottleneck for the downstream system. We exhibit the form of an optimal control, and show that it is characterized by a set of "deferral times," one for each head machine. Machines which are upstream of a head machine simply adopt a "just–in–time" production policy. The head machines initially stay idle for a period equal to their deferral time, and thereafter produce as fast as possible, until the initial shortfall is eliminated. The optimal values of these deferral times are simply obtained by solving a set of quadratic programming problems.

We also exhibit special cases of re–entrant lines, for which the optimal control is similarly computable.

1 Introduction

The manufacture of products generally involves many operations on many machines, and requires many decisions. Modelling the important features of manufacturing systems has to

*Please address all correspondence to the second author.

†The research reported here has been supported in part by the National Science Foundation under Grant Nos. ECS-90-25007 and ECS-92-16487, the U.S. Army Research Office under Contract No. DAAL-03-91-G-0182, the Joint Services Electronics Program under Contract No. N00014-90-J1270, and by the Manufacturing Research Center of the University of Illinois.

‡Boston University, Dept. of Manufacturing Engineering, 44 Cummington Street, Boston, MA 02215.

§University of Illinois, Dept. of Electrical and Computer Engineering, and the Coordinated Science Lab, 1308 West Main Street, Urbana, IL 61801.

be tempered by the need to retain mathematical tractability. Thus, the control of manufacturing systems is often divided into simplified hierarchical levels, see Gershwin [1]. The purpose of this paper is to examine the level of machine scheduling and part release.

Classical scheduling theory considers the deterministic problem of scheduling a fixed number of jobs on a given set of machines, so as to minimize a performance criterion, such as makespan, tardiness, or lateness. There is an extensive body of work concerned with this model; see [2], [3], and [4] for an introduction, and [5], [6], [7], and [8] for overviews of the literature. Determining the minimizing solutions to these static combinatorial optimization problems is usually computationally infeasible, even for a few jobs on a few machines; see, for example, Muth and Thompson [9]. Moreover, even when it is determined, an optimal solution within a combinatorial framework may not provide any qualitative insight into the structure of the problem at hand.

Perkins and Kumar [10] attack the issue of scheduling manufacturing systems from a different perspective. They formulate a "push" model, which combines dynamics and feedback. Since the system is dynamic, the determination of stable scheduling policies and performance bounds is a key point of emphasis. This approach is further pursued in Kumar and Seidman [11], Chase and Ramadge [12], Lou, Sethi and Sorger [13], Perkins, Humes and Kumar [14], and Burgess and Passino [15].

In this paper, we examine the problem of optimal control of a pull model of manufacturing systems. The objective is to determine an optimal control which minimizes the sum of buffer holding costs, and system shortfall/inventory costs, when subjected to an exogenous demand of constant rate. We provide the optimal control for flow shops where the buffer holding costs are non-decreasing along the route. Thereby, we obtain the optimal part release and machine scheduling policies.

We proceed by exhibiting a decomposition of the flow shop into "sections" of contiguous machines. In each section, the head machine is the bottleneck for the downstream system. We solve the optimal control problem by showing the form of an optimal production rate at each machine. It is characterized by a set of "deferral times," one for each head machine. The machines upstream of a head machine in a section simply follow a "just–in–time" production policy. The head machines initially stay idle for a period of time equal to their deferral time, and thereafter produce as fast as possible, until the initial shortfall is eliminated. The optimal values of these deferral times are shown to be obtained from the solution of a set of quadratic programming problems.

We also extend our results to certain re-entrant lines with non-decreasing buffer holding costs. For the case in which the bottleneck for the system is the machine servicing the final buffer, we determine an optimal control.

For some other problem settings related to optimal control of networks, we refer the reader to Hajek and Ogier [16] and Chen and Yao [17], and the references cited in them. In [16], the objective is to empty a network in minimum time. However, there is no drainage from the system, and hence no negative buffer levels. In contrast, the difficulty in our problem lies in the fact that the final buffer can assume negative values, which represent the shortfall in system output. In [17], a push model with buffer holding costs is considered. A certain class of myopically optimal controls, which result in the minimum running cost uniformly at every time instant, are investigated, assuming the existence of such controls.

The optimal control of systems with machine failures has been determined for a single machine; see Akella and Kumar [18], Bielecki and Kumar [19], and Sharifnia [20]. Recently, the cost of certain "aiming" policies is computed in Glasserman [21]. Although we consider machines which do not fail, the results presented here could be used to derive performance bounds on manufacturing systems which are nearly reliable, i.e., the machines rarely fail or fail for short periods.

2 System Description

Consider the manufacturing system shown in Figure 1. It consists of L machines $\{M_0, M_1, \ldots, M_{L-1}\}$ in tandem. The system produces a single part-type. Parts leaving machine M_i flow into a buffer b_{i+1}, where they are held for processing by machine M_{i+1}. b_0 serving machine M_0 is not a buffer, but is an infinite reservoir of raw parts. The output

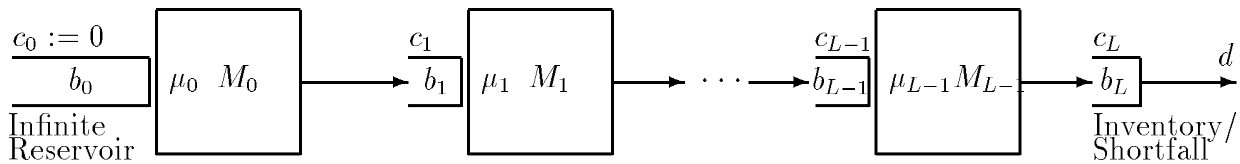

Figure 1: The flow shop with drainage.

buffer b_L is a "virtual" buffer. It is constantly depleted, due to demand, at a rate d, and replenished by parts exiting machine M_{L-1}. If the initial amount in b_L and the cumulative input to b_L up to a time t, together exceed the cumulative depletion, then the level of the buffer b_L at that time t is positive, signifying a positive *inventory* of finished goods awaiting shipping. Otherwise, the buffer level of b_L is negative, signifying a *shortfall*.

We consider a fluid model and suppose that machine M_i can process parts at a rate up to μ_i parts per unit time. We also assume that the system has enough capacity to meet the demand, i.e., $\mu_i > d$ for all i.

We suppose that a unit of material in buffer b_i incurs a holding cost of c_i units per unit time, for $1 \leq i \leq L-1$. No costs accrue for parts in b_0. For the output virtual buffer b_L, each unit of *positive* material (i.e., inventory) incurs a cost of c_L^+ units per unit time, while each unit of *negative* material (i.e., shortfall) incurs a cost of c_L^- units per unit time. We assume $c_i \geq 0$ for $0 \leq i \leq L-1$, and $c_L^+, c_L^- \geq 0$.

The system description is completed by the specification of an initial condition $x(0) = (x_1(0), x_2(0), \ldots, x_{L-1}(0), x_L(0))$, where $x_i(t)$ is the level of buffer b_i at time t. Note that $x_i(0) \geq 0$ for $1 \leq i \leq L-1$, but $x_L(0)$ may be positive or negative.

A control input $u(t)$ is *feasible* if it meets the following conditions:

(i) $u(t) = (u_0(t), \ldots, u_{L-1}(t))$ is a measurable function.

(ii) $0 \leq u_i(t) \leq \mu_i$ for all $t \geq 0$.

(iii) Define the level of buffer b_i at time t by $x_i(t) := x_i(0) + \int_0^t [u_{i-1}(\sigma) - u_i(\sigma)] d\sigma$. We require that $x_i(t) \geq 0$ for all $t \geq 0$ and $1 \leq i \leq L-1$.

(iv) If $x_i(t) = 0$, then $u_i(t) \leq u_{i-1}(t)$.

We shall denote the class of feasible controls by $\mathcal{F}$.

The goal of scheduling is to choose the vector production rate function $u(t) \in \mathcal{F}$ so as to minimize the infinite horizon total cost,

$$\int_0^{+\infty} \sum_{i=0}^{L} c_i x_i(t) dt.$$

Above, and throughout, $c_L x_L(t)$ is a shorthand for $c_L^+ x_L^+(t) + c_L^- x_L^-(t)$, where $x_L^+ = \max(x_L, 0)$ and $x_L^- = \max(-x_L, 0)$. We shall denote the cost of a control u by $J(u)$.

Due to the complex nature of the system, including state constraints, discontinuous right sides, and an infinite horizon, we do not assume a-priori the existence of an optimal control law. Rather, in the sequel, we resolve the problem of existence of an optimal control law concurrently with its determination.

Before proceeding any further, it is important to note the following special case, for which the optimal control problem is trivially solved.

Case 0: All buffers are initially empty.

Suppose $x_i(0) = 0$ for $1 \leq i \leq L$. Then the optimal solution is $u_i(t) \equiv d$ for $0 \leq t < +\infty$ and $0 \leq i \leq L-1$. This maintains $x_i(t) \equiv 0$ for $1 \leq i \leq L$, $0 \leq t < +\infty$, and incurs zero cost, which is clearly optimal.

Let us say that a time T is a *system clearing time* if $x_i(T) = 0$ for $1 \leq i \leq L$. Then, by the Principle of Optimality (see Bellman [22]), we see that after a system clearing time the optimal solution is $u_i(t) \equiv d$, for $0 \leq i \leq L-1$ and all $t \geq T$, i.e., an optimal solution is to *pipeline*[1] all the machines at rate d.

3 The Initial Surplus Case: $x_L(0) \geq 0$

In the rest of the paper, we will analyze systems where the holding costs are non-decreasing, i.e., $0 =: c_0 \leq c_1 \leq \ldots \leq c_{L-1} \leq c_L^+$. This assumption is natural since holding costs usually

[1] If $u_i(t) = u_{i-1}(t) > 0$ and $x_i(t) = 0$, we say that machine M_i is "pipelining" material.

increase as the "value added" increases. We will also see that the problem is more tractable under this assumption.

The form of the optimal control is particularly easy to deduce when $x_L(0) \geq 0$, which we shall refer to as the "initial surplus" case. As the following theorem shows, an optimal policy is for each machine to remain idle until all the buffers downstream from it are empty, and to then process material at rate d, i.e., each machine simply adopts a just–in–time production policy.

Theorem 1 *If $x_L(0) \geq 0$, an optimal control at machine M_i is*

$$u_i(t) = \begin{cases} d & \text{if } x_j(t) = 0 \text{ for all } j > i, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Proof: Clearly, $T = 1/d \sum_{i=1}^{L} x_i(0)$ is the minimum possible value of the system clearing time. The proposed policy indeed clears the system at time T. After time T, the policy pipelines material at rate d from source to output, thus incurring no further cost, since all buffer levels thereafter remain at 0. Now consider an infinitesimal unit of material, Δ, in the system at time 0. Δ cannot leave the system until the material ahead of it is cleared, which occurs at rate d. Thus, the policy clears each unit of material as fast as possible from the system. Since each unit only incurs a cost from being in its original buffer, and buffer costs are non-decreasing, it follows that the proposed policy incurs the minimum cost. □

For the initial surplus case, Theorem 1 extends to "re-entrant lines" of the type shown in Figure 2, i.e., systems where the parts may require processing more than once on a given machine (also see Section 9). This is because one can actually use the same just–in–time policy as for the tandem case.

Theorem 2 *Consider a re-entrant line, such as the system shown in Figure 2. If $x_L(0) \geq 0$, an optimal control at buffer b_i is given by (1).*

If $\mathcal{G} \subseteq \mathcal{F}$ is a class of controls such that for every $u \in \mathcal{F}$ there exists a control $\bar{u} \in \mathcal{G}$ with $J(\bar{u}) \leq J(u)$, then we shall say that $\mathcal{G}$ *dominates* $\mathcal{F}$. Our analysis in the sequel will proceed by identifying a decreasing sequence of classes $\mathcal{F}_i$ which dominate $\mathcal{F}$.

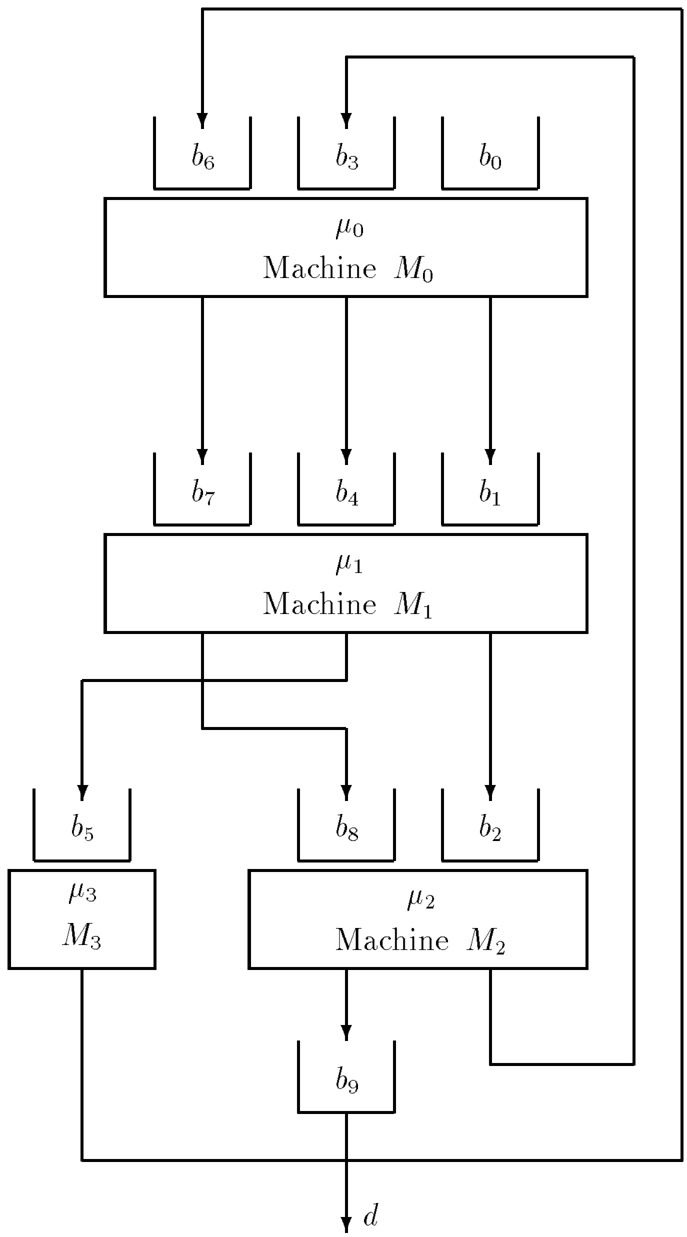

Figure 2: A re-entrant line.

Let $\mathcal{F}_0 \subseteq \mathcal{F}$ be the class of controls u such that,

$$u_i(t) = \begin{cases} d & \text{if } x_j(t) = 0 \text{ for all } j > i, \\ 0 & \text{if } x_j(t) \geq 0 \text{ for all } j > i, \text{ with } x_j(t) > 0 \text{ for some } j > i. \end{cases}$$

The class $\mathcal{F}_0$ specifies the future control fully after x_L becomes non-negative. Theorem 1 says that $\mathcal{F}_0$ dominates $\mathcal{F}$.

4 The Initial Shortfall Case: $x_L(0) < 0$

This case is considerably more complex. There is a tradeoff between producing and not producing. By producing, a machine is able to erase the shortfall of the output virtual buffer, and thus lower shortfall costs. However, by producing, a machine transfers its buffer contents to the next downstream machine, which incurs higher holding costs. Thus, while it is important to keep material in the earlier buffers as long as possible because of the increasing costs, as in the previous case, it is also important to "put out the fire," caused by the output buffer level being negative. The optimal policy must strike a balance between these two conflicting objectives.

Define T^u to be the first time at which the output virtual buffer b_L no longer has a shortfall, when the control $u(t)$ is implemented. That is,

$$T^u := \inf\{\sigma \geq 0 : x_L(\sigma) \geq 0\}. \tag{2}$$

We allow for the possibility that T^u may be infinity. We will call T^u the *shortfall erasure time* under control u. By the Principle of Optimality, an optimal solution for all $t \geq T^u$ is given by Theorem 1, i.e., as restricted by $\mathcal{F}_0$. Thus, we are only interested in determining an optimal control for $t < T^u$.

Given an initial shortfall $x_L(0) < 0$ and a control $u \in \mathcal{F}$, we can determine the number of buffers that are needed to eliminate the shortfall of the output virtual buffer. Define

$$i(u) := \max\{0 \leq i \leq L-1 : \sum_{j=i}^{L-1} x_j(0) \geq -x_L(0) + T^u d \text{ under the control } u\}.$$

If $T^u = +\infty$, then $i(u) := 0$. We will call $i(u)$ the *shortfall erasing buffer* under control u. Thus, the control $u(t)$ can use the material which is initially in buffers $b_{i(u)}, \ldots, b_{L-1}$ to eventually erase the shortage at b_L. Also, define $y_{i(u)}$ to be the amount of material in buffer $b_{i(u)}$ which is necessary to do this, i.e.,

$$y_{i(u)} := -x_L(0) + T^u\, d - \sum_{j=i(u)+1}^{L-1} x_j(0).$$

Note that if $T^u = +\infty$, then $y_{i(u)} = +\infty$. We shall call the material $y_{i(u)}$ in buffer $b_{i(u)}$ at time 0, and all the material downstream of it in buffers $b_{i(u)+1}, \ldots, b_{L-1}$ at time 0, as the *material needed to erase the shortfall* in b_L under u.

As the following lemma shows, before time T^u, there is no need to process any material which will not be used to erase the shortfall.

Lemma 1 *Let $\mathcal{F}_1 \subseteq \mathcal{F}_0$ be the subclass of controls u with the following property: If $x_L(0) < 0$, then $u_i(t) = 0$ for $i < i(u)$ and $0 \le t < T^u$, and $\int_0^{T^u} u_{i(u)}(\sigma)d\sigma = y_{i(u)}$. In words, prior to time T^u, controls in $\mathcal{F}_1$ only process material needed to erase the shortfall. After T^u, they follow the optimal policy of Theorem 1. Then, $\mathcal{F}_1$ dominates $\mathcal{F}$.*

Proof: By Theorem 1, for any control $\bar{u} \in \mathcal{F}$ there exists a control $\tilde{u} \in \mathcal{F}_0$ which has lower (or equal) cost. For $x_L(0) < 0$, assume that under $\tilde{u}$, at machine M_i, $i < i(u)$, an amount Δ of material is processed which does not leave the system before time T^u. From Theorem 1, this material will leave the system at the same time under all controls in $\mathcal{F}_0$. Thus, since $c_j \ge c_i$ for all $j > i$, it is less expensive to keep the material in buffer b_i, until time T^u, and then follow the policy of Theorem 1. □

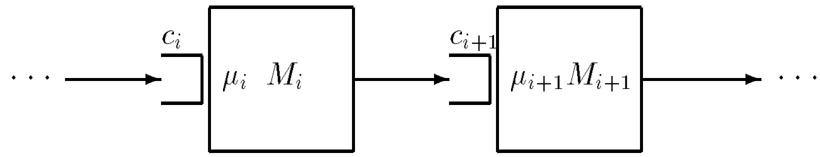

Figure 3: Two consecutive machines.

Consider any two consecutive machines, as shown in Figure 3. If machine M_i can process material as fast as (or faster than) machine M_{i+1}, then there is no need for it transfer the contents of its buffer b_i to buffer b_{i+1}, until the exact moment that the material will be processed by machine M_{i+1} as well, i.e., just–in–time.

Lemma 2 *Let $\mathcal{F}_2 \subseteq \mathcal{F}_1$ be the subclass of controls u with the following property: If $x_L(0) < 0$, then for every i such that $\mu_i \geq \mu_{i+1}$,*

$$u_i(t) = \begin{cases} 0 & 0 \leq t < t_i^u \\ u_{i+1}(t) & t_i^u \leq t, \end{cases} \tag{3}$$

and

$$t_i^u = \inf\{v \geq 0 : \int_0^v u_{i+1}(\sigma)\, d\sigma = x_{i+1}(0)\}. \tag{4}$$

Then, $\mathcal{F}_2$ dominates $\mathcal{F}$.

Proof: For any control $\bar{u} \in \mathcal{F}$, consider a control $u = (u_0(t), \ldots, u_{L-1}(t))^T \in \mathcal{F}_1$, which has lower (or equal) cost. (From Lemma 1, this is always possible.) Consider any i such that $\mu_i \geq \mu_{i+1}$. Define $\tilde{u}(t) = (u_0(t), \ldots, u_{i-1}(t), \tilde{u}_i(t), u_{i+1}(t), \ldots, u_{L-1}(t))^T$, where

$$\tilde{u}_i(t) = \begin{cases} 0 & 0 \leq t < t_i^{\tilde{u}} \\ u_{i+1}(t) & t_i^{\tilde{u}} \leq t, \end{cases}$$

and $t_i^{\tilde{u}}$ is defined as in (4), for the control $\tilde{u}$. That is, $t_i^{\tilde{u}}$ is the first time that b_{i+1} *would* empty, under $u(t)$, if no material were allowed to flow into it. Then, under $\tilde{u}(t)$, b_{i+1} empties for the first time at time $t_i^{\tilde{u}}$, and it will remain empty thereafter. (Note that $\tilde{u}(t)$ may not be an element of $\mathcal{F}_2$, since we have not yet ensured that the properties required by $\mathcal{F}_2$ are met at the buffers other than buffer b_i.) Since $\tilde{u}(t)$ feeds buffer b_{i+1} as *slowly* as possible, while still ensuring that $u_{i+1}(t)$ is feasible, the *cumulative* amount of material processed by machine M_i under $\tilde{u}(t)$ must be less than (or equal to) that processed under $u(t)$, i.e., $\int_0^t [\tilde{u}_i(\sigma) - u_i(\sigma)] d\sigma \leq 0$, for all t.

Thus, since $u(t)$ and $\tilde{u}(t)$ are identical at all machines except machine M_i and $c_i \leq c_{i+1}$, it follows that $J(u) - J(\tilde{u}) = (c_i - c_{i+1}) \int_0^{+\infty} \int_0^t [\tilde{u}_i(\sigma) - u_i(\sigma)] d\sigma\, dt \geq 0$. By applying this argument for each i such that $\mu_i \geq \mu_{i+1}$, starting from the downstream end of the system, we construct a $\hat{u} \in \mathcal{F}_2$ such that $J(\hat{u}) \leq J(\bar{u})$. $\square$

Define

$$u_i^{\max} := \sup_{t \geq 0} u_i(t).$$

Thus, $u_i^{\max}$ is the maximum instantaneous rate at which machine M_i processes material under control u. Clearly, given $\{u_{i+1}(t) : t \geq 0\}$, the result of Lemma 2 remains valid as long as $\mu_i \geq u_{i+1}^{max}$.

Lemma 3 *Let $\mathcal{F}_3 \subseteq \mathcal{F}_2$ be the subclass of controls u with the following property: If $x_L(0) < 0$, then for every i such that $\mu_i \geq u_{i+1}^{\max}$, (3) holds, where t_i^u is again determined by (4). Then, $\mathcal{F}_3$ dominates $\mathcal{F}$.*

5 The Sections

In order to extend Lemma 3 to "sections" containing more than two machines, we partition the set of machines $\mathcal{M} := \{M_0, M_1, \ldots, M_{L-1}\}$ into $N\,(\leq L)$ *sections* as follows:

$$S_1 = \{M_0, \ldots, M_{s_1}\}, S_2 = \{M_{s_1+1}, \ldots, M_{s_2}\}, \ldots, S_N = \{M_{s_{N-1}+1}, \ldots, M_{s_N}\}$$

where $s_N := L - 1$, and s_{i-1} is recursively defined by

$$s_{i-1} := \max\{0 \leq j < s_i : \mu_j < \mu_{s_i}\}.$$

Note that, in each section S_i, the bottleneck is the most downstream machine M_{s_i} with the smallest processing rate μ_{s_i}. Also note that machine M_{s_i} is the bottleneck for its downstream machines, in the sense of feeding them at the rate that they could work. For convenience, we define

$$\bar{\mu}_j := \mu_{s_i} \text{ for all machines } M_j \text{ in section } S_i. \tag{5}$$

As the following lemma shows, the upstream machines in section S_i should not produce until all the downstream buffers in section S_i have emptied, and should then produce at a rate matching the rate of machine M_{s_i}, i.e., just–in–time.

Lemma 4 *Let $\mathcal{F}_4 \subseteq \mathcal{F}_3$ be the subclass of controls u with the following property: If $x_L(0) < 0$, then*

$$u_j(t) = \begin{cases} 0 & 0 \leq t < t_j^u \\ u_{s_i}(t) & t_j^u \leq t, \end{cases}$$

for every machine M_j (except the head machine M_{s_i}) in section S_i. Above, as in (4),

$$t_j^u = \inf\{\sigma \geq 0 : x_{j+1}(\sigma) = 0 \textit{ under the control } u\},$$

where $x(t) = (x_1(t), \ldots, x_L(t))^T$ is the vector of buffer levels resulting from $u(t)$. Then, $\mathcal{F}_4$ dominates $\mathcal{F}$.

Proof: The result is obtained by a repeated application of Lemma 3 to machines $M_{s_i-1}, M_{s_i-2}, \ldots, M_{s_{i-1}+1}$ in each section S_i. □

6 The Bottleneck Machines

Note that controls in $\mathcal{F}_4$ are completely characterized by the production rates of the bottleneck machines $M_{s_1}, M_{s_2}, \ldots, M_{s_N}$. Therefore, it only remains for us to determine the rates $(u_{s_1}(t), \ldots, u_{s_N}(t))$ at these bottleneck machines.

As the following lemma shows, once all the buffers (excluding b_L) downstream from any section are empty, they will remain empty.

Lemma 5 *Let $\mathcal{F}_5 \subseteq \mathcal{F}_4$ be the subclass of controls u with the following property: If, at some time $v \geq 0$,*

$$\sum_{k=s_i+1}^{L-1} x_k(v) = 0, \tag{6}$$

then $\sum_{k=s_i+1}^{L-1} x_k(t) = 0$ for all $t \geq v$. Then, $\mathcal{F}_5$ dominates $\mathcal{F}$.

Proof: For any control $\bar{u} \in \mathcal{F}$, consider a control $u \in \mathcal{F}_4$, which has lower (or equal) cost. Consider an i for which (6) holds, letting v be the the first time at which (6) holds. Define a control $\tilde{u} \in \mathcal{F}_4$, which is the same as u at machine M_j for $0 \leq j \leq s_i$. For $t < v$, also let the controls under $\tilde{u}(t)$ at all the machines downstream from M_{s_i} be equal to those under $u(t)$. Thus, the total buffer holding cost is the same for both controls for $t < v$. For $t \geq v$, define the controls under $\tilde{u}(t)$ at the machines downstream from M_{s_i} such that material coming into buffer b_{s_i+1} is pipelined directly to the output virtual buffer at rate $u_{s_i}(t)$. Note that this can be done since $\mu_j > \mu_{s_i}$ for $s_i + 1 \leq j \leq L-1$. That is, the buffers $b_{s_i+1}, \ldots, b_{L-1}$ remain empty for $t \geq v$, thus incurring no additional cost. Also, since $\int_v^t [\tilde{u}_{L-1}(\sigma) - u_{L-1}(\sigma)]d\sigma \geq 0$, the level of the output virtual buffer b_L is less negative under control $\tilde{u}$ than it is under control u. Thus, $J(\tilde{u}) \leq J(u)$. Using this procedure for sections $S_N, S_{N-1}, \ldots, S_1$ produces the desired $\hat{u} \in \mathcal{F}_5$, with $J(\hat{u}) \leq J(\bar{u})$. □

If b_L has a shortfall, and buffers $b_{L-1}, b_{L-2}, \ldots, b_{s_i+1}$ are empty, but section S_i contains material, then machine M_{s_i} should produce at its maximum rate μ_{s_i}, as the following lemma shows.

Lemma 6 *Let $\mathcal{F}_6 \subseteq \mathcal{F}_5$ be the subclass of controls u with the following property: If, at some time $v \geq 0$, $x_L(v) < 0$,*

$$\sum_{k=s_i+1}^{L-1} x_k(v) = 0, \tag{7}$$

and

$$\sum_{k=s_{i-1}+1}^{s_i} x_k(v) > 0, \tag{8}$$

where $s_0 := -1$, then $u_{s_i}(v) = u_{s_{i+1}}(v) = \ldots = u_{s_N}(v) = \mu_{s_i}$. (For $i = N$, conditions (7) and (8) are replaced by just (8) above.) Then, $\mathcal{F}_6$ dominates $\mathcal{F}$.

Proof: For any control $\bar{u} \in \mathcal{F}$, consider a control $u \in \mathcal{F}_5$, which has lower (or equal) cost. Consider the first time instant v at which buffers $b_{j+1}, \ldots, b_{L-1}$ are empty, but b_j is not empty, for some j. By Lemmas 4 and 5, any material processed by machine M_j will be pipelined to the output virtual buffer b_L. Define $\tilde{u}(t) := (u_0(t), \ldots, u_{j-1}(t), \tilde{u}_j(t), \ldots, \tilde{u}_{L-1}(t))^T \in \mathcal{F}_5$ by

$$\tilde{u}_j(t) = \begin{cases} u_j(t) & 0 \leq t < v \\ \mu_{s_i} & v \leq t < \tau \\ u_{j-1}(t) & \tau \leq t, \end{cases}$$

where $\tau := \inf\{\sigma > v : \tilde{x}_j(\sigma) = 0$, where $\tilde{x}_j(\sigma)$ is the buffer level at time σ under control $\tilde{u}\}$, and, for $j+1 \leq k \leq L-1$,

$$\tilde{u}_k(t) = \begin{cases} u_k(t) & 0 \leq t < v \\ \tilde{u}_j(t) & v \leq t. \end{cases}$$

Then, the buffer levels at all buffers, except buffer b_j and the output virtual buffer, are the same under both $\tilde{u}$ and u. Since, for all $t \geq 0$, $\int_0^t [\tilde{u}_j(\sigma) - u_j(\sigma)]d\sigma \geq 0$, and $\int_0^t [\tilde{u}_{L-1}(\sigma) -$

$u_{L-1}(\sigma)]d\sigma \geq 0$, it follows that $J(\tilde{u}) \leq J(u)$. Now if $\tilde{u}$ does not belong to $\mathcal{F}_6$, then there is some time v' and a $j' < j$ at which buffers $b_{j'+1}, \ldots, b_{L-1}$ are empty, but $b_{j'}$ is not. We can then repeat the above argument to obtain an improved control. In this way, we construct the desired control $\hat{u} \in \mathcal{F}_6$, with $J(\hat{u}) \leq J(\bar{u})$. □

We can now show that the shortfall erasure time T^u, defined in (2), is finite and bounded for all controls $u \in \mathcal{F}_6$. Also, we can thereby obtain a lower bound on the index $i(u)$ of the shortfall clearing buffer $b_{i(u)}$.

Lemma 7 *Let $x_L(0) < 0$. For any control $u \in \mathcal{F}_6$,*

$$T^u \leq T_{\max}$$

and

$$i(u) \geq i_{\min}.$$

Above,

$$i_{\min} := \max\{0 \leq i \leq L-1 : \sum_{j=i}^{L-1} x_j(0) \geq -x_L(0) + d \sum_{j=i}^{L-1} \frac{x_j(0)}{\bar{\mu}_j}\},$$

where $\bar{\mu}_j$ is defined in (5), and

$$T_{\max} = \frac{\sum_{j=i_{\min}+1}^{L-1}[\frac{\bar{\mu}_{i_{\min}}}{\bar{\mu}_j} - 1]x_j(0) - x_L(0)}{\bar{\mu}_{i_{\min}} - d}.$$

Proof: From Lemma 6, under any control in $\mathcal{F}_6$, the most downstream machine with a positive amount of material in its buffer, say $M_j \in S_i$, pipelines material at rate μ_{s_i} to the output virtual buffer b_L, until the shortfall erasure time. Thus, the control $\bar{u} \in \mathcal{F}_6$ that takes the longest time to erase the shortfall in b_L is the one that pipelines material to b_L, from the most downstream section which is not empty, while keeping idle all its upstream sections. Thus, the buffers will empty one at a time from downstream to upstream, i.e.,

$$\bar{u}_{s_i}(t) = \begin{cases} 0 & 0 \leq t < T_{i+1}^{\bar{u}} \\ \mu_{s_i} & T_{i+1}^{\bar{u}} \leq t < T_i^{\bar{u}} \\ \mu_{s_{i-1}} & T_i^{\bar{u}} \leq t < T_{i-1}^{\bar{u}} \\ . & \\ . & \\ \mu_{s_l} & T_{l+1}^{\bar{u}} \leq t < T_{\max} \\ d & T_{\max} \leq t. \end{cases}$$

Above, $T_i^{\bar{u}} = \sum_{j=i}^{N} \frac{\sum_{k=s_{j-1}+1}^{s_j} x_k(0)}{\mu_{s_j}}$, and l is such that $M_{i_{\min}} \in S_l$. It is straightforward to verify that this control attains $T^{\bar{u}} = T_{\max}$ and $i(\bar{u}) = i_{\min}$. All other controls u in $\mathcal{F}_6$ will have smaller T^u and larger $i(u)$. □

From Lemma 6, it also follows that the final machine will process any material in its section as fast as possible until T^u.

Lemma 8 *Let* $T_N^u := \inf\{t \geq 0 : x_{s_{N-1}+1}(t) = x_{s_{N-1}+2}(t) = \ldots = x_{L-1}(t) = 0\} \wedge T^u$, *where "$\wedge$" denotes the "min" operation. For any control in* $\mathcal{F}_6$, *if* $x_L(0) < 0$, *then*

$$u_{L-1}(t) = \begin{cases} \mu_{s_N} & 0 \leq t < T_N^u \\ u_{s_{N-1}}(t) & T_N^u \leq t < T^u \\ d & T^u \leq t. \end{cases}$$

In the special case that the final machine cannot produce faster than *any* of the other machines, i.e., $\mu_i \geq \mu_{L-1}$ for all $0 \leq i \leq L-2$, there is only one section. An optimal control has been determined for this case, since $\mathcal{F}_6$ is a singleton set consisting of only one control.

Theorem 3 Optimal Control When The Final Machine Is The Bottleneck

For an L-machine flow shop with non-decreasing buffer costs (i.e., $c_1 \leq c_2 \ldots \leq c_L^+$*) and all production rates at least as large as* μ_{L-1} *(i.e.,* $\mu_i \geq \mu_{L-1}$ *for all i), an optimal control is the following: At machine* M_i,

$$u_i^*(t) = \begin{cases} 0 & 0 \leq t < t_i^* \\ \mu_{L-1} & t_i^* \leq t < T^* \\ d & T^* \leq t, \end{cases}$$

where

$$T^* := \max\{\frac{-x_0(0)}{\mu_{L-1} - d}, 0\},$$

and

$$t_i^* := \min\{\frac{1}{\mu_{L-1}} \sum_{j=i+1}^{L-1} x_j(0), T^*\}.$$

Proof: The result follows since $\mathcal{F}_6$ consists of only the control u^* defined above. Note that the "max" in T^* is needed to include the case $x_L(0) > 0$. □

7 The Deferral Times for the Bottleneck Machines: The Form of an Optimal Control

We now turn to the general case, where there is more than one section S_i. As the following lemma shows, the first machine in each section is initially idle, and then it processes material as fast as it can, subject to its maximum processing rate, or if its buffer level is zero, at a rate matching its input, until there is no longer a shortfall at the output virtual buffer. We will call the length of time for which the first machine in each section is initially idle as the *deferral time* of that section. For convenience, let us define $l(u)$ as the index of the section that contains $M_{i(u)}$, i.e., $M_{i(u)} \in S_{l(u)}$.

Lemma 9 *Let $\mathcal{F}_7 \subseteq \mathcal{F}_6$ be the subclass of controls u with the following property: There exists a set of "deferral times" $\{\tau^u_{s_{l(u)+1}}, \tau^u_{s_{l(u)+2}}, \ldots, \tau^u_{s_N}\}$, and, for $i = l(u)+1, \ldots, N-1$,*

$$u_{s_i}(t) = \begin{cases} 0 & 0 \le t < \tau^u_{s_i} \\ \mu_{s_i} & \tau^u_{s_i} \le t < T^u_i \\ u_{s_{i-1}}(t) & T^u_i \le t < T^u \\ d & T^u \le t. \end{cases}$$

Above,

$$T^u_i := \inf\{t \ge \tau^u_{s_i} : x_{s_{i-1}+1}(t) = x_{s_{i-1}+2}(t) = \cdots = x_{s_i}(t) = 0\} \wedge T^u. \qquad (9)$$

Then, $\mathcal{F}_7$ dominates $\mathcal{F}$.

Proof: For any control $\bar{u} \in \mathcal{F}$, consider a control $u \in \mathcal{F}_6$, which has lower (or equal) cost. We will use induction. If $l(u) = N-1$, the result is given by Lemma 8. So assume $l(u) \ne N-1$. From Lemma 8,

$$u_{s_N}(t) = \begin{cases} \mu_{s_N} & 0 \le t < T^u_N \\ u_{s_{N-1}}(t) & T^u_N \le t < T^u \\ d & T^u \le t. \end{cases} \qquad (10)$$

Consider $i = N-1 > l(u)$. If $\sum_{j=s_{N-1}+1}^{s_N} x_j(0) = 0$,, then, by Lemma 6, $T^u_N = 0$ and $u_{s_{N-1}}(t)$ has the desired form. Otherwise, from (10), $T^u_N = \frac{\sum_{j=s_{N-1}+1}^{s_N} x_j(0) + \int_0^{T^u_N} u_{s_{N-1}}(\sigma)d\sigma}{\mu_{s_N}}$, noting that, for $j \geq s_{i(u)}+1, T_j^u \leq T^u$. Define a control $\tilde{u}(t) \in \mathcal{F}_6$ such that $\tilde{u}_j(t) = u_j(t)$ for $M_j \notin S_{N-1}$, and

$$\tilde{u}_{s_{N-1}}(t) = \begin{cases} 0 & 0 \leq t < t^{\tilde{u}}_{s_{N-1}} \\ \mu_{s_{N-1}} & t^{\tilde{u}}_{s_{N-1}} \leq t < T^u_N \\ u_{s_{N-1}}(t) & T^u_N \leq t, \end{cases} \tag{11}$$

where the time $t^{\tilde{u}}_{s_{N-1}}$ is given by $t^{\tilde{u}}_{s_{N-1}} = T^u_N - \frac{\int_0^{T^u_N} u_{s_{N-1}}(\sigma)d\sigma}{\mu_{s_{N-1}}}$. Thus, $\tilde{u}$ holds material in the buffers in section S_{N-1} as long as possible, while guaranteeing that all the buffers in section S_N still empty (for the first time) at T^u_N. Also, from (11), it is clear that $T^{\tilde{u}}_N = T^u_N$, and that section S_{N-1} does not empty (after $t^{\tilde{u}}_{s_{N-1}}$) before T^u_N, i.e.,

$$T^{\tilde{u}}_{N-1} \geq T^{\tilde{u}}_N = T^u_N. \tag{12}$$

Also, since $\tilde{u} \in \mathcal{F}_6$, it follows that

$$\tilde{u}_{s_{N-1}}(t) = \begin{cases} 0 & 0 \leq t \leq t^{\tilde{u}}_{s_{N-1}} \\ \mu_{s_{N-1}} & t^{\tilde{u}}_{s_{N-1}} \leq t < T^u_{N-1} \\ u_{s_{N-2}}(t) & T^u_{N-1} \leq t < T^u \\ d & T^u \leq t. \end{cases} \tag{13}$$

Thus, from (12) and (13), the lemma holds for $i = N-1$.

Now suppose that the lemma holds for $i = k+1, \ldots, N-1$. Using the same argument as above, one can defer production at machine M_{s_k} until a time $t^{\tilde{u}}_k$, and then produce at full rate μ_{s_k} to still clear section S_k at time T^u_{k+1}. Thus one can construct a control $\tilde{u}$ such that the lemma is satisfied for $i = k$. By completing the induction for all $i < l(u)$, we construct the desired control $\hat{u} \in \mathcal{F}_7$, with $J(\hat{u}) \leq J(\bar{u})$. □

As shown in the following lemma, given T^u, the first machine of the section that contains machine $M_{i(u)}$ should wait as long as possible before beginning to process material.

Lemma 10 *Let $\mathcal{F}_8 \subseteq \mathcal{F}_7$ be the subclass of controls u satisfying*

$$u_{s_{l(u)}}(t) = \begin{cases} 0 & 0 \leq t < t^u_{s_{l(u)}} \\ \mu_{s_{l(u)}} & t^u_{s_{l(u)}} \leq t < T^u \\ d & T^u \leq t. \end{cases}$$

Then, $\mathcal{F}_8$ dominates $\mathcal{F}$.

Proof: For any control $\bar{u} \in \mathcal{F}$, consider a control $u \in \mathcal{F}_7$, which has lower (or equal) cost. Given T^u, the control defined above holds the material for as long as possible in the buffers of section $S_{l(u)}$, while ensuring that the shortfall in the output virtual buffer is erased at time T^u. $\square$

The following theorem, which combines Lemmas 1-10, shows that any control in $\mathcal{F}_8$ can be characterized by its deferral times $\{\tau^u_{s_1}, \ldots, \tau^u_{s_N}\}$, where $\tau^u_{s_N} = 0$.

Theorem 4 *Consider $x_L(0) < 0$. Every $u \in \mathcal{F}_8$ has the form shown below, for some deferral times $\{\tau^u_{s_i} : 1 \le i \le N\}$. The controls $\{u_{s_i}(t) : i = 1, \ldots, N\}$ fall into three categories.*
For $1 \le i \le l(u) - 1$,

$$u_{s_i}(t) = \begin{cases} 0 & 0 \le t < \tau^u_{s_i} \\ d & \tau^u_{s_i} \le t, \end{cases}$$

where

$$\tau^u_{s_i} := \frac{\sum_{k=s_i+1}^{L} x_k(0)}{d} \text{ for } 1 \le i \le l(u) - 1.$$

For $i = l(u)$,

$$u_{s_{l(u)}}(t) = \begin{cases} 0 & 0 \le t < \tau^u_{s_{l(u)}} \\ \mu_{s_{l(u)}} & \tau^u_{s_{l(u)}} \le t < T^u \\ d & T^u \le t, \end{cases}$$

where T^u is given by (2).
For $l(u) < i \le N$,

$$u_{s_i}(t) = \begin{cases} 0 & 0 \le t < \tau^u_{s_i} \\ \mu_{s_i} & \tau^u_{s_i} \le t < T^u_i \\ \mu_{s_{i-1}} & T^u_i \le t < T^u_{i-1} \\ . & \\ . & \\ \mu_{s_{l(u)+1}} & T^u_{l(u)+2} \le t < T^u_{l(u)+1} \\ \mu_{s_{l(u)}} & T^u_{l(u)+1} \le t < T^u \\ d & T^u \le t, \end{cases}$$

where the T^u_i are given by (9).
For machine $M_j \in S_i$

$$u_j(t) = \begin{cases} 0 & 0 \le t < \tau^u_j \\ u_{s_i}(t) & \tau^u_j \le t, \end{cases}$$

where

$$\tau_j^u := \begin{cases} \tau_{s_i}^u + \frac{\sum_{k=j+1}^{s_i} x_k(0)}{\mu_{s_i}} & i(u) \le j \le L-1 \\ \tau_{s_i}^u + \frac{\sum_{k=j+1}^{s_i} x_k(0)}{d} & 0 \le j < i(u). \end{cases}$$

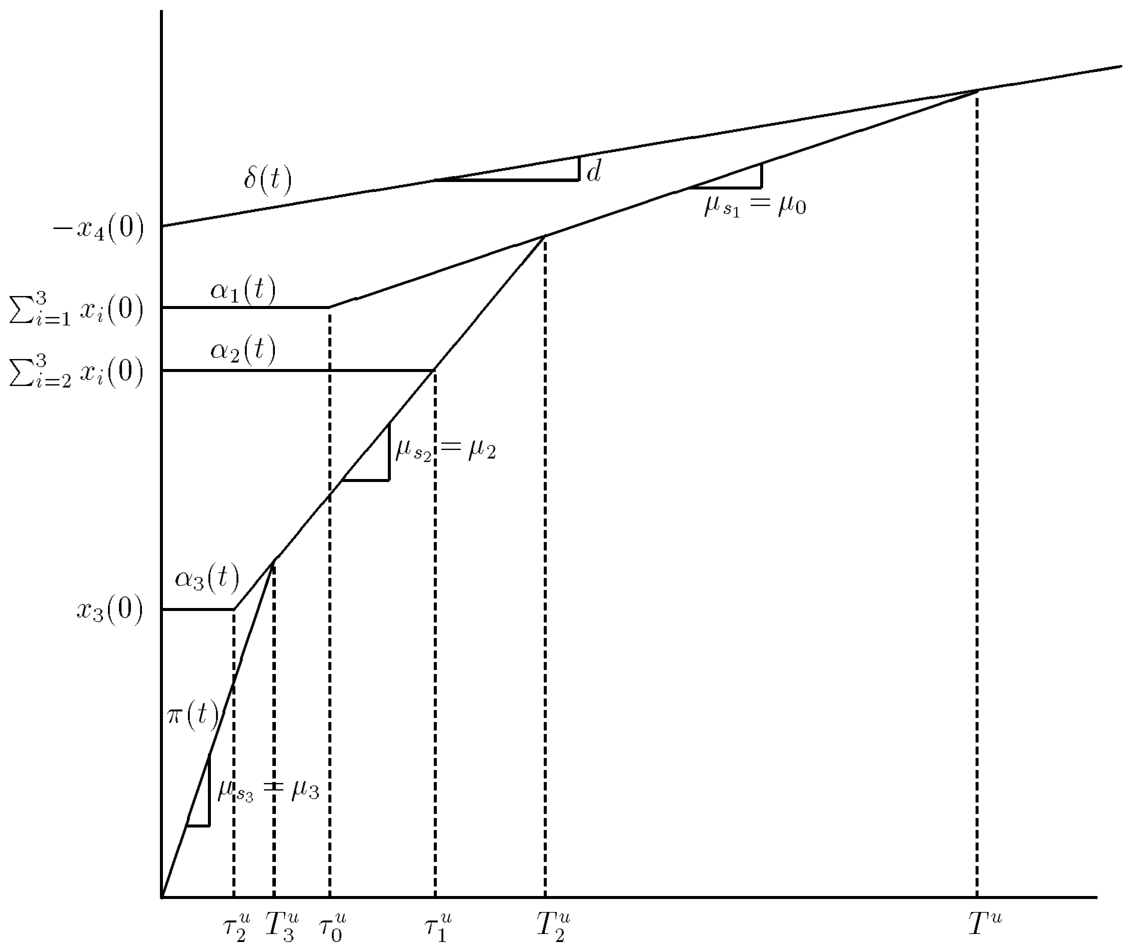

Figure 4: $\pi(t), \alpha_j(t)$, and $\delta(t)$ under a control $u \in \mathcal{F}_8$ for four-machine system of Example 1.

To describe the form of a control $u \in \mathcal{F}_8$, it is convenient to define the *cumulative demand*

$$\delta(t) := -x_L(0) + t\,d,$$

the *cumulative production*

$$\pi(t) := \int_0^t u_{L-1}(\sigma) d\sigma,$$

and the *effective input to the segment* $M_j, M_{j+1}, \ldots, M_{L-1}$ by

$$\alpha_j(t) := \sum_{k=j}^{L-1} x_k(0) + \int_0^t u_{j-1}(\sigma)d\sigma \text{ for } 1 \leq j \leq L-1.$$

Note that the derivatives of $\pi(t)$ and $\alpha_j(t)$ yield the production rates of the various machines. Note also that

$$\pi(t) \leq \alpha_{L-1}(t) \leq \alpha_{L-2}(t) \leq \cdots \leq \alpha_1(t) \leq \delta(t)$$

for every $t \geq 0$.

Example 1 *A four-machine system.*

Figure 4 shows these quantities for a control $u \in \mathcal{F}_8$, for a four-machine system with $\mu_0 < \mu_2 < \mu_3$ and $\mu_1 \geq \mu_2$. Note that $S_1 = \{M_0\}, S_2 = \{M_1, M_2\}$, and $S_3 = \{M_3\}$. □

For any control in $\mathcal{F}_8$, we can also determine a lower bound on how many machines will be used to eliminate the shortfall at the output buffer and on the time required to do this.

Lemma 11 *For any control* $u \in \mathcal{F}_8$*,*

$$T^u \geq T_{\min} \tag{14}$$

and

$$i(u) \leq i_{\max}, \tag{15}$$

where

$$i_{\max} := \max\{0 \leq i \leq L-1 : \sum_{j=i}^{L-1} x_j(0) \geq -x_L(0) + d \max_{i \leq k \leq L-1} \frac{\sum_{j=i}^{k} x_j(0)}{\bar{\mu}_k}\}$$

and $T_{\min} = -\max_{i_{\max} \leq k \leq L-1} \frac{\sum_{j=k+1}^{L} x_j(0)}{\bar{\mu}_k - d}$.

Proof: Suppose $i(u) = i$. Then some of the material in buffer b_i is sufficient to erase the shortfall. Necessarily, *all* the material in $\{b_i, b_{i+1}, \ldots, b_{L-1}\}$, processed as rapidly as it can be, can also erase the shortfall at the output virtual buffer b_L. This implies that $\sum_{j=i}^{L-1} x_j(0) \geq -x_L(0) + d \max_{i \leq k \leq L-1} \frac{\sum_{j=i}^{k} x_j(0)}{\bar{\mu}_k}$, from which the result (15) follows.

Now suppose that y_i is the amount of material from buffer b_i which is used to erase the shortfall. Then

$$T^u \geq \max_{i \leq k \leq L-1} \frac{\sum_{j=i+1}^{k} x_j(0) + y_i}{\bar{\mu}_k} \tag{16}$$

and

$$\sum_{j=i+1}^{L-1} x_j(0) + y_i = -x_L(0) + dT^u. \tag{17}$$

Substituting for y_i from (17) into (16) gives $T^u \geq \max_{i \leq k \leq L-1} \frac{-\sum_{j=k+1}^{L} x_j(0) + dT^u}{\bar{\mu}_k}$. Using $d < \bar{\mu}_k$ and $i \leq i_{\max}$, yields (14). □

Consider a vector $(i, T, T_N, T_{N-1}, \ldots, T_{l(u)+1})$ where $l(u)$ is such that $M_i \in S_{l(u)}$. The following lemma shows the necessary and sufficient conditions to be satisfied by this vector such that it corresponds to a control $u \in \mathcal{F}_8$.

Lemma 12 *Given $(i, T, T_N, T_{N-1}, \ldots, T_{l(u)+1})$ with $l(u)$ such that $M_i \in S_{l(u)}$, there exists a control $u \in \mathcal{F}_8$ with $i(u) = i$, $T^u = T$ and $T_j^u = T_j$ for $l(u) + 1 \leq j \leq N$, if and only if the following conditions are satisfied:*

$$0 \leq T_N \leq T_{N-1} \quad \leq \cdots \leq \quad T_{l(u)+1} \leq T \tag{18}$$

$$0 \leq \tau_{s_j} \leq T_{j+1} \quad \text{for} \quad l(u) \leq j \leq N-1 \tag{19}$$

$$\mu_{s_{l(u)}}(T - \tau_{s_{l(u)}}) \quad \leq \quad X_{l(u)}(0) \tag{20}$$

$$\sum_{j=l(u)+1}^{N} (T_j - T_{j+1})\mu_{s_j} \quad + \quad (T - T_{l(u)+1})\mu_{s_{l(u)}} = -x_L(0) + Td \tag{21}$$

where $T_{N+1} := 0$,

$$X_j(0) \;:=\; \sum_{k=s_{j-1}+1}^{s_j} x_k(0) \; \textit{for } l(u) \leq j \leq N, \tag{22}$$

$$\tau_{s_{N-1}} \;:=\; \frac{T_N(\mu_{s_{N-1}} - \mu_{s_N}) + X_N(0)}{\mu_{s_{N-1}}}, \tag{23}$$

and

$$\tau_{s_j} \;:=\; \frac{T_{j+1}(\mu_{s_j} - \mu_{s_{j+1}}) + \tau_{s_{j+1}}\mu_{s_{j+1}} + X_{j+1}(0)}{\mu_{s_j}} \; \textit{for } l(u) \leq j \leq N-2. \tag{24}$$

Proof: First we prove necessity. Suppose there is such a control $u \in \mathcal{F}_8$. Clearly, (18) is satisfied. Let us determine the condition to be satisfied by T_N for it to be a clearing time for section S_N. Note first that machine $M_{s_{N-1}}$ produces at its maximum rate $\mu_{s_{N-1}}$ in the time interval $[\tau^u_{N-1}, T_{N-1})$, and therefore also in the sub-interval $[\tau^u_{N-1}, T_N)$. Thus, for section S_N to clear at T_N, we will need to have,

$$(T_N - \tau^u_{s_{N-1}})\mu_{s_{N-1}} + X_N(0) \quad = \quad T_N \mu_{s_N}. \tag{25}$$

This shows that $\tau^u_{s_{N-1}}$ will have to equal $\tau_{s_{N-1}}$ given in (23). Moreover $\tau^u_{s_{N-1}}$ clearly has to satisfy (19) for $j = N-1$.

Consider section S_{j+1} for $l(u) \leq j \leq N-2$. Machine M_{s_j} produces at rate μ_{s_j} in the interval $[\tau^u_{s_j}, T_j)$, and hence also in the subinterval $[\tau^u_{s_j}, T_{j+1})$. Thus, for section S_{j+1} to clear at T_{j+1}, we will need to have

$$(T_{j+1} - \tau^u_{s_j})\mu_{s_j} + X_{j+1}(0) \quad = \quad (T_{j+1} - \tau^u_{s_{j+1}})\mu_{s_{j+1}}. \tag{26}$$

Hence $\tau^u_{s_j}$ will have to equal τ_{s_j} defined in (23), and will moreover have to satisfy (19).

Now we turn to section $S_{l(u)}$. First, since machine $M_{s_{l(u)}}$ produces continuously in the interval $[\tau^u_{s_{l(u)}}, T)$ at rate $\mu_{s_{l(u)}}$, without receiving any input from section $S_{l(u)-1}$, we see that (20) will have to hold.

Finally, for T to be the shortfall erasure time, we must have (21). This completes the proof of necessity of (18)-(24).

Now we prove the sufficiency of (18)-(24) by constructing a control $u \in \mathcal{F}_8$. Let machine $M_{s_{l(u)}}$ produce continuously in $[\tau_{s_{l(u)}}, T)$ at rate $\mu_{s_{l(u)}}$. This is feasible, without receiving any input from section $S_{l(u)-1}$, due to (20), by using pipelining within section $S_{l(u)}$.

Next, let machine $M_{l(u)+1}$ commence production at $\tau_{s_{l(u)+1}}$, and then produce continuously at rate $\mu_{s_{l(u)+1}}$ in $[\tau_{s_{l(u)+1}}, T_{l(u)+1})$ and at rate $\mu_{s_{l(u)}}$ in $[T_{l(u)+1}, T)$. To see that this is feasible, note first that machine $M_{s_{l(u)}}$ commences production at rate $\mu_{s_{l(u)}}$ at time $\tau_{s_{l(u)}} \leq T_{l(u)+1}$, by (19). The initial amount $X_{l(u)+1}(0)$ in section $S_{l(u)+1}$ will be depleted at a time $\tau_{s_{l(u)+1}} + \frac{X_{l(u)+1}}{\mu_{s_{l(u)+1}}}$, if the machines in section $S_{l(u)+1}$ pipeline material. However, $\tau_{s_{l(u)}} \leq \tau_{s_{l(u)+1}} +$

$\frac{X_{l(u)+1}(0)}{\mu_{s_{l(u)+1}}}$, as can be seen by solving for $T_{l(u)+1}$ in (24) and substituting it in the second inequality in (19). Thus machine $M_{s_{l(u)}}$ commences production before section $S_{l(u)+1}$ clears, and the clearing time for section $S_{l(u)+1}$ is consequently given by $T_{l(u)+1}$ satisfying (24). So machine $M_{s_{l(u)+1}}$ can produce at rate $\mu_{s_{l(u)+1}}$ in $[\tau_{s_{l(u)+1}}, T_{l(u)+1})$, and thereafter it can pipeline from section $S_{l(u)}$ at rate $\mu_{s_{l(u)}}$, as specified.

Now suppose that $T_{l(u)+1}, T_{l(u)+2}, \ldots, T_{j-1}$ are the clearing times, respectively, for sections $S_{l(u)+1}$, $S_{l(u)+2}, \cdots, S_{j-1}$, and that machines $M_{l(u)+1}, \ldots, M_{j-1}$ produce with deferral times $\tau_{l(u)+1}, \ldots, \tau_{j-1}$, as controls in $\mathcal{F}_8$ do. We will proceed by induction and consider machine M_{s_j} and section S_j.

We wish to show machine M_{s_j} can start producing at τ_{s_j}, produce at rate μ_{s_j} in $[\tau_{s_j}, T_j)$ clearing its section at T_j, and thereafter pipeline from section S_{j-1}. The material $X_j(0)$ in section S_j would be depleted at time $\tau_{s_j} + \frac{X_j(0)}{\mu_{s_j}}$. However machine $M_{s_{j-1}}$ produces at rate $\mu_{s_{j-1}}$ in $[\tau_{s_{j-1}}, T_{j-1})$, and that section S_j will clear at $T_j \leq T_{j-1}$. This completes the induction, where one notes that for s_N, the deferral time is $\tau_{s_N} = 0$.

Finally, from (21) it follows that T is indeed the shortfall erasure time. □

8 Optimal Control Through Quadratic Programming

We first calculate the cost of a control $u \in \mathcal{F}_8$, parameterized by $(i, T, T_N, T_{N-1}, \ldots, T_{l(u)})$.

Lemma 13 *Consider the control $u \in \mathcal{F}_8$ parameterized by $(i, T, T_N, T_{N-1}, \ldots, T_{l(u)})$, with $M_i \in S_{l(u)}$, satisfying (18)-(24). Denote by $V(i, T, T_N, \ldots, T_{l(u)}) := J(u)$ the cost of the control. Then*

$$V(i, T, T_N, \ldots, T_{l(u)}) = \sum_{j=1}^{L} k_j, \tag{27}$$

where

$$k_j = \frac{c_j x_j(0)}{d}\left[\sum_{k=j}^{L} x_k(0) - \frac{1}{2}x_j(0)\right] \quad \text{for} \quad 0 \leq j < i \tag{28}$$

$$= \frac{c_i}{2}\left[(\mu_{s_{l(u)}} - d)T^2 - \mu_{s_{l(u)}}\tau_i^2 + \frac{\left(\sum_{k=i}^{L} x_k(0)\right)^2}{d}\right] \quad \text{for} \quad j = i \tag{29}$$

$$= \frac{c_j}{2}\left[\mu_{s_{n-1}}\tau_{s_{n-1}}^2 - \mu_{s_n}\tau_j^2 + (\mu_{s_n} - \mu_{s_{n-1}})T_n^2\right] \textit{ for } i < j \leq L-1, j = s_{n-1}+1 \tag{30}$$

$$= c_j\left[x_j(0)\tau_j + \frac{1}{2}\frac{x_j^2(0)}{\mu_{s_n}}\right] \quad \textit{for} \quad i < j \leq L-1,\, M_j \in S_n,\, j \neq s_{n-1}+1 \tag{31}$$

$$= \frac{c_L^-}{2}\left[T^2\left(\mu_{s_{l(u)}} - d\right) + \sum_{k=l(u)+1}^{N} T_k^2\left(\mu_{s_k} - \mu_{s_{k-1}}\right)\right] \textit{ for } j = L. \tag{32}$$

Proof: This follows from Figures 5-9, showing the trajectories of the various buffer levels, and using identities such as $(T - \tau_{s_{l(u)}})\mu_{s_{l(u)}} + \sum_{j=l(u)+1}^{N} X_j(0) = -x_L(0) + T\,d$ to simplify some expressions. □

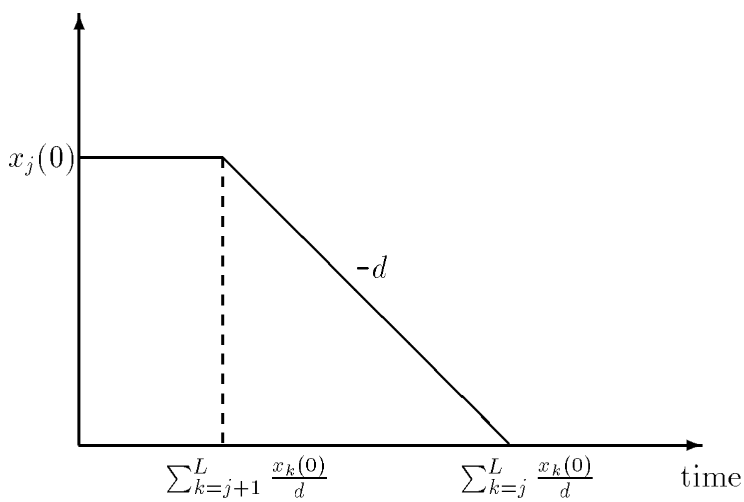

Figure 5: Buffer level trajectory for buffer b_j, $1 \leq j < i$.

The following theorem now shows how the optimal control problem is solved by solving $(i_{max} - i_{min} + 1)$ quadratic programming problems.

Theorem 5 *For each* $i \in \{i_{min}, i_{min}+1, \ldots, i_{max}\}$, *let* $V_{min}(i)$ *be the optimal cost of the quadratic program:*

$$\textit{Minimize } V(i, T, T_N, \ldots, T_{l(u)+1}) \tag{33}$$

subject to (18)-(24), where $l(u)$ *is such that* $M_i \in S_{l(u)}$. *Let* $(T^{(i)}, T_N^{(i)}, \ldots, T_{l(u)+1}^{(i)})$ *be the optimizing solution. Also let*

$$V^* \;=\; \min_{i_{min} \leq i \leq i_{max}} V_{min}(i) \tag{34}$$

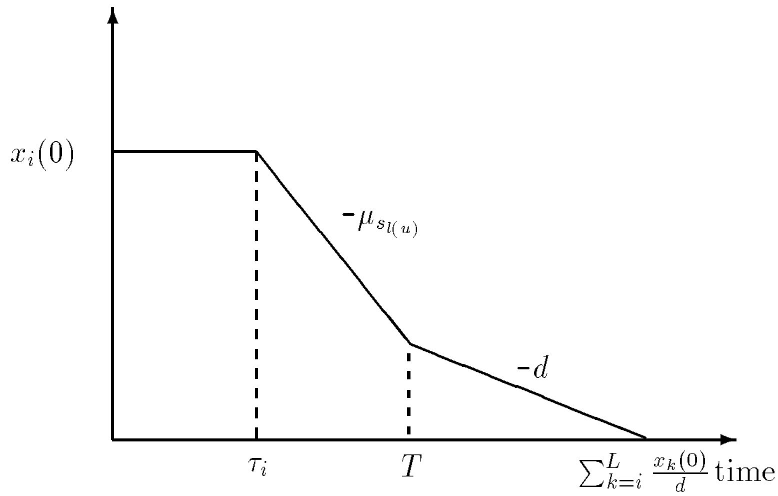

Figure 6: Buffer level trajectory for buffer b_i.

be the minimum of the costs over i, and let i^* be the minimizing choice of i. Then V^* is the optimal cost of the control problem, and the control in $\mathcal{F}_8$ corresponding to $(i^*, T^{(i^*)}, T_N^{(i^*)}, \ldots, T_{l(u)+1}^{(i^*)})$ is an optimal control.

Proof: From Lemmas 1-10, it is clear that if a control $u^*(t)$ attains the minimum cost among the controls $u \in \mathcal{F}_8$, then it is optimal. □

As the following example illustrates, it is not always optimal to attempt erase the shortfall in minimum time.

Example 2 *Filling backlogged orders in minimum time may not be optimal.*

Consider a two-machine system, with $c_2^- = 20, c_1 = 10, c_0 = 0, \mu_1 = 3, \mu_0 = 2, d = 1, x_2(0) = -110$, and $x_1(0) = 55$. Note that $i_{\min} = i_{\max} = 0$. So $i(u) = 0$ for all $u \in \mathcal{F}_8$. The total cost is $J(u) = k_1 + k_2$, where $k_1 = \frac{c_1}{2}\,[\mu_0\,\tau_0^2 + (\mu_1 - \mu_0)\,T_2^2]$ and $k_2 = \frac{c_2^-}{2}\,[(\mu_1 - \mu_0)\,T_2^2 + (\mu_0 - d)\,T^2]$. The optimal solution has deferral time $\tau_0 = 5, T_2 = 45$, the level of the output virtual buffer reaches 0 at time $T = 65$, and the optimal cost is $J(u^*) = 72,875$. However, the minimum time it takes to erase the shortfall in the output virtual buffer b_2 is $T_{\min} = 55$. But, the cost in order to achieve this minimum time is $J(T_{\min}) = 75,625$. Thus, it is not optimal to process material as quickly as possible when a backlog exists. □

The following example illustrates the application of Theorem 5, using Lemmas 12 and 13.

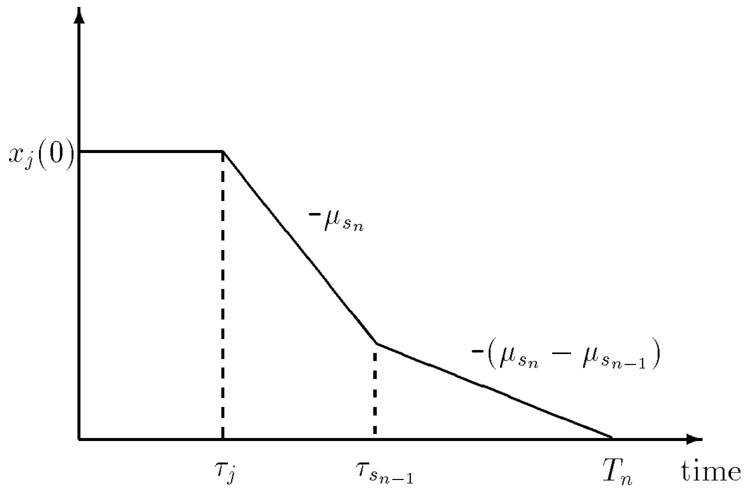

Figure 7: Buffer level trajectory for buffer b_j, $i < j \leq L-1, M_j \in S_n, j = s_{n-1}+1$, and $\tau_j \leq \tau_{s_{n-1}}$.

Example 3 *Determination of an optimal control.*

Consider again the four-machine system of Example 1. Assume that the buffer holding costs are $c_1 = 1, c_2 = 2, c_3 = 3$, and $c_4^- = 4$, and the initial buffer levels are $x_1(0) = 6, x_2(0) = 12, x_3(0) = 24$, and $x_4(0) = -24$. Also assume the demand is $d = 1$, and the processing rates are $\mu_0 = 3/2, \mu_1 = 3, \mu_2 = 2$, and $\mu_3 = 3$. Then, from Lemmas 7 and 11, it follows that $i_{\min} = 1$ and $i_{\max} = 2$. Thus, from Theorem 4, it immediately follows that the optimal deferral time for machine M_0 is $\tau_0 = 18$. Since $\tau_3 = 0$, and machines M_1 and M_2 are both in section S_2, the only deferral time which remains to be determined is τ_2.

By Theorem 5, in order to determine the optimal control, we must consider the two cases $i(u) = 1$ and $i(u) = 2$. Since machine M_1 and machine M_2 both belong to section S_2, the constraints given by Lemma 12 are the same for both cases. Specifically, (18)-(24) become

$$0 \leq \tau_2 \leq T_3 \leq T, \tag{35}$$

$$T - \tau_2 \leq 9, \tag{36}$$

$$T_3 + T = 24, \tag{37}$$

and

$$2\,\tau_2 = 24 - T_3. \tag{38}$$

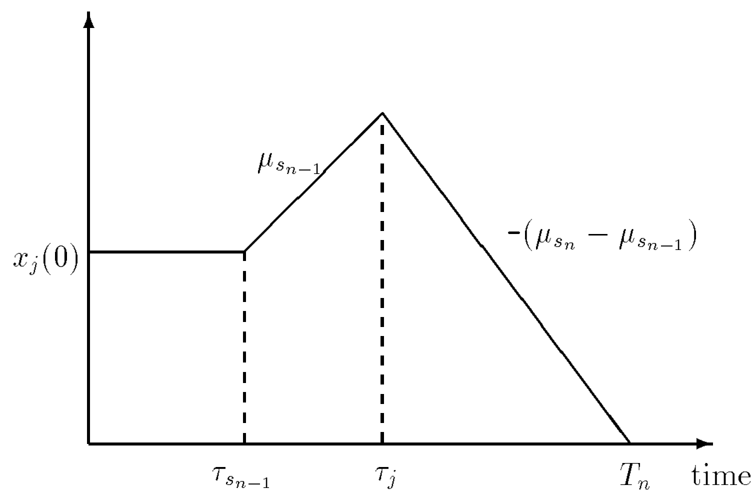

Figure 8: Buffer level trajectory for buffer b_j, $i < j \leq L-1, M_j \in S_n, j = s_{n-1}+1$, and $\tau_j > \tau_{s_{n-1}}$.

From Lemma 13, using (37) and (38), the total cost at each buffer for $i(u) = 1$ is $k_1 = \tau_2^2 - 12\,\tau_2 + 126$, $k_2 = 24\,(\tau_2 + 3)$, $k_3 = 9\,(\tau_2^2 - 16\,\tau_2 + 96)$, and $k_4 = 16\,(\tau_2^2 - 12\,\tau_2 + 72)$. Given the constraints (35) and (36), the solution to the quadratic programming problem (33) is $V_{\min}(1) = \frac{15660}{13} \approx 1204.62$, where $\tau_2 = \frac{81}{13}, T_3 = \frac{150}{13}$, and $T = \frac{162}{13}$.

Similarly, for $i(u) = 2$, the total cost at each buffer is $k_1 = 90$, $k_2 = 2\,(\tau_2^2 + 72)$,, $k_3 = 9\,(\tau_2^2 - 16\,\tau_2 + 96)$, and $k_4 = 16\,(\tau_2^2 - 12\,\tau_2 + 72)$. The solution to (33) is $V_{\min}(2) = 1206$, where $\tau_2 = 6, T_3 = 12$, and $T = 12$.

Thus, the optimal deferral time for section S_2 is $\tau_2 = 81/13$, and the resulting optimal cost is $V^* := \min\{V_{\min}(1), V_{\min}(2)\} = \frac{15660}{13}$. $\square$

9 Re-Entrant Lines

We now exhibit the extension of our results to re–entrant lines. Consider a re-entrant manufacturing system, of the type shown in Figure 2, which produces a single part-type on R machines. Parts follow a fixed route through the system, requiring processing at machines $M_{\sigma_0}, M_{\sigma_1}, \ldots, M_{\sigma_L}$, where $\sigma_i \in \{0, 1, \ldots, R-1\}$. Let buffer b_i, served by machine M_{σ_i}, be the i-th buffer visited. Let μ_j be the number of parts that machine M_j can process in

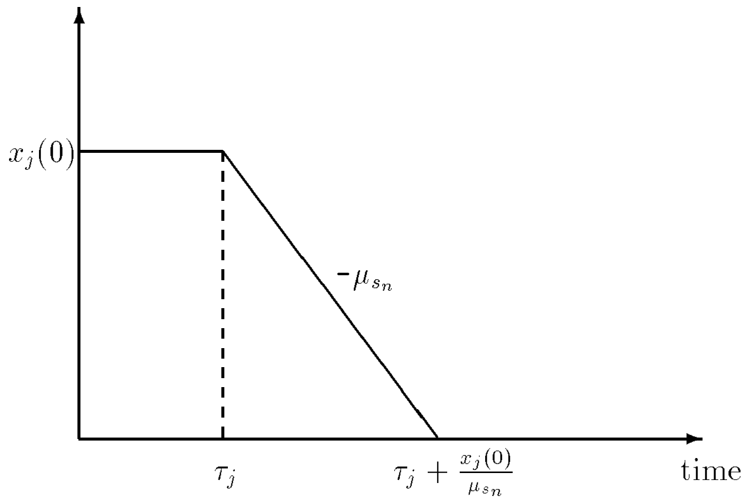

Figure 9: Buffer level trajectory for buffer b_j, $i < j \leq L-1, M_j \in S_n, j \neq s_{n-1}+1$.

unit time. Note that parts in different buffers served by the same machine require the same processing time. Further assume, as in Section 2, that the buffer costs are non-decreasing downstream, i.e., if $i \geq j$, then $c_i \geq c_j$.

As noted in Section 3, if there is a surplus of material in the virtual output buffer b_L, then an optimal control is given by Theorem 1. Thus, we will assume that there is an initial shortfall in b_L, i.e., $x_L(0) < 0$.

Define $n(i,j) := \sum_{k=i}^{L-1} 1(\text{buffer } b_k \text{ is served by machine } M_j)$, to be the the total number of visits which the part-type makes to machine M_j from, and including, buffer b_i to the virtual output buffer b_L. Thus, each unit of material at buffer b_i requires an additional $n(i,j)\mu_j^{-1}$ units of processing time at machine M_j before it leaves the system.

Suppose that machine $M_{\sigma_{L-1}}$ serves N buffers. Then partition the set of buffers $\{b_0, b_1, \ldots, b_{L-1}\}$ into N sections as follows:

$$S_1 = \{b_0, \ldots, b_{s_1}\}, S_2 = \{b_{s_1+1}, \ldots, b_{s_2}\}, S_N = \{b_{s_{N-1}+1}, \ldots, b_{s_N}\},$$

where $s_N := L-1$, and s_{i-1} is recursively defined by

$$s_{i-1} = \max\{0 \leq j < s_i : \sigma_j = \sigma_{L-1}\}.$$

Thus, buffer b_{s_i} is served by machine $M_{\sigma_{L-1}}$, and it is the only buffer in section S_i which is so served.

We will now determine an optimal control for the case where the "final" machine, i.e, the machine which serves buffer b_{L-1}, is the bottleneck. It is hoped that the derivation will provide insight into the general problem, which remains an open problem.

Theorem 6 *If, for $i = 1, \ldots, N$ and $j = s_{i-1} + 1, \ldots, s_i$,*

$$\frac{\mu_{\sigma_{L-1}}}{n(s_i, \sigma_{L-1})} \leq \frac{\mu_{\sigma_j}}{n(j, \sigma_j)},$$

then an optimal control at buffer b_{s_i}, $l \leq i \leq N$ is

$$u^*_{s_i}(t) = \begin{cases} 0 & 0 \leq t < t(i) \\ \frac{\mu_{\sigma_{L-1}}}{n(s_i, \sigma_{L-1})} & t(i) \leq t \leq t(i-1) \\ \cdot & \\ \cdot & \\ \frac{\mu_{\sigma_{L-1}}}{n(s_{l+1}, \sigma_{L-1})} & t(l+1) \leq t \leq t(l) \\ \frac{\mu_{\sigma_{L-1}}}{n(s_l, \sigma_{L-1})} & t(l) \leq t \leq T \\ d & T \leq t, \end{cases} \tag{39}$$

Above, $t(N) = 0$, and $t(i-1)$ is recursively defined by

$$t(i-1) = t(i) + \frac{(N-i+1)X_i(0)}{\mu_{\sigma_{L-1}}}.$$

where

$$X_i(0) := \sum_{s_{i-1}+1}^{j=s_i} x_j(0).$$

Also

$$T := \frac{-x_L 0 - \sum_{i=l}^{N-1} \frac{\mu_{\sigma_{L-1}}}{(N-i)(N-i+1)} t(i)}{\frac{\mu_{\sigma_{L-1}}}{N-l+1} - d}$$

and

$$l := \max\{0 \leq i \leq N : \sum_{k=1}^{N} X_k(0) \geq -x_L(0) + d \sum_{k=i}^{N} \frac{(N-k+1)X_k(0)}{\mu_{\sigma_{L-1}}}\}.$$

In addition, the optimal control at buffer b_j, for $j = s_{i-1} + 1, \ldots, s_i$, is

$$u^*_j(t) = \begin{cases} 0 & 0 \leq t < \tau(j) \\ u_{s_i}(t) & \tau(j) \leq t, \end{cases} \tag{40}$$

where

$$\tau(j) := t(i) + \frac{\sum_{k=j+1}^{s_i} x_j(0)}{\frac{\mu_{\sigma_{L-1}}}{n(s_i, \sigma_{L-1})}}.$$

Proof: The proof centers around the fact that machine $M_{\sigma_{L-1}}$ will work on buffer b_{L-1} at rate $\mu_{\sigma_{L-1}}$ until time T. To see this, suppose machine $M_{\sigma_{L-1}}$ were to work on any of its other buffers. Then, the material processed would not reach the output virtual buffer b_L any faster than it will under $u^*(t)$. This is because machine $M_{\sigma_{L-1}}$ is a bottleneck for every section, and it works at its maximum rate under $u^*(t)$.

Since all the machines can match the rate of machine $M_{\sigma_{L-1}}$, they will not work until all of the buffers downstream from them have emptied. Thus, an optimal control is given by (39) and (40). The other equations follow after some computation, using the fact that $n(s_i, \sigma_{L-1}) = N - i + 1$. □

Example 4 *The last machine is the system bottleneck.*

Consider the re-entrant line shown in Figure 2. Note that $N = 2, L = 9, S_1 = \{b_0, b_1, b_2\}$, and $S_2 = \{b_3, b_4, \ldots, b_8\}$. If $\mu_2 \leq \frac{\mu_0}{2}, \mu_2 \leq \frac{\mu_1}{2}$, and $\mu_2 \leq \mu_3$, then the conditions of Theorem 6 are met, i.e., machine M_2 is the bottleneck machine. An optimal control is therefore given by Theorem 6. □

10 Concluding Remarks

We have studied a pull model of a manufacturing system. For a flow shop, we have shown that when the holding costs are non-decreasing, the problem of determining the optimal control to minimize the total buffer costs plus the system shortfall/inventory cost, can be reduced to a set of quadratic programming problems. We have also determined the optimal control for certain re-entrant lines, when the machine serving the final buffer is the bottleneck.

It would be useful to obtain extensions of these results to cases where the buffer costs are unrestricted, where there are many part–types, and to re–entrant lines.

We have not incorporated any uncertainties in our model, either in processing times or in machine availabilities. The results presented here could be used to derive bounds on manufacturing systems that are nearly reliable, i.e., systems in which the machines rarely fail or fail for only a short time compared to the system dynamics. We refer the reader to Sethi, Taksar and Zhang [23] for other such asymptotic results. Recently, Cornelius [24] has considered systems of unreliable re–entrant lines operating under policies which "aim" each buffer at a desired level, using a priority ordering on buffers to resolve contentions. Optimization is performed through an iteration of two simulations to estimate gradients, and coordinate descent. He has noted that in the examples considered, the cost function is unimodal in the levels. It would useful to analytically investigate such results, and to obtain bounds on performance, perhaps by extending the ideas in Glasserman [21].

References

[1] S. B. Gershwin, "A hierarchical framework for discrete-event scheduling," in *Manufacturing Systems* (P. Varaiya and A. B. Kurzhanski, eds.), no. 103 in Lecture Notes in Control and Information Sciences, Springer-Verlag, 1987. Presented at the IIASA Workshop on Discrete Event Systems: Models and Applications, Sopron, Hungary.

[2] R. W. Conway, W. L. Maxwell, and L. W. Miller, *Theory of Scheduling.* Reading, MA: Addison-Wesley, 1967.

[3] K. Baker, *Introduction to Sequencing and Scheduling.* New York, NY: Wiley, 1974.

[4] S. French, *Sequencing and Scheduling.* New York, NY: Wiley, 1982.

[5] S. C. Graves, "A review of production scheduling," *Mathematics of Operations Research*, vol. 29, pp. 646–675, July-August 1981.

[6] M. Dempster, J. K. Lenstra, and A. Rinooy Kan, eds., *Deterministic and Stochastic Scheduling.* Dordrecht: D. Reidel, 1982.

[7] I. Khosla, "Scheduling in manufacturing." Graduate School of Business, Boston University, November, 1987.

[8] E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, and D. B. Shmoys, "Sequencing and scheduling: Algorithms and complexity," Technical Report BS-R8909, Centre for Mathematics and Computer Science, Amsterdam, The Netherlands, November 1989.

[9] J. F. Muth and G. L. Thompson, *Industrial Scheduling.* Englewood Cliffs, NJ: Prentice-Hall, 1963.

[10] J. R. Perkins and P. R. Kumar, "Stable distributed real-time scheduling of flexible manufacturing/assembly/disassembly systems," *IEEE Trans. Automat. Control*, vol. AC-34, pp. 139–148, February 1989.

[11] P. R. Kumar and T. I. Seidman, "Dynamic instabilities and stabilization methods in distributed real-time scheduling of manufacturing systems," *IEEE Trans. Automat. Control*, vol. AC-35, pp. 289–298, March 1990.

[12] C. J. Chase and P. J. Ramadge, "On real-time scheduling policies for flexible manufacturing systems," *IEEE Trans. Automat. Control*, vol. AC-37, pp. 491–496, April 1992.

[13] S. Lou, S. Sethi, and G. Sorger, "Analysis of a class of real-time multiproduct lot scheduling policies," *IEEE Trans. Automat. Control*, vol. AC-36, pp. 243–248, February 1991.

[14] J. R. Perkins, C. Humes, Jr., and P. R. Kumar, "Distributed control of flexible manufacturing systems: Stability and performance," tech. rep., University of Illinois, Urbana, IL, 1993. To appear in *IEEE Transactions on Robotics and Automation*, 1994.

[15] K. L. Burgess and K. M. Passino, "Flexible manufacturing system analysis and design," tech. rep., The Ohio State University, Columbus, Ohio, 1994.

[16] B. Hajek and R. G. Ogier, "Optimal dynamic routing in communication networks with continuous traffic," *Networks*, vol. 14, pp. 457–487, 1984.

[17] H. Chen and D. D. Yao, "Dynamic scheduling of a multiclass fluid network," *Operations Research*, vol. 41, pp. 1104–1115, November–December 1993.

[18] R. Akella and P. R. Kumar, "Optimal control of production rate in a failure prone manufacturing system," *IEEE Trans. Automat. Control*, vol. AC-31, pp. 116–126, February 1986.

[19] T. Bielecki and P. R. Kumar, "Optimality of zero-inventory policies for unreliable manufacturing systems," *Operations Research*, vol. 36, pp. 532–541, July-August 1988.

[20] A. Sharifnia, "Production control of a manufacturing system with multiple machine states," *IEEE Trans. Automat. Control*, vol. AC-33, pp. 600–626, July 1988.

[21] P. Glasserman, "Hedging–point production control with multiple failure modes," tech. rep., Columbia University, New York, N.Y., December 1993.

[22] R. Bellman, *Dynamic Programming.* Princeton, NJ: Princeton University Press, 1957.

[23] S. Sethi, M. Taksar, and Q. Zhang, "Capacity and production decisions in stochastic manufacturing systems; an asymptotic optimal hierarchical approach," *Production and Operations Management*, vol. 1, pp. 367–392, Fall 1992.

[24] A. Cornelius, "*An Iterative Simulation Method for Determining the Optimal Buffer Level Regulation Policy for Unreliable Manufacturing Systems*," Master's thesis, University of Illinois, Urbana, IL, 1992.