WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

A Dynamic Stochastic Stock Cutting Problem

E. Krichagina, R. Rubio, M. Taksar
and
Lawrence M. Wein

WP# 3512-93-MSA January, 1993

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

A Dynamic Stochastic Stock Cutting Problem

E. Krichagina, R. Rubio, M. Taksar
and
Lawrence M. Wein

WP# 3512-93-MSA January, 1993

A DYNAMIC STOCHASTIC STOCK CUTTING PROBLEM

Elena V. Krichagina, *Institute of Control Sciences, Moscow*

Rodrigo Rubio, *Operations Research Center, M.I.T.*

Michael I. Taksar *Department of Applied Mathematics, SUNY Stony Brook*

and

Lawrence M. Wein, *Sloan School of Management, M.I.T.*

Abstract

We consider a stock cutting problem for a paper plant that produces sheets of various sizes for a finished goods inventory that services customer demand. The controller decides when to shut down and restart the paper machine and how to cut completed paper rolls into sheets of paper. The objective is to minimize long run expected average costs related to paper waste (from inefficient cutting), shutdowns, and backordering and holding finished goods inventory. A two-step procedure (linear programming in the first step and Brownian control in the second step) is developed that leads to an effective, but suboptimal, solution. The linear program greatly restricts the number of cutting configurations that can be employed in the Brownian analysis, and hence the proposed policy is easy to implement and the resulting production process is considerably simplified. In an illustrative numerical example using representative data from an industrial facility, the proposed policy outperforms several policies that use a larger number of cutting configurations. Finally, we discuss some alternative production settings where this two-step procedure may be applicable.

December 1992

A DYNAMIC STOCHASTIC STOCK CUTTING PROBLEM

Elena V. Krichagina, *Institute of Control Sciences, Moscow*

Rodrigo Rubio, *Operations Research Center, M.I.T.*

Michael I. Taksar *Department of Applied Mathematics, SUNY Stony Brook*

and

Lawrence M. Wein, *Sloan School of Management, M.I.T.*

We consider a problem that is commonly faced in a paper factory: how to cut completed rolls of paper into individual sheets for customers. This *stock cutting* problem is traditionally modeled as an integer program: given a set of orders of various sizes. such as $8.5'' \times 11''$, and an unlimited amount of paper on a roll of specified width. cut the roll into the individual orders so as to minimize the amount of paper waste. A huge literature has emerged on various generalizations of this problem; readers are referred to the recent special issue of *European Journal of Operations Research* (1990) and references therein.

Our formulation differs greatly from the existing literature and is driven by the simple fact that many paper factories cut rolls *in anticipation* of customer demand. The particular paper plant that motivates this study guarantees same day delivery of orders. and because sheet cutting is a time-consuming operation, completed paper rolls are immediately cut into sheets of various sizes and placed in a finished goods inventory that services the actual customer demand. This *make-to-stock* view of the problem leads to three complexities. First, since demand is uncertain, the problem is more realistically formulated as a *dynamic and stochastic* one, rather than as a static and deterministic mathematical program. Second, the stock cutting decisions are intimately related to the *amount* of paper produced. and consequently these two decisions need to be considered jointly. Finally, the problem faced by this factory is *multi-criteria* in nature. In addition to wasted paper that cannot be cut into sheets, there are several other costs, which are described below, that are of significant concern.

Since paper machines are extremely expensive, plant managers ideally like to keep them running continually. However, at this particular factory, plant capacity is slightly higher than demand, and hence inventory would build up indefinitely if the machines were never shut down. Since several hours are required to turn an activated machine off and at least one working shift is needed to get an idle machine into working order, many pounds of paper are wasted (although the paper is partially recycled) and significant labor costs are incurred during a shutdown.

Furthermore, as in many industries, prompt and reliable customer delivery is of utmost concern. Managers at this facility believe that customers who do not receive same day delivery often pursue future orders elsewhere. These large shutdown and backorder costs have led this facility to carry millions of pounds of paper sheets in finished goods inventory. Our goal in this study is to find a dynamic *scheduling policy* to minimize shutdown, waste and inventory backorder and holding costs, where a scheduling policy simultaneously decides when to turn the paper machine on and off and how to cut the completed paper rolls.

The problem considered here, which is formulated in Section 1, is idealized in several ways. Whereas most plants have many paper machines, we will consider only a single paper machine. Also, a finished sheet of paper is characterized by its grade, color and size, and changing grades on a paper machine can take about an hour and changing colors can take several minutes. We will ignore these two set-ups and assume that the paper machine under consideration produces only a single color of a single grade. However, in most paper plants, a popular grade, or even color-grade combination, is often processed on a single dedicated paper machine. Finally, shutdown times will not be explicitly modeled.

Rather than attempting to find an optimal solution to this very difficult problem, we look instead for an effective solution that is relatively easy to implement. A *two-step procedure* is taken: in the first step, which is carried out in Section 2, we find the average *activity rates* to minimize the average paper waste subject to meeting average demand. In this linear program, a different activity is defined for each possible *cutting configuration* (that is, combination of sheets that simultaneously fit on the width of the paper roll), and an activity rate is simply the fraction of time that the paper machine employs this activity. In the second step, which is described in Section 3, these activity rates are taken as given, and we consider the problem of finding the dynamic scheduling policy that minimizes the long run expected average shutdown and inventory backorder and holding costs.

Under the heavy traffic conditions that the paper machine must be busy the great majority of the time to satisfy customer demand *and* the shutdown cost is sufficiently large relative to the inventory costs, the dynamic scheduling problem is approximated by a dynamic control problem involving Brownian motion. The Brownian control problem is equivalently reformulated in terms of workloads, and in the course of finding an optimal solution to the *workload formulation*, we provide the first complete solution to an impulse control problem addressed in Bather's (1966) classic paper. The workload formulation solution is easily interpreted in terms of the original problem to propose an (s, S) shutdown policy, where the paper machine is shutdown when the *weighted inventory process*, which dynamically measures the amount of machine time invested in the current inventory, reaches the value S, and the machine is restarted when the weighted inventory process drops to the level s. The determination of the values of s and S is reduced to the solution of two equations that are expressed solely in terms of the original problem parameters.

Since a one-to-one correspondence does not exist between cutting configurations and sheet sizes, interpreting the solution of the workload formulation to obtain a cutting policy is not straightforward. Consequently, we propose a heuristic cutting policy in Section 4 that incorporates two main concepts. First, we adapt Zipkin's (1990) myopic look ahead policy from the setting of a multiclass make-to-stock queue to our more complicated setting. The second idea is to find *target inventory levels* for each sheet size at the moment of shutdown to minimize the inventory costs incurred during the shutdown period. The proposed policy attempts to attain the target inventory levels when the machine is close to shutting down (that is, when the weighted inventory process starts to approach S), and uses the myopic policy otherwise.

The proposed policy is tested on a simulation model that uses representative data from the facility that motivated this study. The machine makes three different sheet sizes, and our proposed policy uses only three of the many possible cutting configurations. Our simulation results show that the average cost under the derived values of s and S are very close to the corresponding cost under the best values of s and S, which were found via a search procedure. Surprisingly, our proposed policy outperforms more complicated heuristics that employ more than three cutting configurations in a seemingly intelligent manner; details are given in Section 5.

The two-step procedure proposed here can be applied to a broad range of multi-criteria scheduling problems in a make-to-stock setting. The approach is useful when a large set of possible activities, or *production processes*, can be employed to produce a set of products, where each activity generates a certain number of completed units of each product, in either a deterministic (as in the stock-cutting case) or stochastic (for example, when random yield is present) manner. Each different production process has an associated variable cost; although in our setting this cost is associated with paper waste, in many other settings the cost reflects the variable cost of employing the process. Average production costs are minimized in the first step of the procedure and inventory backorder and holding costs (and shutdown costs, if appropriate) are minimized in the second step. In Section 6, we briefly describe some alternative settings (from the steel, semiconductor, and blood separation industries) where our procedure may be applicable.

Although the number of potential activities, say J, will typically be much greater than the number of different products, say K, no more than K production processes have positive activity rates in the optimal LP solution in the first step of the procedure, and hence no more than K production processes are considered in the second step of the procedure. In fact, only the processes selected in the first step are required for system stability: for rigorous stability proofs along this line for somewhat related problems, see Courcoubetis et al. (1989) and Courcoubetis and Rothblum (1991). Therefore, the proposed solution is very easy to implement and will lead to a production system that is relatively easy to manage. Moreover, Johnson and Kaplan (1987) and others have observed that significant set-up costs are often incurred when an additional production process is employed. These costs, which until recently have been largely ignored by the cost accounting community, include administrative, labor, purchasing and engineering costs. Although little or no cost is incurred when using an additional activity in our stock cutting example, these set-up costs will be incurred in all the examples sited in Section 6. Hence, the top level of our hierarchical approach, by greatly restricting the number of activities employed, aids in reducing these often considerable costs and leads to a *simplification of the production process.* Of course, this restriction on the number of activities, or processes, that can be employed may lead to a suboptimal solution to the original multi-criteria problem. However, in our computational study in Section 5, we were unable to improve upon the proposed policy by employing more than K of the J activities.

1. Problem Formulation

Consider a paper machine that produces K different types of sheets. Each type of sheet will be referred to as a *product* and is characterized by the sheet dimensions, d_{k1} and d_{k2}, where $d_{k1} \leq d_{k2}$. We will call d_{k1} the product *width* and d_{k2} the *length.* These dimensions are measured in inches so that for standard writing paper, $d_{k1} = 8.5$ and $d_{k2} = 11$. In keeping with industry conventions, we will measure the quantity of paper in terms of pounds, not sheets. Let $D_k(t)$ denote the cumulative number of pounds of product k demanded up to time t. No specific distributional assumption is required; we merely assume that D_k satisfies a functional central limit theorem, where λ_k and v_k^2 denote the asymptotic demand rate and squared coefficient of variation (variance divided by the square of the mean) of the demand process. For simplicity, we assume that the demand for individual products is independent, although this assumption can be easily relaxed.

Let $S(t)$ be the cumulative number of paper rolls produced if the paper machine is continuously busy during the time interval $[0, t]$. We assume that $S(t)$ is a renewal process with asymptotic service rate μ and squared coefficient of variation v^2. The cutting policy dictates the cutting configuration for each completed roll, and we define a different *activity* for each possible cutting configuration. Since each product can have its width dimension d_{k1} or its length dimension d_{k2} placed along the roll's width, activity $j = 1, ..., J$ is characterized by $n_{j,k1}$, the number of sheets of product k placed along its width dimension, and $n_{j,k2}$, the number of sheets placed along its length. Let each paper roll weigh p pounds and be w inches wide. Since each activity must fit on the roll. the *waste* w_j associated with activity j must be nonnegative, where

$$w_j = w - \sum_{k=1}^{K}(n_{j,k1}d_{k1} + n_{j,k2}d_{k2}). \tag{1}$$

Define the $K \times J$ matrix $F = (F_{kj})$ by

$$F_{kj} = (\frac{n_{j,k1}d_{k1} + n_{j,k2}d_{k2}}{w})p. \tag{2}$$

so that F_{kj} is the number of pounds of product k on a roll cut according to activity j. A paper roll cut according to activity j will often be referred to as a *type j roll.*

Recall that our scheduling policy dynamically decides whether the paper machine is busy or idle and how to configure each completed roll. In practice. the paper machine is the bottleneck, and the rolls are typically cut directly after they are produced by the paper machine. However, we find it convenient to assume that the production and cutting of a roll are undertaken simultaneously. In particular, the scheduling policy is defined by a $J-$dimensional vector of nondecreasing processes $\{T_j(t), t \geq 0\}$, where $T_j(t)$ is the cumulative amount of time that the paper machine allocates to type j rolls during $[0, t]$.

Let $Z_k(t)$ be the number of pounds of product k in finished goods inventory at time t. where a negative quantity represents backordered demand: the vector $Z = (Z_k)$ will be referred to as the *inventory process*. If we assume that $Z(0) = 0$, then

$$Z_k(t) = \sum_{j=1}^{J} F_{kj} S(T_j(t)) - D_k(t) \quad \text{for} \quad k = 1, \ldots, K \quad \text{and} \quad t \geq 0. \tag{3}$$

Also, let the *cumulative idleness process* $I(t)$ be the cumulative amount of time that the paper machine is idle in $[0, t]$, where

$$I(t) = t - \sum_{k=1}^{K} T_k(t) \quad \text{for } t \geq 0. \tag{4}$$

Let $J(t)$ denote the cumulative number of times that the paper machine is shut down during $[0, t]$; the *cumulative shutdown process* J can be recovered from the cumulative idleness process I. For concreteness, we assume that shutdown costs are incurred when the machine is turned off. As in Harrison (1988), a scheduling policy T must satisfy

$$T \text{ is continuous, nondecreasing and } T(0) = 0. \tag{5}$$

$$T \text{ is nonanticipating with respect to } Z, \tag{6}$$

$$I \text{ and } J \text{ are nondecreasing with } I(0) = J(0) = 0. \tag{7}$$

where constraint (6) implies that the scheduler cannot observe future demands or service times.

Define the cost function c_k for $k = 1, \ldots, K$ by

$$c_k(x) = \begin{cases} -b_k x & \text{if } x \leq 0, \\ h_k x & \text{if } x > 0, \end{cases} \tag{8}$$

where b_k represents the backorder cost per pound per unit time for product k, and h_k is the inventory holding cost per pound per unit time for product k. Let C_s denote the shutdown cost and let C_w denote the cost per pound of wasted paper. Then the scheduling problem is to find a policy T to

$$\min \limsup_{T \to \infty} \frac{1}{T} E\Big[\int_0^T \sum_{k=1}^{K} c_k(Z_k(t))dt + C_s J(T) + \frac{C_w p}{w} \sum_{j=1}^{J} w_j S(T_j(T))\Big] \tag{9}$$

subject to constraints (3)-(7).

2. The First Step: A Linear Program

The number of possible cutting configurations, J, can be huge, and the goal of the first step of our procedure is to select a small subset of these configurations, or activities, that will actually be employed in the Brownian analysis in the second step. These activities will be selected by solving a linear program that has J decision variables and K constraints. Hence, at most K activities will be used in the Brownian analysis, which greatly simplifies both the analysis and the resulting production process.

The decision variables of the linear program are the *activity rates* x_j, where x_j is the fraction of total time (not just busy time) that the paper machine produces type j rolls. Let

$$R_{kj} = \mu F_{kj}, \tag{10}$$

so that R_{kj} is the rate, in pounds per unit time, that product k is generated when type j rolls are produced. The linear program finds the activity rates that minimize the average paper waste subject to meeting average demand:

$$\min_{x_j} \sum_{j=1}^{J} w_j x_j \tag{11}$$

$$\text{subject to} \quad \sum_{j=1}^{J} R_{kj} x_j = \lambda_k \quad \text{for} \quad k = 1, \ldots, K, \tag{12}$$

$$x_j \geq 0 \quad \text{for} \quad j = 1, \ldots, J. \tag{13}$$

Hereafter, we assume that a unique nondegenerate solution $\rho_1, \ldots, \rho_J$ exists, although the subsequent analysis can be employed with any optimal, possibly nondegenerate, solution. Without loss of generality, assume that the original J activities are indexed so that $\rho_j > 0$ for $j = 1, \ldots, K$ and $\rho_j = 0$ for $j = K+1, \ldots, J$. We also assume that $\sum_{j=1}^{K} \rho_j$, which we denote by the *traffic intensity* ρ, is less than one, so that the paper machine is able to meet average customer demand over the long run.

3. The Second Step: A Brownian Analysis

In this section, we use the LP solution $\rho_1, ..., \rho_K$ to simplify the scheduling problem (3)-(9) in two ways: only these K activities can be employed and the waste cost, which was minimized in the long run average sense in (11), will be ignored. Since the scheduling policy proposed here will automatically enforce the LP solution, our procedure minimizes the average waste cost. Hence, the remaining control problem is identical to (3)-(9) with $J = K$ and $C_w = 0$. Following the procedure taken in Sections 3 through 5 of Harrison (1988), we approximate the scheduling problem as a control problem for Brownian motion. Many of the details of the formulation and analysis of the Brownian control problem will be omitted since they are similar to those described in Wein (1992) and Ou and Wein (1992).

3.1. The Limiting Control Problem. To obtain a limiting Brownian control problem, we need to assume that the server is busy the great majority of the time and the shutdown cost is sufficiently large relative to the inventory costs. We suspect that these heavy traffic conditions typically hold in practice, since paper machines are very capital intensive and shutdown costs are substantial. Formally, one defines a sequence of systems indexed by n where the traffic intensity ρ^n and shutdown cost C_s^n of the n^{th} system are such that $\sqrt{n}(1-\rho^n)$ and $C_s^n/n^{3/2}$ converge to positive constants, while the inventory costs remain unscaled. The original system under study is identified as a particular system in this sequence. Since we will not be attempting to prove convergence of our optimal controlled processes to limiting controlled diffusions, the limiting control problem will be informally stated without introducing the substantial amount of additional notation that is required to define a sequence of systems. Readers are referred to Krichagina et al. (1992a,b) for a proof of convergence for Brownian control problems that approximate single product examples.

Define $\alpha_k = \rho_k/\rho$ to be the proportion of the paper machine's busy time that is devoted to producing type k rolls. Although activities and products are now both indexed by $k = 1, ..., K$, no special relationship exists between activity k and product k. Define the scaled centered allocation process Y by

$$Y_k(t) = \frac{\alpha_k nt - T_k(nt)}{\sqrt{n}} \quad \text{for } k = 1, ..., K \text{ and } t \geq 0, \tag{14}$$

and defined the scaled processes (the same symbols are used on both sides of these equations to reduce the amount of notation required)

$$Z_k(t) = \frac{Z_k(nt)}{\sqrt{n}} \quad \text{for } k = 1, ..., K \text{ and } t \geq 0, \tag{15}$$

$$I(t) = \frac{I(nt)}{\sqrt{n}} \quad \text{for} \quad t \geq 0, \text{ and} \tag{16}$$

$$J(t) = \frac{J(nt)}{\sqrt{n}} \quad \text{for} \quad t \geq 0. \tag{17}$$

The Brownian approximation is obtained by letting the parameter $n \to \infty$, and we will refer to the limiting scaled processes Z, I and J as the inventory, idleness, and shutdown process, respectively. Let $X = (X_1, ..., X_k)$ be a K−dimensional Brownian motion process with drift $\sqrt{n}(\sum_{j=1}^{K} R_{kj}\alpha_j - \lambda_k), k = 1, ..., K$ and covariance matrix Σ, where

$$\Sigma_{kj} = \lambda_k v_k^2 I_{kj} + \mu v^2 \sum_{l=1}^{K} \sum_{m=1}^{K} F_{jl} F_{km} \min(\rho_l, \rho_m) \tag{18}$$

and $I = (I_{kj})$ is the $K \times K$ identity matrix.

The *limiting control problem* is to choose a K−dimensional RCLL (right continuous with left limits) process Y to

$$\min \limsup_{T \to \infty} \frac{1}{T} E\Big[\int_0^T \sum_{k=1}^{K} c_k(Z_k(t))dt + C_s J(T)\Big] \tag{19}$$

$$\text{subject to} \quad Z_k(t) = X_k(t) - \sum_{j=1}^{K} R_{kj} Y_j(t) \text{ for } k = 1, ..., K \text{ and } t \geq 0. \tag{20}$$

$$I(t) = \sum_{k=1}^{K} Y_k(t) \text{ for } t \geq 0, \tag{21}$$

$$I \text{ and } J \text{ are nondecreasing with } I(0) = J(0) = 0, \text{ and} \tag{22}$$

$$Y(0) = 0 \text{ and } Y \text{ is nonanticipating with respect to } X. \tag{23}$$

3.2. The Workload Formulation. The next step is to reformulate (19)-(23) in terms of workloads. Let $R = (R_{kj})$ denote the invertible $K \times K$ matrix obtained by restricting attention in definition (10) to activities $j = 1, ..., K$. Let R_{kj}^{-1} denote the elements of the matrix R^{-1}, and define the *expected effective resource consumption* m_k for product k by

$$m_k = \sum_{j=1}^{K} R_{jk}^{-1} \quad \text{for } k = 1, ..., K. \tag{24}$$

Define the one-dimensional Brownian motion B by

$$B(t) = \sum_{k=1}^{K} m_k X_k(t) \quad \text{for} \quad t \geq 0, \tag{25}$$

so that B has drift $\theta = \sqrt{n}(1-\rho) > 0$ and variance $\sigma^2 = m\Sigma m^T$, or

$$\sigma^2 = \sum_{k=1}^{K} \lambda_k m_k^2 v_k^2 + \mu v^2 \sum_{j=1}^{K} m_j \sum_{k=1}^{K} m_k \sum_{l=1}^{K} \sum_{m=1}^{K} F_{jl} F_{km} \min(\rho_l, \rho_m). \tag{26}$$

Since R is invertible, it follows from Proposition 1 of Wein (1992) that the workload formulation of the limiting control problem (19)-(23) is to choose the K-dimensional process Z and the one-dimensional process I to

$$\min \limsup_{T \to \infty} \frac{1}{T} E\Big[\int_0^T \sum_{k=1}^{K} c_k(Z_k(t))dt + C_s J(T)\Big] \tag{27}$$

subject to $$\sum_{k=1}^{K} m_k Z_k(t) = B(t) - I(t) \quad \text{for } t \geq 0, \tag{28}$$

I and J are nondecreasing with $I(0) = J(0) = 0$, and (29)

Z and I are nonanticipating with respect to X. (30)

The two problem formulations (19)-(23) and (27)-(30) are equivalent in that their objective functions are identical and every feasible (that is, all the constraints are satisfied) policy Y for the limiting control problem yields a corresponding feasible policy (Z, I) for the workload formulation, and every feasible policy (Z, I) for the workload formulation yields a corresponding feasible policy Y for the limiting control problem. The cumulative shutdown process J is not an explicit control in (27)-(30), since it is dictated by the cumulative idleness process I.

3.3. The Solution to the Workload Formulation. Not only is the workload formulation easier to solve than the limiting control problem, but the workload formulation solution, which consists of the scaled inventory process Z and cumulative idleness process I, is easier to interpret in terms of the original problem than the scaled centered allocation process Y. The solution to the workload formulation requires two steps. The optimal control process Z is derived in terms of the control process I in the first step, and the optimal control process I is derived in the second step. Since the first step has been carried out in Section 4 of Wein (1992), only the solution Z^* will be displayed. Define the one-dimensional *weighted inventory process* W by

$$W(t) = \sum_{k=1}^{K} m_k Z_k(t) \quad \text{for} \quad t \geq 0. \tag{31}$$

This quantity is the weighted sum of the finished goods inventory for each product. where the weight is the expected effective resource consumption. By (28). we also have

$$W(t) = B(t) - I(t) \quad \text{for} \quad t \geq 0. \tag{32}$$

Without loss of generality, define the indices j and l by

$$\min_{1 \leq k \leq K} \frac{h_k}{m_k} = \frac{h_j}{m_j} \quad \text{and} \tag{33}$$

$$\min_{1 \leq k \leq K} \frac{b_k}{m_k} = \frac{b_l}{m_l}, \tag{34}$$

where it is possible for $j = l$. Then the optimal solution $Z^*(t)$ is

$$Z_k^*(t) = \begin{cases} \frac{W(t)}{m_k} & \text{if } k = j \text{ and } W(t) \geq 0, \\ 0 & \text{if } k \neq j \text{ and } W(t) \geq 0, \end{cases} \tag{35}$$

and

$$Z_k^*(t) = \begin{cases} \frac{W(t)}{m_k} & \text{if } k = l \text{ and } W(t) < 0, \\ 0 & \text{if } k \neq l \text{ and } W(t) < 0. \end{cases} \tag{36}$$

Notice that the optimal control process Z^* is expressed in terms of the control process I via (32).

To describe the remaining control problem for I, let us define $h = h_j/m_j$ and $b = b_l/m_l$, where the indices j and l are defined in (33)-(34), and let

$$f(x) = \begin{cases} hx & \text{if } x \geq 0, \\ -bx & \text{if } x < 0. \end{cases} \tag{37}$$

The second step of the workload formulation solution is to find a nondecreasing, nonanticipating (with respect to X) control process I to

$$\min \limsup_{T \to \infty} \frac{1}{T} E\Big[\int_0^T f(W(t))dt + C_s J(T)\Big] \tag{38}$$

$$\text{subject to} \quad W(t) = B(t) - I(t) \ \text{ for } t \geq 0. \tag{39}$$

Recall that the shutdown process J measures the cumulative number of times that the control process I is exerted. Problem (38)-(39) is a one-dimensional impulse control problem for Brownian motion. The word *impulse* stems from the fact that, due to the fixed cost C_s, the optimal control process I takes the form of instantaneous jumps, or impulses, exerted at certain points in time. A nearly identical problem (essentially a mirror image of our problem), the only differences being that (39) is changed to $W(t) = B(t) + I(t)$ and the drift of the Brownian motion is negative, was solved by Bather (1966). The solution to Bather's problem is an (s, S) policy: the controlled process $W(t)$ is displaced to the level S whenever the level s is reached, where $S > s$; this suggests that an optimal policy to our problem displaces $W(t)$ to the level s whenever the level S is reached, where $S > s$. A closed form solution to s and S does not appear possible for this problem. Bather reduces the determination of s and S (and the optimal long run average cost) to the solution of equations (6.1), (6.2) and (6.4) of his paper, and asserts that $S > 0 > s$. Under this assertion, Puterman (1975) employs a regenerative argument to reduce the determination of s and S to the solution of two equations that are displayed on page 156 of his paper, and notes that equation (6.4) of Bather appears to contain a minor misprint.

The same procedures can be used to derive analogous equations for our problem. Using Puterman's approach and Bather's assertion that $S > 0 > s$, we find that S is the solution to

$$\frac{2\theta\phi^2 C_s}{h+b} = (\frac{h+b}{b})(1-\phi S - e^{-\phi S})^2 - 2(1-\phi S + \frac{\phi^2 S^2}{2} - e^{-\phi S}), \tag{40}$$

and s can be determined from

$$S - s = (\frac{h+b}{b})\Big(S - \phi^{-1}(1-e^{-\phi S})\Big), \tag{41}$$

where $\phi = 2\theta/\sigma^2$. If we let g denote the long run average optimal cost of (38)-(39). then it satisfies

$$g = b(\phi^{-1} - s). \tag{42}$$

However, upon numerically solving various problem instances of (40)-(41). we encountered several cases where $S > s > 0$. Closer examination reveals that Bather's original assertion that $S > 0 > s$ does not always hold, and hence his solution is incomplete: since our problem is the mirror image of his problem, (40)-(41) is only a partial solution to (38)-(39). The reason that s and S can be of the same sign is explained as follows. Although the drift of B is positive in (38)-(39), the controlled process can still attain values below s. Hence, if b is much greater than h in (37), the shutdown cost is not very large. and the drift is not far from zero, it can be optimal for $s > 0$. An analogous argument shows that it is possible for $s < S < 0$ in Bather's problem; for the inventory problem motivating Bather's study, it is highly unlikely for the parameter values to satisfy the conditions that imply $s < S < 0$, and so it is not surprising that he did not consider this case.

To complete the solution to (38)-(39), we simply use Puterman's regenerative approach under the assumption that $S > s > 0$. In this case, the parameter $\Delta = S - s$ can be found by solving

$$\frac{\theta C_s}{h} = \Delta^2\Big(\frac{1}{1-e^{-\phi\Delta}} - \frac{1}{2}\Big) - \frac{\Delta}{\phi}. \tag{43}$$

and then

$$S = -\frac{1}{\phi}\ln\Big[(\frac{h}{h+b})(\frac{\phi\Delta}{e^{\phi\Delta}-1})\Big]. \tag{44}$$

The general theory of variational inequalitiies, which is now well developed, shows that the Hamilton-Jacobi-Bellman equation of the impulse control problem has a unique solution and this solution yields an (s, S) optimal policy. The latter implies the existence of a feasible (i.e., the sign constraint on s and S is satisfied) solution to either (40)-(41) or (43)-(44). This, however, does not exclude the possibility that two local minima, corresponding to feasible solutions to both (40)-(41) and (43)-(44), exist. In this case, the minimal cost solution must be employed. Whether or not exactly one local minimum exists is a rather tangential issue for this problem, and therefore we have made no effort to prove this fact. It should be mentioned, however, that all of our numerical computations with equations (40)-(41) and (43)-(44) have yielded one solution that violates the sign constraint and one solution that satisfies the sign constraint. Moreover, the two solutions coincide when the optimal value of S equals zero.

For completeness, the solution to Bather's problem when $s < S < 0$ is given by (using Bather's notation)

$$\frac{\lambda K}{p} = (S-s)^2\Big(\frac{1}{1-e^{-\lambda(S-s)}} - \frac{1}{2}\Big) - \frac{S-s}{\lambda}. \tag{45}$$

and

$$s = \frac{1}{\lambda}\ln\Big[(\frac{p}{h+p})(\frac{\lambda(S-s)}{e^{\lambda(S-s)}-1})\Big]. \tag{46}$$

Finally, note that if the shutdown cost $C_s = 0$ in problem (38)-(39), then the solution to (43) is $S = s$ (i.e., $\Delta = 0$). If we let $\Delta \to 0$ in (44), then L'Hopital's rule yields the limiting solution

$$S = \frac{1}{\phi}\ln(1+\frac{b}{h}). \tag{47}$$

When $C_s = 0$, problem (38)-(39) becomes a singular control problem, and the solution (see Section 6 in Wein) is

$$I(t) = \sup_{0\le s\le t}\big[B(s) - \frac{1}{\phi}\ln(1+\frac{b}{h})\big]^+, \quad t \ge 0. \tag{48}$$

so that the controlled process is a reflected Brownian motion in the interval $(-\infty, \phi^{-1}\ln(1+(b/h))$. Hence, as expected, the two solutions coincide as the shutdown cost $C_s \to 0$.

4. The Proposed Scheduling Policy

In this section, we propose a scheduling policy that is partially based on the preceding analysis. Recall that a scheduling policy dictates the shutdown/startup times for the machine, and dynamically chooses a cutting configuration for each roll of paper that is produced.

The shutdown policy is easily interpreted in terms of the optimal control process I^*. which represents the scaled cumulative idleness process. By (31) and (39). the $(s. S)$ policy derived in (40)-(41) and (43)-(44) suggests that the machine should be shut down when the scaled weighted inventory process $W(t)$ reaches the level S, and should be reactivated when $W(t)$ drops to s. Moreover, by reversing the heavy traffic scalings. the (s, S) policy for the original system can be expressed solely in terms of the problem parameters. More specifically, if we substitute $s/\sqrt{n}$ for s, $S/\sqrt{n}$ for S, $\Delta/\sqrt{n}$ for Δ. $C_s/n^{3/2}$ for C_s and $\sqrt{n}(1-\rho)$ for θ in (40)-(41) and (43)-(44), then the scaling parameter n vanishes. and these four equations become, respectively,

$$\frac{8(1-\rho)^3 C_s}{\sigma^4(h+b)} = \Big(\frac{h+b}{b}\Big)\Big(1-\frac{2(1-\rho)S}{\sigma^2}-\epsilon^{-2(1-\rho)S/\sigma^2}\Big)^2 - 2\Big(1-\frac{2(1-\rho)S}{\sigma^2}+2\Big(\frac{(1-\rho)S}{\sigma^2}\Big)^2-\epsilon^{-2(1-\rho)S/\sigma^2}\Big). \tag{49}$$

$$S - s = \Big(\frac{h+b}{b}\Big)\Big(S-\frac{\sigma^2}{2(1-\rho)}(1-\epsilon^{-2(1-\rho)S/\sigma^2}\Big). \tag{50}$$

$$\frac{(1-\rho)C_s}{h} = \Delta^2\Big(\frac{1}{1-\epsilon^{-2(1-\rho)\Delta/\sigma^2}}-\frac{1}{2}\Big)-\frac{\sigma^2\Delta}{2(1-\rho)}. \tag{51}$$

and

$$S = -\frac{\sigma^2}{2(1-\rho)}\ln\Big[\Big(\frac{h}{h+b}\Big)\Big(\frac{2(1-\rho)\Delta}{\sigma^2(\epsilon^{2(1-\rho)\Delta/\sigma^2}-1)}\Big)\Big]. \tag{52}$$

The proposed shutdown/startup policy calculates s and S from these equations, and employs the resulting (s, S) policy with respect to the original (that is, unscaled) weighted inventory process $\sum_{k=1}^{K} m_k Z_k(t)$.

In Brownian approximations to queueing system scheduling problems (see, for example, Harrison 1988 and Wein 1992), the policy for "who to serve next," which corresponds to the cutting policy here, is typically interpreted in terms of the pathwise solution to the LP that is embedded in the workload formulation. In our case, the pathwise solution is given by (35)-(36). However, unlike the previous problems that have been analyzed, there is no one-to-one correspondence between the K cutting configurations and the K products. Consequently, we have been unable to interpret (35)-(36) to obtain a reliable cutting policy.

Instead, we propose a heuristic policy that incorporates two essential ideas. The first idea, which is due to Zipkin, is to use a myopic look ahead policy. He proposes a service time look ahead policy for a single server multiclass make-to-stock queue, which is essentially a special case (where $K = J, C_s = 0$ and the matrix R is diagonal) of the problem considered here. This policy dynamically serves the class that minimizes the expected reduction in cost rate per unit of time after one service completion. The policy has its roots in Miller's (1974) transportation look ahead policy for the decision of which base to send an item repaired at a central depot, where transportation time plays the role of service time. Numerical results in Veatch and Wein (1992) show that Zipkin's policy performs very well for multiclass make-to-stock queues.

We propose a simple extension to Zipkin's policy that makes it adaptable to the cutting problem considered here. Since a paper roll's service time is independent of how it is eventually cut, the policy chooses the cutting configuration $j = 1, \ldots, K$ (only the configurations identified by the LP in Section 2 will be considered) that minimizes the expected inventory cost rate incurred after one service time; that is, at time t, we cut a type j^* roll, where

$$j^* = \arg\min_j E\left[\sum_{k=1}^{K} c_k\Big(Z_k(t) + F_{kj} - D_k(S)\Big)\right], \tag{53}$$

and the expectation is taken over both the service time S and the demand processes $D_1, \ldots D_K$.

Although this policy works well when no shutdown costs are incurred, an effective cutting policy for this problem needs to look beyond one service time and, in particular, must prevent the vector inventory process from entering a vulnerable position (for example, incurring many costly backorders) during the potentially long shutdown period. Hence, we derive a *target inventory level* ϵ_k for each product k at the moment of shutdown (i.e., when $W(t) = S$) to minimize inventory costs incurred during the shutdown period. In order to obtain a closed form expression for ϵ_k, we assume that demand is deterministic during the shutdown period. Under this assumption, the length of the shutdown period is

$$\tau = \frac{S - s}{\sum_{k=1}^{K} \lambda_k m_k}, \tag{54}$$

and the target inventory levels are found by solving

$$\min_{\epsilon_k} \sum_{k=1}^{K} \int_0^{\tau} c_k(\epsilon_k - \lambda_k t)dt \tag{55}$$

$$\text{such that } \sum_{k=1}^{K} m_k \epsilon_k = S, \tag{56}$$

where S is the unscaled maximum weighted inventory level derived in (49)-(52). The solution to (55)-(56) is

$$\epsilon_k = \frac{b_k \lambda_k \tau}{h_k + b_k} + \frac{m_k \lambda_k}{h_k + b_k}\left(\frac{S - \tau \sum_{k=1}^{K} \frac{b_k \lambda_k m_k}{h_k + b_k}}{\sum_{k=1}^{K} \frac{\lambda_k m_k^2}{h_k + b_k}}\right). \tag{57}$$

When the weighted inventory process gets sufficiently close to S, we use the cutting configuration that minimizes the Euclidean distance between the expected resulting inventory levels and the target levels; that is, cut a type j^* roll, where

$$j^* = \arg\min_j \sqrt{\sum_{k=1}^{K} (Z_k(t) + F_{kj} - \frac{\lambda_k}{\mu} - \epsilon_k)^2}. \tag{58}$$

Our proposed cutting policy uses the myopic policy (53) when

$$\sum_{k=1}^{K} m_k Z_k(t) < s + \gamma(S - s), \tag{59}$$

and uses the target inventory policy (58) when

$$\sum_{k=1}^{K} m_k Z_k(t) \geq s + \gamma(S - s), \tag{60}$$

where the parameter $\gamma \in [0, 1]$ quantifies the proximity to machine shutdown. In the simulation study in the the next section, we set the parameter $\gamma = 0.9$, and found that the policy's performance was quite robust with respect to the specification of γ.

5. An Example

In this section, we compare our proposed scheduling policy against several other heuristics on a simulation model of a hypothetical system.

The Problem Data. The data for the simulation model are based on conversations with a manager of the facility under study. We consider a single paper machine that produces three different products. Table I provides the size, demand rate and inventory cost rates for each product. The demand rate for each product is assumed to be a compound Poisson process; the average demand rate of orders is given in Table I, and all orders for all products are independent Poisson random variables with mean 200 pounds. Hence, if we denote the average demand rate of product k orders by $\bar{\lambda}_k$, then the term $\lambda_k v_k^2$ in (18) equals $40,200\bar{\lambda}_k$.

PRODUCT	DIMENSIONS (inches)	DEMAND RATE (orders/hr)	HOLDING COST ($/lb/yr)	BACKORDER COST ($/lb/yr)
1	8.5×11	1.509	0.20	1.00
2	17×22	2.465	0.30	1.50
3	28×35	1.056	0.40	2.00

TABLE I. Product Data.

Each paper roll is $w = 60$ inches wide and weighs $p = 1250$ pounds. When the paper machine is turned on, the production rate is 0.875 rolls per hour. Hence, if the paper machine experienced no paper waste, then it would need to be busy 92.0% of the time over the long run to meet average demand. The service times are Erlang with squared coefficient of variation $v^2 = 0.05$. Notice that nearly all of the statistical variability in this system is due to uncertain demand. Finally, the shutdown and waste costs are estimated to be $C_s = \$10,000$ and $C_w = \$0.70$ per pound, respectively.

The Proposed Policy. The first step in our procedure is to solve the LP (11)-(13). We first solved this problem using an explicit enumeration of the 31 activities, or cutting configurations, with waste less than 8.5″; all other activities waste paper that could be made into final product. Using a column generation approach, this solution was verified as optimal over all activities. The LP solution (ρ_1, ρ_2, ρ_3) is displayed in Table II. Summing up the last column of this table yields $\rho = 0.935$; hence, under the proposed activity mix, the paper machine needs to be busy 93.5% of the time to meet average demand. The integers in Table II refer to the variables $n_{j,k1}$ and $n_{j,k2}$; for example, activity 1 has three 8.5″ × 11″ sheets, one of which is placed along its width and two along its length, and one 28″ × 35″ sheet placed along its width.

ACTIVITY	8.5″	11″	17″	22″	28″	35″	WASTE	ρ_j
1	1	2	0	0	1	0	1.5″	0.392
2	1	0	3	0	0	0	0.5″	0.521
3	1	0	0	1	1	0	1.5″	0.022

TABLE II. The Solution to LP (11)-(13).

Further calculations show that the expected effective resource consumption values defined in (24) are $m_1 = m_2 = 0.00092$ and $m_3 = 0.00095$, which are in units of hours per pound of paper. Hence, the indices defined in (33)-(34) are $j = l = 1$, and the cost parameters appearing in (37) are $h = 216.92$ and $b = 1084.60$, both in dollars per pound per year. Also, the variance in (26) is $\sigma^2 = 0.2778$. Finally, the (s, S) policy is derived from (49)-(50) and is $s = -39.74$ and $S = 211.17$, which are both in terms of hours. This policy roughly corresponds to shutting down the machine every five months for about ten days.

Under our distributional assumptions (Erlang service times and compound Poisson demands), the expected cost rate in (53) does not have a closed form expression. and the numerical computation required to compute this quantity would be too cumbersome to embed in our simulation model. Instead, we computed the quantity $E\Big[c_k\Big(Z_k(t) + F_{kj} - D_k(S)\Big)\Big]$ in (53) under the assumptions of exponential service times and Poisson demands, which yields

$$h_k(Z_k(t) + F_{kj} - \frac{\lambda_k}{\mu}) + \frac{\lambda_k}{\mu}(h_k + b_k)\left(\frac{\lambda_k}{\lambda_k + \mu}\right)^{Z_k(t)+F_{kj}} \quad \text{for} \quad Z_k(t) + F_{kj} \geq 0. \tag{63}$$

and

$$-b_k(Z_k(t) + F_{kj} - \frac{\lambda_k}{\mu}) \quad \text{for} \quad Z_k(t) + F_{kj} < 0. \tag{64}$$

The optimal target inventory levels are given by $\epsilon_1 = 68,441, \epsilon_2 = 111,280$ and $\epsilon_3 = 47,609$ pounds.

Straw Policies. Since our problem formulation is new, there are no obvious straw policies to test against the proposed policy, which is denoted by PROPOSED in Table III. We consider three other policies; all three policies use a shutdown/startup policy characterized by an (s, S) policy on the total unweighted inventory process, where the values of s and S are found via a search on the computer simulation model. The first and most simplistic policy is called the LP policy in Table III, and it deterministically enforces the solution (ρ_1, ρ_2, ρ_3) given in Table II. More specifically, the cutting policy continually repeats a cycle of ten rolls, where the first nine rolls are activities (2,1,2,1,2,1,2,1,2), and the tenth roll of each cycle is randomized according to the probabilities (0.193, 0.572, 0.235). The second policy, called MYOPIC, uses the myopic policy defined in (53), but considers not the three cutting configurations specified by the LP, but all configurations j with waste $w_j < \bar{w}$. The cost minimizing value of the parameter $\bar{w}$ is found using a search procedure. The third policy, called MYOPIC+TARGET, uses our proposed cutting policy, but as in the MYOPIC policy, considers all configurations j with waste $w_j < \bar{w}$.

Simulation Results. For three of the four scheduling policies, we ran ten independent runs, each consisting of six years. To obtain sufficiently small confidence intervals, we made 50 independent runs for the LP policy. Although we do not expect the demand facing this facility to remain stable for six years, a sufficiently long period was required to eliminate round off effects due to shutdowns. All runs began with the system empty and the machine activated. The cost minimizing values of the parameters for the scheduling policies (the parameter γ for the PROPOSED policy, the parameters s and S for the other three policies, and the parameter $\bar{w}$ for the two myopic policies) were found by making three runs of six years.

POLICY	WASTE COST	INVENTORY COST	SHUTDOWN COST	TOTAL COST
PROPOSED	99,331(±195)	34,640(±575)	18,967(±361)	152,938(±663)
MYOPIC+TARGET	121,039(±383)	21,180(±380)	24,865(±589)	167,084(±535)
LP	99,267(±80)	67,840(±8431)	20,434(±254)	187,541(±8484)
MYOPIC	130,938(±370)	34,970(±703)	26,618(±551)	192,526(±850)

TABLE III. Simulation Results: Average Annual Costs.

The simulation results are reported in Table III, which contains the average total annual cost, and its breakdown into inventory (holding and backorder), waste and shutdown costs; 95% confidence intervals are reported for all cost quantities. Notice that paper waste accounts for roughly two-thirds of the total cost. For ease of comparison, Table IV reports average costs as a percentage of the costs of the PROPOSED policy.

POLICY	WASTE COST	INVENTORY COST	SHUTDOWN COST	TOTAL COST
PROPOSED	100%	100%	100%	100%
MYOPIC+TARGET	122%	61%	131%	109%
LP	100%	196%	108%	123%
MYOPIC	132%	101%	140%	126%

TABLE IV. Simulation Results: Relative Costs.

The PROPOSED policy outperforms the other three policies, and the relative cost increases experienced by the other three policies range from 9% to 26% in Table IV. To test the accuracy of our derived values for s and S, a two-dimensional search was undertaken for the cost minimizing values. The search yielded an average annual cost of \$149,932, which is less than 2% lower than the cost under our derived values. However, the cost minimizing values, $s = -17.5$ and $S = 175.0$, are significantly different than the derived values, $s = -39.74$ and $S = 211.17$. Consequently, the cost minimizing s and S resulted in lower inventory costs and higher shutdown costs relative to the derived values of s and S; its four normalized costs for Table IV are 100%, 75%, 131% and 98%.

The cost minimizing value of γ in (59)-(60) was 0.9. The PROPOSED policy's performance was remarkably robust with respect to this parameter. The average total cost increased by 0.3% when $\gamma = 0.95$ and increased by less than 0.2% when $\gamma = 0.7$.

We also tested the PROPOSED policy under one slight variant; the (s, S) policy was based on the total unweighted inventory $\sum_{k=1}^{K} Z_k(t)$ rather than the weighted inventory $\sum_{k=1}^{K} m_k Z_k(t)$. A search for the best (s, S) policy gave essentially the same cost (\$150.037) as the search under the weighted inventory policy (\$149,932). This close agreement is hardly surprising since the three m_k values are very similar. Hence, any discrepancy among the various policies cannot be attributed to the fact that the PROPOSED policy bases its (s, S) policy on the weighted inventory, and the other policies base their (s, S) policy on the unweighted inventory.

As expected, the LP policy achieves the same waste cost as the PROPOSED policy. However, since the LP policy is an open loop policy (i.e., the cutting policy is independent of the state of the system), very high inventory costs are incurred. The cost minimizing values of s and S are $s = -15,000$ and $S = 235,000$. Recall that s and S are expressed in units of hours for the PROPOSED policy, and in units of pounds of paper for the other three policies.

The MYOPIC policy achieves the worst performance of the four policies. Since the MYOPIC and PROPOSED policies have similar inventory costs, it appears that the inventory cost reductions gained by the MYOPIC policy while the machine was operating were offset during the shutdown periods. Moreover, since the MYOPIC policy employs activities that do not minimize average waste costs, a significant increase in waste cost is incurred. Finally, the high backorder cost incurred during the long shutdown periods forces this policy to employ more frequent shutdowns and hence incur higher shutdown costs. The cost minimizing values of s and S are $s = -50,000$, $S = 100,000$.

For both myopic policies, the cost minimizing value of $\bar{w}$, which is the maximum allowable waste on a paper roll, was found to be 2.75″, which was the lowest value that included any activities producing product 3. This parameter value allowed ten cutting configurations to be employed. When $\bar{w}$ was raised to 3.0″, two more activities were included and the overall cost increased by 6.3% for the MYOPIC policy and 9.4% for the MYOPIC+TARGET policy. When $\bar{w}$ was raised to the next level of 4.75″, eight additional activities were included for a total of 20, and the cost roughly doubled for both policies.

The use of the target inventory levels clearly enhances performance, since the MYOPIC+TARGET policy achieves a much lower cost than the MYOPIC policy. The MYOPIC+TARGET policy achieves significantly lower inventory costs than the other three policies. However, relative to the PROPOSED policy, this cost reduction is more than offset by increased waste and setup costs. We were initially somewhat surprised that the PROPOSED policy, which uses only three activities, outperformed the MYOPIC+TARGET policy, which can employ any number of activities; after all, the primary reason for restricting the number of cutting configurations was to obtain a mathematically tractable problem. The discrepancy in performance between these two policies can best be seen by comparing the MYOPIC+TARGET policy to the PROPOSED policy under the cost minimizing values (as opposed to the derived values) of s and S: the two policies have identical shutdown costs, the PROPOSED policy has 21.8% higher inventory costs than the MYOPIC+TARGET policy, and the MYOPIC+TARGET policy has 21.9% higher waste costs than the PROPOSED policy. Hence, the 11% difference in average total cost between the two policies is due to the fact that waste costs comprise roughly two-thirds of the total cost. Therefore, since waste costs are minimized in the first step of our procedure, the PROPOSED policy may not work as well when waste costs are small relative to inventory and shutdown costs. However, we reiterate that the problem parameters employed in this simulation study are based upon our best estimates from an actual facility.

It is worth noting that not only does the PROPOSED policy outperform the other three policies in the simulation study, but it is also *simpler to employ* than the other policies. First, the PROPOSED policy is based on derived values of s and S, whereas the other policies require a two-dimensional search using a simulation model. Furthermore, the PROPOSED policy's performance is very robust with respect to the parameter γ, which specifies the proximity to shutdown. In contrast, the performance of the two myopic policies was very sensitive to the parameter $\bar{w}$, which is the maximum allowable waste per roll. Hence, the parameter γ is much easier to specify than the parameter $\bar{w}$.

6. Other Applications

In this section, we briefly describe several other settings where the two-step procedure may be applicable.

Semiconductor Manufacturing. In semiconductor wafer fabrication, a batch of wafers produced according to a particular process randomly yields chips of many different product types, which are partially ordered with respect to quality. Here, F_{kj} represents the expected number of type k chips in a batch of wafers produced according to process j. These facilities usually have many more possible production processes than product types. Each production process has its own variable cost, and costlier processes will typically yield higher quality chips. The LP in the first step of the procedure minimizes average variable processing cost subject to meeting average demand. Wafer fabrication facilities often operate in a make-to-stock mode and often have only one bottleneck station, the photolithography workstation, and wafers visit this station ten to twenty times during processing. Hence, by focusing on the bottleneck workstation, we obtain a scheduling problem for a production/inventory system with customer feedback and random yield. The Brownian control problem in the second step of the procedure was explicitly solved in Ou and Wein (1992) and readers are referred there for details. Similar problems also occur in fiber optics, ingot and crystal cutting, and blending operations in the petroleum industry.

Blood Separation. A variety of separation processes can be used to separate blood into its various components (plasma, red blood cells, etc.), which are maintained in a finished goods inventory servicing hospitals and other health care facilities. Here, each separation process has its own variable cost, which are minimized in the first step LP, and produces specific amounts of each component. One key aspect that is not captured in our model is the perishability of the blood; typically, blood components (except for plasma) have to be discarded after sitting in inventory for a certain number of days.

Steel Industry. In a tube mill, each product type can be made from several different blooms, or starting stocks, and a particular starting stock can be used to produce several different products. If we define an activity for each feasible combination of starting stock and product type, then the activity cost includes both raw material costs for the starting stock and processing costs, which differ by activity. The second step analysis develops a dynamic production policy for the bottleneck operation of the tube mill.

Acknowledgments

We are deeply indebted to Tim Loucks, plant manager at Strathmore Paper Company, for generously sharing his time and expertise. We also thank Anant Balakrishnan for engaging in helpful conversations about the steel industry. This research was partially supported by a grant from the Leaders for Manufacturing Program at MIT and by National Science Foundation Grant Award No. DDM-9057297.

REFERENCES

Bather, J. 1966. A Continuous Time Inventory Model. *J. Applied Probability* **3**, 538-549.

Courcoubetis, C., P. Konstantopoulos, J. Walrand, and R. R. Weber. 1989. Stabilizing an Uncertain Production System. *Queueing Systems* **5**, 37-54.

Courcoubetis, C. and U. G. Rothblum. 1991. On Optimal Packing of Randomly Arriving Objects. *Mathematics of Operations Research* **16**, 176-194.

European Journal of Operations Research **44**, 133-306, 1990.

Harrison, J. M. 1988. Brownian Models of Queueing Networks with Heterogeneous Customer Populations, in W. Fleming and P. L. Lions (eds.), *Stochastic Differential Systems, Stochastic Control Theory and Applications*. IMA Volume **10**, Springer-Verlag, New York, 147-186.

Johnson, H. T. and R. S. Kaplan. 1987. *Relevance Lost: The Rise and Fall of Management Accounting*. Harvard Business School Press, Boston.

Krichagina, E.V., S. X. C. Lou. S. P. Sethi and M. I. Taksar 1992a. Production Control in a Failure-Prone Manufacturing System: Diffusion Approximation and Asymptotic Optimality. Submitted for publication.

Krichagina, E. V., S. X. C. Lou and M. I. Taksar 1992b. Double Band Policy for Stochastic Manufacturing Systems in Heavy Traffic. Submitted for publication.

Miller, B. L. 1974. Dispatching from Depot Repair in a Recoverable Item Inventory System: On the Optimality of a Heuristic Rule. *Management Science* **21**. 316-325.

Ou, J. and L. M. Wein. 1992. Dynamic Scheduling of a Production/Inventory System with By-Products and Random Yield. Submitted to *Management Science*.

Puterman. M. L. 1975. A Diffusion Process Model for a Storage System. In *Logistics*. M. A. Geisler. ed., North-Holland, Amsterdam, 143-159.

Veatch, M. H. and L. M. Wein. 1992. Scheduling a Make-to-Stock Queue: Index Policies and Hedging Points. Technical Report, Operations Research Center, MIT.

Wein, L. M. 1992. Dynamic Scheduling of a Multiclass Make-to-Stock Queue. *Operations Research* **40**, 724-735.

Zipkin, P. 1990. An Alternative Dynamic Scheduling Policy. Working notes, Graduate School of Business, Columbia Univ., New York.