CDC01-REG2025

Stochastic Production Systems: Production and Maintenance Scheduling with Finite Buffers[*]

J. J. Westman
Department of Mathematics
University of California
Box 951555
Los Angeles, CA 90095-1555, USA
E-Mail: jwestman@math.ucla.edu
URL: http://www.math.ucla.edu-/~jwestman/

E. K. Boukas [†]
Mechanical Engineering Department
École Polytechnique de Montréal
P.O. Box 6079, station “centre-ville”
Montréal, Québec, Canada, H3C 3A7
E-Mail: boukas@meca.polymtl.ca
URL: http://www.meca.polymtl.ca-/boukas/boukas.html/

F. B. Hanson [‡]
Laboratory for Advanced Computing
University of Illinois at Chicago
851 Morgan St.; M/C 249
Chicago, IL 60607-7045, USA
E-Mail: hanson@math.uic.edu
URL: http://www.math.uic.edu-/~hanson/

2001 CDC

July 22, 2001

Abstract

Consider the production of a single consumable product that is fabricated in a process of k stages that is subject to an uncertain environment. There are a number of workstations on each stage that have different operating parameters. The workstations are subject to the discrete events of repair, failure, and preventive maintenance that generate a jump in the state of the system. Between each stage of the manufacturing process is a finite buffer that holds pieces before they can be processed by the next stage. If a buffer is full, then the preceding stage cannot produce pieces since there will be no place for them to go. This formulation of a manufacturing system is a hybrid system consisting of the meeting the global production demand while locally managing the operational status of the workstations. This formulation of the optimal scheduling of production is a quasi-LQGP problem whose jumps are generated by State Dependent Poisson Processes (SDPP). A numerical example is presented to illustrate the model.

1. Introduction

There are two classes of manufacturing systems based on perspective of the plant manager. The first type is a flexible manufacturing system (FMS) that has a local perspective of part routing, following the part as it undergoes the manufacturing process, which is inherently a discrete process. The other type, is a multistage manufacturing system (MMS), which takes a global perspective in order to determine the production rates necessary to achieve the desired production goal. In an MMS the parts flowing through the various stages of the manufacturing process are typically viewed a fluid and therefore a continuous model is used to approximate the physical system. Each stage in an MMS may be viewed as an FMS. Kimemia and Gershwin [9] describe in detail the differences and similarities between FMS and MMS while providing a hierarchical scheme and algorithm for the operational control of an FMS.

In Westman, Hanson and Boukas [14], an optimal production scheduling was presented for a manufacturing system consisting of k stages that produces a single consumable good. The model for the production scheduling

[*] Preprint of paper accepted in *Proceedings of of 41st Conference on Decision and Control, Orlando, Florida, December 4-7, 2001*

[†] Work supported in part by the Natural Sciences and Engineering Research Council of Canada under grants OGP0036444

[‡] Work supported in part by the National Science Foundation Grant DMS-99-73231.

was formulated as a quasi-LQGP problem. On each stage, there were a number of workstations that were subject to the events of failure, repair, and preventive maintenance. These event transitions, jumps in the value of the state, are modeled using state dependent Poisson processes (SDPP) whose coefficients were parameterized by the current value of the state. This manufacturing system is a hybrid system in the sense that the continuous model for meeting the production goal is that of an MMS, with local considerations for the discrete event transitions of the individual workstations on each stage that cause jumps in the value of the state, which is similar to an FMS.

In this paper we reformulate the optimal production scheduling model presented in [14] to include finite buffers between the stages. In [14], the stages are assumed to have infinite buffers between stages. This causes a decoupling of the stages and additional handling is needed to clear the buffers. In this paper, all of the stages are coupled and the production of pieces for stage i is restricted to be less than or equal to the total capacity, production plus buffer capacity, of stage $i+1$; see Figure 1. This new formulation greatly changes the characteristics of the manufacturing system. The state space decomposition of the LQGP and quasi-LQGP problem allow for larger dynamical systems to be modeled since they are not subject to the *Curse of Dimensionality* [1] of dynamic programming in state space.

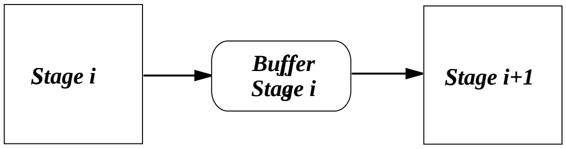

Figure 1: In between the stages there is a buffer with finite capacity to hold pieces that have completed i stages but can not be processed immediately by stage $i+1$.

The model formulation here is similar to that of Sethi, Zhang, and Zhang [10] and Sethi and Zhang [11] of the dynamical or N-machine flowshop. In these models, each stage only consists of one machine and no maintenance is performed. The models are analyzed using asymptotics for the hierarchal production planning for the flowshop with machines that can fail and be repaired. In the model presented here, each stage can have a number of different machines and therefore the constraints are more complex and change with the value of the state, which is subject to a larger number of discrete jumps. The use of the quasi-LQGP problem allows for the ability to model larger and more complex manufacturing systems than those presented in [11, 10].

Motivation for the manufacturing model presented here as well as that of [14] can be found in the fabrication of semiconductors (see the online tutorials [3, 8, 7]). The fabrication process can be viewed globally as two operations that are commonly referred to as the front- and back-end. The front-end is where the actual wafers are produced and the back-end is where the wafer processed and packaged into individual integrated circuits. The overall fabrication with two stages would be modeled as that of paradigm presented in [14]. However, the back-end and subtasks of the front-end would best be modeled as the manufacturing system paradigm presented here with finite buffers between the processing stages. The determination of which paradigm to use depends on the relative processing times, ability to store excess pieces, and on the physical parameters and layout of the workstations in the various stages.

The front-end is where the wafer containing many of the individual dies (each of which will become an integrated circuit) is fabricated. The process begins by growing the silicon ingot from a seed crystal using a pulling technique. Thin wafers are sliced off of the ingot, ground to the desired diameter, polished, a layer of additional ultra-pure crystalline silicon is deposited by epitaxy in order to insure the surface is free of defects, know as the *epi-layer*, and finally the epi-layer is exposed to high temperatures to form an oxide layer. After the wafer is prepared, a series processes are performed to grow the 16 to 24 layers that make up the integrated circuit. The first step of the process is an application of photoresist material, next photolithography using a reticle or mask outlines the patterns for the material to be deposited, which is done for one or several dies at a time and then is *stepped* to the next position, then the photoresist is developed to generate the patterns. Next an etching process removes the oxide in places where the photoresist pattern is not present generating channels in the substrate, then the photoresist material is stripped from the from the wafer. Dopant material of the desired type is then diffused onto the surface of the channels or implanted into the surface of the silicon via an ion beam. The dopant material produces the source and drain for the CMOS transistor. The gate of the transistor is formed by repeating the photolithography process, chemical vapor deposition is used to grow a thin oxide layer to act as an insulator between the gate and the silicon, finally the gate region is grown using *sputtering* (physical vapor deposition) of a conductive polysilicon. Oxides are the grown to isolate and protect the transistors from the effects of surrounding transistors. Connections to the regions of the transistor are formed by using photolithography to drill holes or *vias*, aluminum is then deposited to fill the vias, the excess aluminum is then removed, a dielectric isolation oxide is grown and chemical mechanical planarization is used to polish the surface completing the layer. Subsequent layers are then grown and vias are used to connect the various layers. During the growth of the layers testing is performed to insure the quality of the dies.

The back-end is where the wafers are processed and packaged. The first step in this procedure is to test the dies on the wafer and mark all of the unacceptable dies. The wafer is then *diced* to remove the good dies (unacceptable dies are discarded). The dies are then bonded or mounted to the frame of the package. The connect pads on the surface of the dies need to be connected to the corresponding pin of the packaging. This is traditionally done using a fine gold or aluminum wire, however new technologies such as flip-chip can also be used to make these connections. The bonded die and frame are then encapsulated finishing the fabrication process. Finally the complete integrated circuits are tested.

The back-end can be viewed as an FMS or an MMS where the various subtasks are the stages. Normally in the automated fabrication plant an assembly line is used to package the integrated circuits where there is a finite buffer between the stages. The manufacturing paradigm presented here can be used to determine the optimal scheduling of production for the back-end. Production scheduling for the entire fabrication process using two stages, the front and back ends, would fall under the paradigm of [14]. A combination of these two paradigms can be utilized for a multistage fabrication process using the material presented here, where decoupled stages have a large buffer size.

The paper is arranged as follows. In Section 2., a summary of the quasi-LQGP Problem [14] with state dependent Poisson processes [13] is presented for completeness. In Section 3., a quasi-LQGP problem utilizing state dependent Poisson processes is used to formulate the dynamical system for the manufacturing system and in Section 4., a numerical example is presented.

2. Summary of Quasi-LQGP Problem

The canonical form for the LQGP problem used here appears in Westman and Hanson [12], for the case with state independent Poisson noise, and in [13] for state dependent Poisson processes. The LQGP problem is characterized by *linear* deterministic dynamics, *quadratic* costs, *Gaussian* noise disturbance, and *Poisson* jumps in the state value. In the *quasi-LQGP* problem [14] the LQGP problem is expanded with the dynamic and cost coefficients to be parameterized by the value of the state. This allows for greater flexibility of modeling in a convenient notational form. Additionally, the cost functional used is the full quadratic form which extends the classic LQGP cost functional. Considerations for modeling physical systems are summarized, as well as formal solution to the LQGP problem.

The quasi-linear dynamical system for the quasi-LQGP problem is governed by the stochastic differential equation (SDE) subject to a Gaussian process and state dependent Poisson processes (SDPPs) disturbances is given by

$$\begin{aligned} d\mathbf{X}(t) &= [A(t;\mathbf{X}(t))\mathbf{X}(t) + B(t;\mathbf{X}(t))\mathbf{U}(t) + \mathbf{C}(t;\mathbf{X}(t))]dt + G(t;\mathbf{X}(t))d\mathbf{W}(t) \qquad (1) \\ &+ [H_1(t;\mathbf{X}(t))\mathbf{X}(t)]d\mathbf{P}_1(\mathbf{X}(t),t) + [H_2(t;\mathbf{X}(t))\mathbf{U}(t)]d\mathbf{P}_2(\mathbf{X}(t),t) + H_3(t;\mathbf{X}(t))d\mathbf{P}_3(\mathbf{X}(t),t), \end{aligned}$$

for general Markov processes in continuous time, with m × 1 state vector **X**(t), n × 1 control vector **U**(t), r × 1 Gaussian noise vector d**W**(t), and q_ℓ × 1 space-time Poisson noise vectors $d\mathbf{P}_\ell(\mathbf{X}(t),t)$, for $\ell = 1$ to 3. Note that the term $[H_1(t;\mathbf{X}(t)) \cdot \mathbf{X}(t)]d\mathbf{P}_1(\mathbf{X}(t),t)$ is not linear in the state. The dimensions of the respective coefficient matrices are: $A(t;\mathbf{x})$ is $m \times m$, $B(t;\mathbf{x})$ is $m \times n$, $\mathbf{C}(t;\mathbf{x})$ is $m \times 1$, $G(t;\mathbf{x})$ is $m \times r$, while the $H_\ell(t;\mathbf{x})$ are dimensioned, so that $[H_1(t;\mathbf{x}) \cdot \mathbf{x}] = [\sum_k H_{1ijk}(t;\mathbf{x})x_k]_{m \times q_1}$, $[H_2(t;\mathbf{x}) \cdot \mathbf{u}] = [\sum_k H_{2ijk}(t;\mathbf{x})u_k]_{m \times q_2}$ and $H_3(t) = [H_{3ij}(t;\mathbf{x})]_{m \times q_3}$. The coefficients for the dynamical system can depend on the value of the state as a parameter, and numerically must be evaluated first, i.e., we assume the coefficients are subdominant or are locally in the state. Note that the space-time Poisson terms are formulated to maintain the linear nature of the dynamics, but the first two are actually bilinear in either $\mathbf{X}(t)$ or $\mathbf{U}(t)$ with $d\mathbf{P}_\ell$ for $\ell = 1$ or 2, respectively.

The SDPP can be viewed as a sequence of events that is represented by its ith couple $\{T_i(\mathbf{X}(T_i)), M_i(\mathbf{X}(T_i))\}$, for $i = 1$ to k, where $T_i(\mathbf{X}(T_i))$ is the time for the occurrence of the ith jump with state dependent mark amplitude $M_i(\mathbf{X}(T_i))$. The form for the SDPP allows for a great deal of realism to be included in the model for deterministic and random jumps in the evolution of the state values. A wider range of stochastic control applications can be accurately modeled since the arrival rates and amplitudes can depend on the value of the state. The state dependent vector valued marked Poisson noises are related to the Poisson random measure (see Gihman and Skorohod [4] or Hanson [6]) and are defined as $d\mathbf{P}_\ell(\mathbf{X}(t),t) = \left[\int_{\mathcal{Z}_{\ell,i}} z_i \mathcal{P}_{\ell,i}(dz_i, \mathbf{X}(t), dt)\right]_{q_\ell \times 1}$ for $\ell = 1$ to 3 which consists of q_ℓ independent differentials of space-time Poisson processes that are functions of the state, $\mathbf{X}(t)$, where z_i is the Poisson jump amplitude random variable or the mark of the $dP_{\ell,i}(\mathbf{X}(t),t)$ Poisson process where $\ell = 1$ to 3 and $i = 1$ to q_ℓ. The mean or expectation is given by $\mathrm{Mean}[d\mathbf{P}_\ell(\mathbf{X}(t),t)] = \Lambda_\ell(\mathbf{X}(t),t)\overline{\mathbf{Z}}_\ell(\mathbf{X}(t),t)dt$ where $\Lambda_\ell(\mathbf{X}(t),t)$ is the diagonal matrix representation of the state dependent Poisson arrival rates $\lambda_{\ell,i}(\mathbf{X}(t),t)$ for $\ell = 1$ to 3 and $i = 1$ to q_ℓ, $\overline{\mathbf{Z}}_\ell(\mathbf{X}(t),t)$ is the mean of the jump amplitude mark vector and $\phi_{\ell,i}(z_i, \mathbf{X}(t), t)$ is the density of the (ℓ, i)th amplitude mark component. Assuming component-wise independence, $d\mathbf{P}_\ell(\mathbf{X}(t),t)$ has covariance given by Covar$[d\mathbf{P}_\ell(\mathbf{X}(t),t), d\mathbf{P}_\ell^\top(\mathbf{X}(t),t)] = \Lambda_\ell(*)\sigma_\ell(*)dt$ with, for instance, $\sigma_\ell(*) = \sigma_\ell(\mathbf{X}(t),t) = [\sigma_{\ell,i,j}\delta_{i,j}]_{q_\ell \times q_\ell}$ denoting the diagonalized covariance of the amplitude mark distribution for $d\mathbf{P}_\ell(\mathbf{X}(t),t)$. Again, the mark vector is not assumed to have a zero mean, i.e., $\overline{\mathbf{Z}}_\ell \neq 0$, permitting additional modeling complexity. Note, that for discrete distributions the above integrals need to be replaced by the appropriate sums. The Gaussian white noise term, $d\mathbf{W}(t)$, consists of r independent, standard Wiener processes $dW_i(t)$, for $i = 1$ to r. These Gaussian components have zero infinitesimal mean, Mean$[d\mathbf{W}(t)] = \mathbf{0}_{r\times 1}$ and and diagonal covariance. Covar$[d\mathbf{W}(t), d\mathbf{W}^T(t)] = I_r dt$. It is further assumed that all of the individual component terms of the Gaussian noise are independent of all of the Poisson processes.

The impact of stochastic processes on a physical system has a great influence on how the state values will evolve. It is difficult to determine coefficients for the stochastic processes in (1). Using historical data from the process to be modeled allows for the determination of all of the coefficients for the dynamical system. The data points need to be divided into sets in which jumps are not present and the value of the state is stable. The statistical moments of the individual sets can then be used to determine the coefficients for the deterministic dynamics and the background Gaussian process by matching them with the moments for the dynamical system presented below (3,4). After these coefficients have been determined, the data points corresponding transitions or jumps between the sets are used to determine the coefficients for the SDPP based on the current value for the state. These coefficients need to capture the effects of these impulses on the system properly. To do this, we need to match the moments of the jumps to that of the dynamical system (3,4) and the impact of the jumps on the value of the state (2). An important feature of the quasi-LQGP problem is the ability for these coefficients to change as the system evolves in time. This is also important if the amount of initial data is small and estimates are used, since the coefficients can be dynamically modified as the physical system evolves over time. The jth jump of the $\{\ell, i\}$th space-time Poisson process at time $t_{\ell,i,j}$ with mark amplitude $z_i = M_{\ell,i,j}$ causes the following jump from $t^-_{\ell,i,j}$ to $t^+_{\ell,i,j}$ in the state:

$$[\mathbf{X}](t_{\ell,i,j}) = \left\{ \begin{array}{ccc} [H_1(t_{\ell,i,j};\mathbf{X}(t^-_{\ell,i,j}))\mathbf{X}(t^-_{\ell,i,j})]_i M_{\ell,i,j}, & \text{if} & \ell = 1 \\ [H_2(t_{\ell,i,j};\mathbf{X}(t^-_{\ell,i,j}))\mathbf{U}(t^-_{\ell,i,j})]_i M_{\ell,i,j}, & \text{if} & \ell = 2 \\ [H_3(t_{\ell,i,j};\mathbf{X}(t^-_{\ell,i,j}))]_i M_{\ell,i,j}, & \text{if} & \ell = 3 \end{array} \right\}. \tag{2}$$

From the above statistical properties of the stochastic processes, $d\mathbf{W}$ and $d\mathbf{P}_\ell$, it follows that the first two conditional infinitesimal moments of the state, fundamental for modeling applications, are

$$\text{Mean}\left[d\mathbf{X}(t) \middle| \begin{array}{c} \mathbf{X}(t) = \mathbf{x} \\ \mathbf{U}(t) = \mathbf{u} \end{array} \right] = \begin{array}{l} \big[A(t;\mathbf{x})\mathbf{x} + B(t;\mathbf{x})\mathbf{u} + \mathbf{C}(t;\mathbf{x}) + [H_1(t;\mathbf{x})\mathbf{x}](\Lambda_1\overline{\mathbf{Z}}_1)(\mathbf{x},t) \\ +\, [H_2(t;\mathbf{x})\mathbf{u}](\Lambda_2\overline{\mathbf{Z}}_2)(\mathbf{x},t) + H_3(t;\mathbf{x})(\Lambda_3\overline{\mathbf{Z}}_3)(\mathbf{x},t)\big]\, dt, \end{array} \tag{3}$$

and the conditional infinitesimal covariance,

$$\text{Covar}\left[d\mathbf{X}(t) \middle| \begin{array}{c} \mathbf{X}(t) = \mathbf{x} \\ \mathbf{U}(t) = \mathbf{u} \end{array} \right] = \left\{ \begin{array}{l} \big[(GG^T)(t;\mathbf{x}) + [H_1(t;\mathbf{x})\mathbf{x}](\Lambda_1\sigma_1)(\mathbf{x},t)[H_1(t;\mathbf{x})\mathbf{x}]^\top \\ +[H_2(t;\mathbf{x})\mathbf{u}](\Lambda_2\sigma_2)(\mathbf{x},t)[H_2(t;\mathbf{x})\mathbf{u}]^\top \\ +\, H_3(t;\mathbf{x})(\Lambda_3\sigma_3)(\mathbf{x},t)H_3^\top(t;\mathbf{x})\big]\, dt. \end{array} \right\} \tag{4}$$

Notice that the conditional infinitesimal mean of the evolution of the state depends only on the linear deterministic dynamics and the contributions from jumps due to the SDPP, whereas the conditional infinitesimal covariance depends only on the small fluctuations from the Gaussian process and the jumps from the SDPP.

The cost functional or performance index that is used is given by the *time-to-go* or *cost-to-go* functional form:

$$V[\mathbf{X},\mathbf{U},t] = \frac{1}{2}\mathbf{X}^\top(t_f)S(t_f)\mathbf{X}(t_f) + \int_t^{t_f} C(\mathbf{X}(\tau),\mathbf{U}(\tau),\tau)d\tau, \tag{5}$$

where the time horizon is (t, t_f), with $S(t_f) \equiv S_f$ is the quadratic final cost coefficient matrix and $C(\mathbf{x},\mathbf{u},t)$ is quadratic running cost function. The form for the instantaneous cost employed here is the general, quasi-quadratic form:

$$C(\mathbf{x},\mathbf{u},t) = \frac{1}{2}\left[\mathbf{x}^\top Q_2(t;\mathbf{x})\mathbf{x} + \mathbf{u}^\top R_2(t;\mathbf{x})\mathbf{u} + \mathbf{x}^\top C_2(t;\mathbf{x})\mathbf{u}\right] + \mathbf{Q}_1^\top(t;\mathbf{x})\mathbf{x} + \mathbf{R}_1^\top(t;\mathbf{x})\mathbf{u} + C_0(t;\mathbf{x}) \tag{6}$$

where the coefficient matrices are of the appropriate dimension. In order to minimize (5) with instantaneous cost function (6) requires that the quadratic control cost coefficient the quadratic control cost coefficient, $R_2(t;\mathbf{x})$, is a symmetric positive definite $n \times n$ array, while the quadratic state control coefficient, $Q_2(t;\mathbf{x})$, is assumed to be a symmetric positive semi-definite $m \times m$ array.

The quasi-LQGP problem is defined by (1, 5, 6) and the approach of stochastic dynamic programming is used to obtain a solution. Here only the material for determining the regular control is presented, for complete details see [14]. In order to determine a solution, the state domain is decomposed into subdomains, $\mathcal{D}_\mathbf{x} = \bigcup_i \mathcal{D}_{x_i}$, where the arrival rates and moments for the Poisson processes and the coefficients for the dynamics and costs are independent of the state. The regular, unconstrained optimal control, $\mathbf{u}^* = \mathbf{u}_{\rm reg}$, for the region $\mathcal{D}_{x_i}$ dropping the subscripts i is given by

$$\mathbf{u}_{\rm reg}(t) = -\widehat{R}_2^{-1}(t)\left[\widehat{B}(t)\mathbf{x} + \widehat{\mathbf{D}}(t)\right], \tag{7}$$

where $\widehat{R}_2(t)$, given below, is related to $R_2(t)$ but with Poisson term corrections. For the region $\mathcal{D}_{x_i}$, dropping the subscripts i and assuming regular control, the coefficients for the optimal expected performance are given by

$$0_{m\times m} = \dot{S}(t) + \left[A^\top S + SA + Q_2 + \widetilde{\Gamma}_1 - \widehat{B}^\top\widehat{R}_2^{-1}\widehat{B}\right](t), \tag{8}$$

$$\mathbf{0}_{m\times 1} = \dot{\mathbf{D}}(t) + \left[\left(A + H_1^\top\Lambda_1\overline{\mathbf{Z}}_1\right)^\top\mathbf{D} + \mathbf{Q}_1 + S\left(\mathbf{C} + H_3\Lambda_3\overline{\mathbf{Z}}_3\right) - \widehat{B}^\top\widehat{R}_2^{-1}\widehat{\mathbf{D}}\right](t), \tag{9}$$

$$0 = \dot{E}(t) + \left[\left(\mathbf{C} + H_3\Lambda_3\overline{\mathbf{Z}}_3\right)^T\mathbf{D} + C_0 + \tfrac{1}{2}\widetilde{\Gamma}_3 - \tfrac{1}{2}\widehat{\mathbf{D}}^\top\widehat{R}_2^{-1}\widehat{\mathbf{D}}\right](t), \tag{10}$$

where

$$\Gamma_1(t) \equiv \left[\left([H_1^T]_i S[H_1]_j\right) : \left(\Lambda_1\overline{ZZ}_1\right)(t)\right]_{m\times m} + 2\left[\left(\Lambda_1\overline{\mathbf{Z}}_1\right)^T H_1^T S\right](t),$$

$$\Gamma_2(t) \equiv \left[\left([H_2^T]_i S[H_2]_j\right) : \left(\Lambda_2\overline{ZZ}_2\right)(t)\right]_{n\times n}, \ \Gamma_3(t) \equiv \left[\left(H_3^T S H_3\right) : \left(\Lambda_3\overline{ZZ}_3\right)\right](t), \ \widetilde{\Gamma}_\ell(t) \equiv \left(\Gamma_\ell + \Gamma_\ell^T\right)(t)$$

$$\overline{ZZ}_\ell(t) \equiv \overline{\sigma}_\ell(t) + \left(\overline{\mathbf{Z}}_\ell\overline{\mathbf{Z}}_\ell^T\right)(t) = \left[\sigma_{\ell,i}\delta_{i,j} + \overline{Z}_{\ell,i}\overline{Z}_{\ell,j}\right]_{q_\ell\times q_\ell}, \quad \widehat{R}_2(t) \equiv R_2(t) + \widetilde{\Gamma}_2(t),$$

$$\widehat{B}(t) \equiv \left[\left(B(t) + H_2\Lambda_2\overline{\mathbf{Z}}_2\right)^T S + \tfrac{1}{2}C_2^\top\right](t), \qquad \widehat{\mathbf{D}}(t) \equiv \left[\left(B(t) + H_2\Lambda_2\overline{\mathbf{Z}}_2\right)^T\mathbf{D} + \mathbf{R}_1\right](t),$$

for $\ell = 1$ to 3, where $A : B = \sum_i\sum_j A_{i,j}B_{i,j} = \text{Trace}[AB^\top]$, with initial conditions $S(t_f) = S_f$, $\mathbf{D}(t_f) = \mathbf{0}$, and $E(t_f) = 0$. Since the matrix R_2 is positive definite, R_2^{-1} exists and then so does $\widehat{R}_2^{-1}$. Note (8) appears to have Riccati-like quadratic form, but in general is highly nonlinear due to the $S(t)$ dependence on $\widehat{R}_2(t)$ through $\Gamma_2(t)$. If $H_\ell = [H_{\ell,i,j,k}]_{m\times q_\ell\times m_\ell}$, then $H_\ell^T = [H_{\ell,j,i,k}]_{q_\ell\times m\times m_\ell}$. Due to uni-directional coupling of these matrix differential equations, it is assumed that the nonlinear matrix differential equation (8) for $S(t)$ is solved first and the result for $S(t)$ is substituted into equation (9) for $\mathbf{D}(t)$, which is then solved, and then both results for $S(t)$ and $\mathbf{D}(t)$ are substituted into equation (10) for $E(t)$.

3. Production Scheduling Manufacturing System Model with Finite Buffers

Consider a manufacturing process that requires k sequential steps to fabricate a single consumable commodity, for example integrated circuits. The planning horizon for the production horizon is $[0, t_f]$ with a demand of $d(t)$ pieces per unit time. The *loading* and *unloading* stages, the means by which raw materials are introduced and finished goods exit the manufacturing system, respectively, are not considered. The model employs MMS-like criteria for determining the optimal production rates for the workstations on all stages and utilizes FMS-like considerations for the local management of the workstations. The model presented here is based on the work of Westman, Hanson, and Boukas [14] that uses the quasi-LQGP problem as the basis for the model including uniform penalties for shortfalls and surpluses of production on all stages and a scheme was presented to insure that pieces were not left in buffers. In [14], a measure for the excess production on each stage that could be positive or negative was used that decoupled all stages of the manufacturing system. In this paper, the manufacturing system is formulated as an assembly line or flowshop (see [11, 10]) with physical buffers between the stages and a measure for the excess production on the last stage. The number of pieces in a physical buffer can never be negative and imposes additional constraints on the production rates not present in [14].

In [11, 10], each stage consists of a single machine that can fail and repair that introduces restrictions on the number of pieces that can be produced. This fluctuation in production is implemented as by an inequality constraint in which the upper bound is a stochastic process. A significant difference with this work is that each stage consists of a number of workstations that have different operating parameters and are subject to the events of repair, failure and preventive maintenance. This greatly complicates the problem since each workstation on each stage can be in a greater number of states and therefore the operational status must be tracked. This leads to a very large dimensional system, which in the model presented here does not suffer from the *Curse of Dimensionality* [1] since the quasi-LQGP problem is the basis of the model. This allows for larger systems to be modeled and implemented in near real time (the results presented in the next section required 11 seconds wall clock on a workstation). Additionally, the processes that effect the operational status of the workstations are modeled using a collection of state dependent Poisson processes (SDPP) [13] which can be dependent on time and the status of the workstation and therefore the processes themselves can evolve in time, for example aging machines may fail more frequently which can be incorporated in the model. A great advantage of this formulation is that the quasi-LQGP problem allows for the coefficients of the dynamics to be parameterized by the value of the state and that SDPPs moments can be functions of the state thereby removing problems associated with changes in the operational dynamics when a jump occurs.

3.1. Local Workstation State Equations

The modeling and evolution of the operational status of all of the workstations is the same as in Westman, Hanson, and Boukas [14]. It is important to note that the local considerations of the workstations for all stages are very similar to that of a FMS, except we follow the machines and not individual pieces. The operational status a given workstation evolves according a stochastic differential equation utilizing SDPPs to generate the impulses for transitions between the different states which is represented in Figure 2. For stage k, assume there are N_k workstations that have different

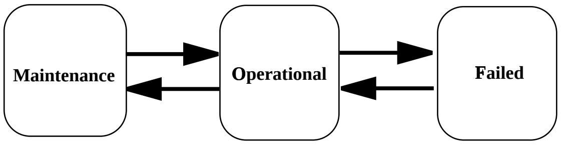

Figure 2: Any given workstation can be in one of the three states (Operational, Failed, Maintenance) which are mutually exclusive with transitions between the states are modeled as SDPPs.

operational parameters, therefore, the status for each workstation must be followed. The tracking of workstation events along with the operational age of the workstation are the state variables for a given workstation on a given stage. This leads to a high dimensional state space, however since the model is in the form of the expanded LQGP problem presented in Section 2. and does not suffer from the *Curse of Dimensionality*. The state variables for workstation i on stage k are the available production capacity, $r_{ki}(t)$, the operational status (repair/failure), $o_{ki}(t)$, the preventive maintenance status, $m_{ki}(t)$, and the operational age, $a_{ki}(t)$, or in vector form,

$$\mathbf{x}_{ki}(t) = \left[r_{ki}(t),\ \ o_{ki}(t),\ \ m_{ki}(t),\ \ a_{ki}(t)\right]^{\top} . \tag{11}$$

The production for a given stage is distributed across all of the workstations evenly. The production rate, $c_k(t)$, is a parameter for each stage k, representing the utilization or the fraction of time busy. The goal of the optimal control problem for the production scheduling is to determine the production rates for each stage for the planning horizon. The production rates need to compensate for changes in the status of the workstations while maintaining the desired constraints on meeting the production demand in the specified way.

The arrival rate, mean time till an event occurs, for failures is dependent on the operational age of the workstation. This implies that the probability of a failure is an increasing function of the operational age of the workstation. Therefore, preventive maintenance is performed periodically based on the operational age of the workstation, which reduces the operational age of the workstation and thereby reduces the probability for a failure to occur. It is assumed that the election to perform preventive maintenance is rational, that is the amount of time perform preventive maintenance is much less than that of repairing a failed machine and/or the cost for preventive maintenance is much less than that of repair. In this treatment, preventive maintenance reduces the available production capacity for the workstation by a fixed percentage assumed to be greater than 0. The operational and preventive maintenance status values evolve according to stochastic differential equations using SDPPs with coefficients that are parameterized by the value of the state. These coefficients partition the state space into regions in which events can occur based on the value of the state ensuring that events can only occur when allowable. The status values are in the range $[0, 1]$ which represents the maximum percentage of available production capacity. The available production capacity is an indicator that determines that the state of a workstation. The available production capacity for workstation i on stage k should be given by $r_{ki}(t) = \mathrm{Min}[o_{ki}(t), m_{ki}(t)]$, but unfortunately this nonlinear expression is not allowable under the LQGP problem. Instead, using the mutually exclusivity of the events, we introduce another variable for the production state of each workstation defined by

$$dr_{ki}(t) \quad = \quad do_{ki}(t) + dm_{ki}(t). \tag{12}$$

Let M_{ki} represent the maximum number of pieces that can be produced per unit time on workstation i, then the total number of pieces that can be produced on stage k at time t is $\widehat{M}_k(t) = \sum_{i=1}^{N_k} M_{ki} r_{ki}(t) \equiv \mathbf{M}_k^\top \mathbf{r}_k(t)$.

The mean time between failures and the time for repair are assumed to be exponentially distributed and dependent on the operational age of the workstation. The form of the defining equation for the operational status of the workstations is modeled from (1) as the following term:

$$do_{ki}(t) = H_3(t; \mathbf{x}_{ki}(t)) d\mathbf{P}_3(\mathbf{x}_{ki}(t), t) = \left[\begin{array}{cc} \delta_{o_{ki}(t),0} & 0 \\ 0 & -\delta_{r_{ki}(t),1} \end{array} \right] \left[\begin{array}{c} d\mathbf{P}_{ki}^R(\mathbf{x}_{ki}(t), t) \\ d\mathbf{P}_{ki}^F(\mathbf{x}_{ki}(t), t) \end{array} \right], \tag{13}$$

where the superscripts R and F denote repair and failure processes, respectively. The coefficient matrix, $H_3(t; \mathbf{X}(t))$, is parameterized by the state so that only allowable events may occur which partitions the state space into regions. The $d\mathbf{P}_3(\mathbf{X}(t), t)$ is the SDPP providing the jumps for workstation repair and failure processes with the following properties, $1/\lambda^R(\mathbf{x}_{ki}(t), t) = T_{ki}^R$, $1/\lambda^F(\mathbf{x}_{ki}(t), t) = T_{ki}^F - a_{ki}(t)$, $\overline{Z}_{ki}^R = \overline{Z}_{ki}^F = 1$ and $\sigma_{ki}^R = \sigma_{ki}^F = 0$, where T_{ki}^R and T_{ki}^F are the mean times between repair and failure, respectively. The operational status formulated here can either have a value of 1 which denotes an operational workstation or 0 which denotes a failed workstation not capable of production.

The effects of maintenance are considered only if the maintenance will be performed in the remaining production horizon. It is assumed that preventive maintenance reduces the amount of available production capacity, but does not necessarily disable production. The SDPPs are used to generate the jumps in the state for the events of beginning maintenance (denoted with a superscript of M) and for the completion of maintenance (denoted with a superscript of D). The defining equation for the preventive maintenance status is given by

$$dm_{ki}(t) = \left[\begin{array}{cc} 1 - \delta_{r_{ki}(t),1} & 0 \\ 0 & -c_{ki}^M(t; a_{ki}(t)) \delta_{r_{ki}(t),1} \end{array} \right] \left[\begin{array}{c} d\mathbf{P}_{ki}^D(\mathbf{x}_{ki}(t), t) \\ d\mathbf{P}_{ki}^M(\mathbf{x}_{ki}(t), t) \end{array} \right], \tag{14}$$

where

$$c_{ki}^M(t; a_{ki}(t)) = \left\{ \begin{array}{lll} 1, & \text{if} & T_{ki}^M - a_{ki}(t) < ttg(t) \\ 0, & \text{if} & T_{ki}^M - a_{ki}(t) \geq ttg(t) \end{array} \right\} \tag{15}$$

is used to ensure that maintenance occurs with in the production horizon with the time-to-go for the production horizon given by $ttg(t) = T - t$. The sojourn times for these processes are given by $1/\lambda^D(\mathbf{x}_{ki}(t), t) = T_{ki}^D$ and $1/\lambda^M(\mathbf{x}_{ki}(t), t) = T_{ki}^M - a_{ki}(t)$, where T_{ki}^D and T_{ki}^M are the average duration of the maintenance and the mean time between preventive maintenance, respectively. It is assumed that preventive maintenance should be performed before a failure which implies $T_{ki}^M < T_{ki}^F$. The moments for the SDPP for preventive maintenance should be modeled as the average loss of production capacity, $\overline{Z}_{ki}^M$, and the variance of the loss of production capacity σ_{ki}^M. The duration of preventive maintenance process (D) is used to restore the value for the preventive maintenance status to 1 and has moments given by $\overline{Z}_{ki}^D = \overline{Z}_{ki}^M$ and $\sigma_{ki}^D = 0$.

The current operational age of the workstation is a monotone increasing function of time and of the number of pieces produced which can be reset to a lower level by the performance of repair or preventive maintenance. The age of the workstation evolves according to:

$$da_{ki}(t) = f(c_k(t), t)dt - H_{ki}^D(t; \mathbf{x}_{ki}(t))dP_{ki}^D(\mathbf{x}_{ki}(t), t) - H_{ki}^R(t; \mathbf{x}_{ki}(t))dP_{ki}^R(\mathbf{x}_{ki}(t), t), \tag{16}$$

where $H_{ki}^D(t; \mathbf{x}_{ki}(t))$ and $H_{ki}^R(t; \mathbf{x}_{ki}(t))$ are the coefficients that are used to reset the operational age due to maintenance and repair to a specified lower level, respectively, with $a_{ki}(\tau_{ki}) = \overline{a}(\tau_{ki})$ where τ_{ki} is the time of the last reset.

3.2. Global Buffer State Equation and Cost Functional

In this treatment, the buffer level, $b_k(t)$, represents the physical number of pieces in the buffer for each stage $k = 1, \ldots, N-1$ at time t and is bounded by $0 \leq b_k(t) \leq B_k$, where B_k is the size of the buffer. The buffer level for the final stage $k = N$ represents the surplus, if positive, and shortfall, if negative, of production and is bounded above by $b_N(t) \leq B_N$, where B_N is the size of the holding buffer for the finished goods. The ideal for the manufacturing system is to have $b_k(t) = 0$ for all k and t. The equation for the buffer level for stage k is given by

$$db_k(t) = \left[\mathbf{M}_k^\top \mathbf{r}_k(t)c_k(t) + u_k(t) - \hat{d}_k(t)\right] dt + g_k(t)dW_k(t). \tag{17}$$

A change in the buffer level is determined by the number of pieces that have successfully completed k stages of the manufacturing process, that are not defective, and are not consumed by stage $k+1$. The term $u_k(t)$, expressed as the number of parts per unit time, is used to adjust the production rates $c_k(t)$ to compensate for changes in workstation status and small local effects modeled as Gaussian process. The constraints on the control, $u_k(t)$, are so that the buffer levels are physically acceptable, and therefore induce constraints on the production rate. The term, $g_k(t)dW_k(t)$, is used to model the continuous random fluctuations in the number of pieces produced.

The model presented here consists of two levels. The upper or global level is where the production demand (pieces per unit time), $d(t)$, and production horizon, t_f is determined by the upper levels of management. The bottom or local level is the actual fabrication plant managed by the plant manager who is responsible for maintaining the workstations and meeting the global production demands in the specified way. A manufacturing disciple needs to be established to consume the pieces in the buffers in a specified time frame, $\tau(t)$, during the remaining production horizon. For simplicity, it is assumed that the pieces are consumed during the remaining production horizon, $\tau(t) \equiv t_f - t$, without loss of generality. Modifications to this consumption policy will result in the need to add additional events in the method described in the next section (see [14]). The effective demand for stage k, $\hat{d}_k(t)dt$, is the number of pieces per unit time demanded by stage $k+1$ ($\mathbf{M}_{k+1}^\top \mathbf{r}_{k+1}(t)c_{k+1}(t) + u_{k+1}(t)$) plus the consumption of pieces in the buffer before stage k ($b_{k-1}(t)/\tau(t)$) less the availability of pieces from the buffer after stage k ($b_k(t)/\tau(t)$) which needs to be modified for the first and last stages and is determined as:

$$\hat{d}_k(t) = \left\{ \begin{array}{ccc} \mathbf{M}_2^\top \mathbf{r}_2(t)c_2(t) - b_k(t)/\tau(t), & \text{if} & k = 1 \\ \mathbf{M}_{k+1}^\top \mathbf{r}_{k+1}(t)c_{k+1}(t) + b_{k-1}(t)/\tau(t) - b_k(t)/\tau(t), & \text{if} & i = 1, \ldots, N-1 \\ d(t) + b_{N-1}(t)/\tau(t) - b_N(t)/\tau(t), & \text{if} & i = N \end{array} \right\}, \tag{18}$$

The cost functional used is the *time-to-go* or *cost-to- go* form (5), using the state parameterized full quadratic instantaneous cost (6) that is motivated by a *zero inventory* or *Just in Time* manufacturing discipline (see Hall [5] and Bielecki and Kumar [2]) while utilizing minimum control effort. In this formulation, the cost functional employed is:

$$V[\mathbf{x}, \mathbf{s}, \mathbf{u}, t] = \frac{1}{2}\left(\mathbf{s}^\top S \mathbf{s}\right)(t_f) + \int_t^{t_f} \left[\frac{1}{2}\mathbf{u}^\top R_2(t;\mathbf{x})\mathbf{u} + \mathbf{Q}_1^\top(t;\mathbf{x})\mathbf{s}\right] d\tau \tag{19}$$

with only the surplus aggregate level, $\mathbf{s}$, of the state and the control, $\mathbf{u}$, used for the cost. The salvage cost, $S(t_f)$, is used to impose a penalty on surplus or shortfall of production at the end of the planning horizon. The positive definite vector $\mathbf{Q}_1(t;\mathbf{x})$ is the cost used to penalize shortfall and surplus production during the planning horizon, maintaining a strict regimen on when the consumable goods are to be produced and is parameterized by the state given by

$$Q_{1,k}(t;\mathbf{x}) = \left\{ \begin{array}{ccc} -\overline{Q_{1,k}^-}, & \text{if} & s_k(t) < 0 \\ \overline{Q_{1,k}^+}, & \text{if} & s_k(t) \geq 0 \end{array} \right\}, \tag{20}$$

where $\overline{Q_{1,k}^-}$ and $\overline{Q_{1,k}^+}$ are positive constant coefficients, with $\overline{Q_{1,k}^-} = 0$ for $k = 1, \ldots, (N-1)$. The positive definite matrix $R_2(t;\mathbf{x})$ is used to enforce a minimum control effort penalty similar to that of (20).

3.3. Computational Considerations

The numerical solution for this optimal control problem is based on the current value of the state, which then can be used to determine the regular, or unconstrained control (7). Since the regular control used, the resulting production rates may not be physically realizable and therefore need to be restricted (see [14]). Additionally, the production rates need to be restricted further so that the buffer levels are never exceeded during the remainder of the production horizon. It is assumed that the status of the manufacturing system is initially known. The algorithm follows the discrete events of the manufacturing system and computes the production rates for the remaining production horizon after events occur. The events considered are the start of production, workstation repair and failure, and the start and end of preventive maintenance. The trajectory following method presented here yields a better control and outcome of the manufacturing system than that of static policies in which only the start event is considered. Let $n_k(t)$ denote the maximum number of pieces that can be produced by stage k based on the current operational status of workstations for the remainder of the time horizon that is given by $n_k(t) = \mathbf{M}_k^\top \mathbf{r}_k(t)\tau(t)$. The method given below is used to determine the production rates, where Step 1 is the initialization, and the remaining steps need to be performed for the occurrence of each event. It is assumed that for all k and t that $n_k(t) > 0$. In the degenerate case, $n_k(t) = 0$, which corresponds to total loss of production capacity, for some k and t all the production rates are set to consume the pieces in the buffers and then set to 0 until some of the production capacity is restored.

1. The state space is partitioned into subdomains, $\mathcal{D}_{\mathbf{x}} = \bigcup_i \mathcal{D}_{x_i}$, so that all coefficients and stochastic processes of (1) are locally state independent.
2. Based on the current state of the manufacturing system the appropriate locally state independent subdomain is selected and the coefficients for the dynamics and cost functional as well as the moments for the stochastic processes are determined. The values for effective demand rates (18) are viewed as constants.
3. The system of equations (8,9,10) is solved to determine $S(t)$, $\mathbf{D}(t)$ and $E(t)$, respectively, and then are used to calculate the regular control, $\mathbf{u}_{\text{reg}}$ (7).

4. The regular controlled production rates are determined by:

$$c_k^{\text{reg}}(t) = \left\{ \begin{array}{ccc} c_k(t) + u_k^{\text{reg}}(t)/n_k(t), & \text{if} & k = 1, \ldots, N-1 \\ c_k(t) + (u_k^{\text{reg}}(t) + b_{k-1}(t) - b_k)/n_k(t), & \text{if} & k = N \end{array} \right\}, \tag{21}$$

where the regular control is modified for stage N to account for pieces in the buffers.

5. The physical production rates, $c_k^{\text{phy}}(t)$, are the restriction of $c_k^{\text{reg}}(t)$ to be admissible, the minimum value of the physical production rate, 1.00 or full utilization, and production limitations that arise due to a shortfall of production from the previous stage due to either machine failure, maintenance, or defective pieces as well as clearing pieces in the buffers, determined by $c_k^{\text{phy}}(t) = \min[c_k^{\text{reg}}(t), c_k^{max}(c_{k-1}^{\text{phy}}(t), t)]$ where

$$c_k^{max}(c_{k-1}^{\text{phy}}(t), t) = \left\{ \begin{array}{ccc} 1, & \text{if} & k = 1 \\ \min\left[1, \left(c_{k-1}^{\text{phy}}(t) n_{k-1}(t) + b_{k-1}(t)\right)/n_k(t)\right], & \text{if} & 1 < k \end{array} \right\}. \tag{22}$$

6. The restriction of $c_k^{\text{phy}}(t)$ so that maximum capacity of the buffers, B_k, is never exceeded given by $\hat{c}_k^*(t) = \min[c_k^{\text{phy}}(t), (d_k(t)\tau(t) + B_k)/n_k(t)]$. The constrained controlled production rates which ensure a valid production rate guaranteeing the integrity of the buffers, given by $c_k^*(t) = \min[\hat{c}_k^*(t), c_k^{max}(\hat{c}_{k-1}^*(t), t)]$, is used as the production rates for operational workstations on each stage, that is $c_k(t) = c_k^*(t)$.

4. Numerical Example for Production Scheduling

For numerical concreteness, consider a manufacturing system with $k = 2$ stages with a planning horizon of $T = 80$ hours. Initially the buffers are clear ($b_1(0) = b_2(0) = 0$) and have maximum capacities of $B_1 = 50$ and $B_2 = 100$, the demand be $d(t) = 145$ pieces per hour, the total number of workstations, N_i, for each stage be 3 and 2, respectively, the Gaussian random fluctuations of production is assumed absent ($g_1(t) = g_2(t) = 0$). The operational characteristics for the workstations are summarized in the table below. During preventive maintenance and workstation failure no production occurs. Therefore, the moments for the moments for the state dependent Poisson processes in (13), (14), (12), and (16) are given by $\overline{Z}_{ki}^F = \overline{Z}_{ki}^R = \overline{Z}_{ki}^D = \overline{Z}_{ki}^M = 1$ with all covariances being 0. Assume that when the operational age of a workstation (16) is reset either due to a repair or preventive maintenance the operational age of the workstation is set to zero and that the aging process is based on the amount of time operational only. This implies that, $f(c_k(t), t)dt = 1$ and $H_{ki}^D(t) = H_{ki}^R(t) = \tau_{ki} - t - a_{ki}(\tau_{ki})$, where τ_{ki} is the time of the last reset (initial value is 0) and $a_{ki}(\tau_{ki})$ is viewed as a parameter that represents the age of the workstation at the last reset, which is zero for all $\tau_{ki} \neq 0$ and is specified in the table below for $\tau_{ki} = 0$.

Stage k	Workstation i	Production Capacity, M_{ki} (pieces/hour)	Operational Age, $a_{ki}(0)$ (hours)	Mean Times (hours)			
				T_{ki}^F	T_{ki}^R	T_{ki}^M	T_{ki}^D
1	1	65	13	170	6	105	2
1	2	70	60	180	8	90	2
1	3	75	0	220	6	110	2
2	1	135	6	190	8	95	2
2	2	115	50	170	7	85	2

This manufacturing system consists of 20 local and 2 global state variables for a state of dimension 22. Define the local state vectors as $*(t) = [*_{11}(t), *_{12}(t), *_{13}(t), *_{21}(t), *_{22}(t)]^\top$, for the available production capacity $* = r$, operational status $* = o$, preventive maintenance status $* = m$, and current operational age $* = a$. Define the global state vector for the surplus aggregate level as $\mathbf{s}(t) = [s_1(t), s_2(t)]^\top$. The total state and control vectors are given by

$$\mathbf{X}(t) = [\mathbf{x}(t),\ \mathbf{s}(t)]^\top = [\mathbf{o}(t),\ \mathbf{m}(t),\ \mathbf{r}(t),\ \mathbf{a}(t),\ \mathbf{s}(t)]^\top, \qquad \mathbf{u}(t) = [u_1(t),\ u_2(t)]^\top. \tag{23}$$

The cost functional used is (19) where the coefficient matrices are given by

$$S(t_f) = S_f = \begin{bmatrix} 0.011 & 0 \\ 0 & 0.014 \end{bmatrix}, \quad R(t) = \begin{bmatrix} 2.0 \times 10^8 & 0 \\ 0 & 3.0 \times 10^8 \end{bmatrix}, \quad \begin{array}{c} \overline{Q^+_{1,1}} = 0.8 \times 10^7 \\ \overline{Q^-_{1,2}} = 1.4 \times 10^7. \\ \overline{Q^+_{1,2}} = 1.0 \times 10^7 \end{array}$$

By comparing the coefficients of (1) with the state equations for the manufacturing system (13), (14), (12), (16), and (17) the deterministic coefficients are given by

$$A(t) = \begin{bmatrix} 0_{20\times10} & 0_{20\times5} & 0_{20\times5} & 0_{20\times2} \\ 0_{2\times10} & M_A(t) & 0_{2\times5} & B_A(t) \end{bmatrix}, \quad B(t) = \begin{bmatrix} 0_{20\times2} \\ \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \end{bmatrix}, \quad C(t) = \begin{bmatrix} \mathbf{0}_{15\times1} \\ \mathbf{1}_{5\times1} \\ 0 \\ d(t) \end{bmatrix}, \tag{24}$$

where

$$M_A(t) = \begin{bmatrix} M_{11}c_1(t) & M_{12}c_1(t) & M_{13}c_1(t) & -M_{21}c_2(t) & -M_{22}c_2(t) \\ 0 & 0 & 0 & M_{21}c_2(t) & M_{22}c_2(t) \end{bmatrix}, \quad B_A(t) = \begin{bmatrix} \hat{\tau}(t) & 0 \\ -\hat{\tau}(t) & \hat{\tau}(t) \end{bmatrix},$$

where $\hat{\tau}(t) = 1/\tau(t)$. Define the set $\gamma = \{11,\ 12,\ 13,\ 21,\ 22\}$ which is an index set for the stage and workstation, respectively. Define the diagonal matrix $\Delta_{*_{\gamma_i},c}$ such that the (i,i) component is given by $\delta_{*_{\gamma_i},c}$. The only *nonzero* stochastic process and corresponding coefficient matrix given by

$$d\mathbf{P}_3(\mathbf{X}(t),t) = \begin{bmatrix} d\mathbf{P}^R(\mathbf{X}(t),t) \\ d\mathbf{P}^F(\mathbf{X}(t),t) \\ d\mathbf{P}^D(\mathbf{X}(t),t) \\ d\mathbf{P}^M(\mathbf{X}(t),t) \end{bmatrix}, \qquad H_3(t) = \begin{bmatrix} \Delta_{o_{\gamma_i},0} & -\Delta_{r_{\gamma_i},1} & 0_{5\times5} & 0_{5\times5} \\ 0_{5\times5} & 0_{5\times5} & I_{5\times5} - \Delta_{m_{\gamma_i},1} & -\Delta_{r_{\gamma_i},1} \\ \Delta_{o_{\gamma_i},0} & -\Delta_{r_{\gamma_i},1} & I_{5\times5} - \Delta_{m_{\gamma_i},1} & -\Delta_{r_{\gamma_i},1} \\ -H_3^R(t) & 0_{5\times5} & -H_3^D(t) & 0_{5\times5} \\ 0_{2\times5} & 0_{2\times5} & 0_{2\times5} & 0_{2\times5} \end{bmatrix},$$

with $-H_3^R(t) = -H_3^D(t) = \operatorname{diag}\left[\tau_{\gamma_i} - t - a_{\gamma_i}(\tau_{\gamma_i})\right]_{5\times5}$, where $\operatorname{diag}[\mathbf{v}] = [v_i\delta_{i,j}]_{k\times k}$ is the diagonal matrix representation of the $k \times 1$ vector $\mathbf{v}$ and the state dependent Poisson processes for $* \in \{R,\ F,\ D,\ M\}$ are given by $d\mathbf{P}^*(\mathbf{X}(t),t) = \left[d\mathbf{P}^*_{\gamma_i}(\mathbf{X}(t),t)\right]_{5\times1}$. Using the above numerical values the algorithm presented in Section 3.3. can be used to determine the production rates for the manufacturing system. Consider the sample path trajectory described in the table below.

Event Time (hours)	Stage	Workstation	Event
0			start
15	2	2	failure
22			repair
30	1	1	start maintenance
32.5			end maintenance

The production rates, buffer levels, and percent relative error for the manufacturing system are given in Figure 3.

The production rates anticipate workstation repair, failure, and maintenance. At the final time of the planning horizon

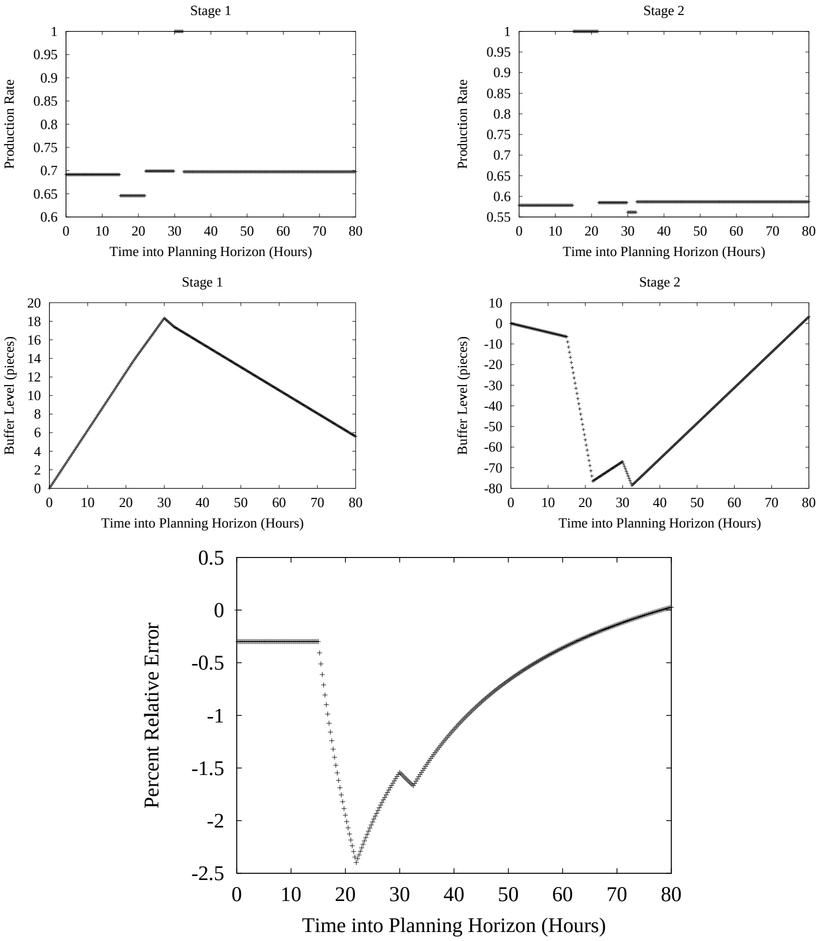

Figure 3: Production rates, buffer levels, and percent relative error for manufacturing system.

the percent relative error is 0.0276% with $b_1(t_f) = 5.582$ and $b_2(t_f) = 3.203$ pieces left in the buffers. The results presented here required approximately 11 wall clock seconds to complete on a Sun Ultra 5, with a memory demand of 2.136 megabytes.

5. Conclusions

Many manufacturing systems utilize assembly lines or flowshops to fabricate a consumable good. These manufacturing systems place buffers between the stages in order to accommodate excess production from one stage to the next. If a buffer becomes full, production on the prior stages is halted. The optimal production scheduling presented here ensures that the production on all stages compensates for workstation failure, repair, and maintenance without exceeding the buffer levels and therefore a continuous stream of pieces is produced. The optimal production scheduling uses the quasi-LQGP problem with state dependent Poisson processes (SDPP) as its paradigm. This formulation allows for a great deal of realism to be included for a large scale manufacturing system with minimal computational and memory demands. The discrete event trajectory following algorithm presented provides a better way to control the manufacturing system and could be easily applied to automated systems. The numerical example presented illustrates the functionality and power of this method.

References

[1] R. E. Bellman and S. E. Dreyfus, *Applied Dynamic Programming*, Princeton University Press, Princeton, NJ, 1962.

[2] 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.

[3] Fullman-Kinetics, *The Semiconductor Manufacturing Process*, available from http://www.fullman.com/semiconductors/_crystalgrowing.html, accessed 20 February 2001.

[4] I. I. Gihman and A. V. Skorohod, *Stochastic Differential Equations*, Springer-Verlag, New York, 1972.

[5] R. W. Hall, *Zero Inventories*, Dow Jones-Irwin, Homewood, Illinois, 1983.

[6] F. B. Hanson, “Techniques in Computational Stochastic Dynamic Programming,” *Digital and Control System Techniques and Applications*, edited by C. T. Leondes, Academic Press, New York, pp. 103-162, 1996.

[7] INFRASTRUCTURE, *Semiconductor Tour: The Chip-Making Process*, available from http://www.infras.com/Tutorial/sld001.htm, accessed 20 February 2001.

[8] International SEMATECH, *Semiconductor Mfg. Process*, available from http://www.sematech.org/public/news/mfgproc/mfgproc.htm, accessed 20 February 2001.

[9] J. Kimemia and S. B. Gershwin, “An Algorithm for the Computer Control of a Flexible Manufacturing System,” *IIE Trans.*, vol. 15, pp. 353-362, December 1983.

[10] S. P. Sethi, H. Zhang, and Q. Zhang, “Hierarchal Production Control in a Stochastic N-Machine Flowshop with Limited Buffers,” *Journal of Mathematical Analysis and Applications*, vol. 246, no. 1, pp. 25-57, June 2000.

[11] S. P. Sethi and Q. Zhang, *Hierarchal Decision Making in Stochastic Manufacturing Systems*, Birkhauser Boston, Cambridge, 1994.

[12] J. J. Westman and F. B. Hanson, “The LQGP Problem: A Manufacturing Application,” *Proceedings of the 1997 American Control Conference*, vol. 1, pp. 566-570, June 1997.

[13] J. J. Westman and F. B. Hanson, “State Dependent Jump Models in Optimal Control,” *Proceedings of 38th Conference on Decision and Control*, pp. 2378-2383, December 1999.

[14] J. J. Westman, F. B. Hanson, and E.K. Boukas, “Optimal Production Scheduling for Manufacturing Systems with Preventive Maintenance in an Uncertain Environment,” to appear in *Proceedings of 2001 American Control Conference*, 6 pages in ms., June 2001.