

# Operation of Manufacturing Systems with Work-in-process Inventory and Production Control

Yuan-Hung (Kevin) Ma, Yoram Koren (1)

NSF Engineering Research Center for Reconfigurable Manufacturing Systems, University of Michigan, Ann Arbor MI 48109-2125, USA

## Abstract

The operation of large manufacturing systems with buffers has two goals: to meet production target and minimize the work-in-process (WIP) inventory. This paper introduces a novel approach, based on optimal control theory, to achieve both goals simultaneously by on-line adjustment of the production rate of each machine. In this method the state variables are the buffer levels, the control variables are the machine production rates, and the output variable is the target production (the demand). The method is evaluated through simulations under various conditions, and compared with other methods in the literature. The results demonstrate that the proposed method can successfully produce low WIP inventory than other methods, while the required production demand is still fulfilled. It is also shown that the method is capable of providing feasible solutions for large manufacturing systems – a goal that is harder to achieve with the current known method.

**Keywords:** Manufacturing, Control, Optimization

## 1 INTRODUCTION

In manufacturing systems, random occurrences of machine failure cause disturbances to production. To reduce the impact of the disturbances, buffers are implemented between machines. However, the introduction of buffers raises the level of work-in-process (WIP) inventory, and, in turn, WIP increases the operating cost. Therefore, while the goal of the system operation is aimed at meeting the production requirements, it is also desired to simultaneously reduce the WIP.

A fuzzy-logic controller to minimize both WIP and production surplus was designed utilizing evolution strategies, but under the (unrealistic) assumption that machines do not fail [1]. In the research done by CIRP members on Production Planning and Control (PPC) systems [2, 3], the dynamics of the PPC systems were analyzed, and a controller for throughput and WIP adjustment was presented. However, this research considered the whole system as a one-machine-one-buffer system, and did not study the failure of each machine in large systems.

For a system consisting of unreliable machines and finite buffers, Kimemia and Gershwin proposed a feedback control policy of continuous production flow based on solving a stochastic optimal control problem [4, 5]. Given a constant desired demand rate, the production rate of each machine is computed in real time to meet a specific production surplus level, called the hedging point. Through an approximation of the discrete material flow with a continuous flow, two models of manufacturing systems were analyzed in [6], which have advantages over the conventional scheduling approaches in two aspects: complexity of the manufacturing system and vulnerability to schedule disruptions. These methods, however, are not applicable for large systems.

To minimize the buffer levels and meet the demand rate, a control policy called two-boundary control was developed, based on dynamic programming, for a two-machine-one-buffer tandem line where the buffer size is infinite [7]. Bai and Gershwin [8] approximated the hedging point control policy to a linear program. In addition, an algorithm for determining desired buffer sizes and hedging points was presented. Simulation results were displayed for a two-machine line and for a five-

machine line. Nevertheless, there is a need for a production flow control policy which is capable of being applied to large systems, as well as providing a trade-off between reducing WIP and keeping the production close to the target demand.

In this paper a novel control policy utilizing optimal control theory is introduced to obtain a low-level WIP while keeping the system throughput close to the required demand. The control policy generates off-line a library of optimal controllers which are selected according to the actual states of the machines (working and not working) and control in real time the production rates of the machines that are operational. The proposed policy is shown to be effective for large manufacturing systems. The new control policy is evaluated through simulations and compared with results presented in the literature [8].

## 2 APPROACH

### 2.1 Stochastic Manufacturing Systems and Control

Consider a serial manufacturing line with unreliable machines and finite buffers (Figure 1).

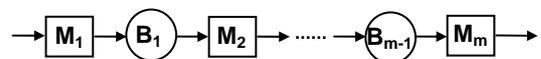


Figure 1: A serial manufacturing line.

The system is subject to random events: machine failures and repairs, which cause abrupt fluctuations in the production. The machine states are binary: machine operates = 1; machine fails = 0. The system has state variables ( $x_i$ ): the buffer levels (which are proportional to the WIP) and the output production surplus. The production rate of each machine is a control variable ( $u_i$ ). Thus, a serial system with  $m$  machines and  $m-1$  buffers has  $m$  states ( $m-1$  buffer states and one output state) and  $m$  control variables.

Since machine failures are random and subject to a probability distribution, the manufacturing system is considered to be a stochastic system. To minimize WIP and fulfill the production demand, the problem was usually formulated as a stochastic optimal control problem in the literature [4, 7]. However, the major

difficulty of this problem is the solution technique. Dynamic programming is generally employed to find the optimal solution. However, the solution to this problem becomes infeasible when the system consists of many machines and buffers, which, in practice, is the general situation in the manufacturing industry.

Our approach is to divide the stochastic optimal control problem into multiple deterministic optimal control sub-problems, based on the instantaneous machine states. The machine state combination randomly changes. When a change of the machine states is detected, controllers are selected from a library of optimal controllers that were pre-determined according to the machine states. The selected optimal controllers determine the production rates of the machines, based on the demand rate and on-line measurement of the buffer levels, as shown in Figure 2.

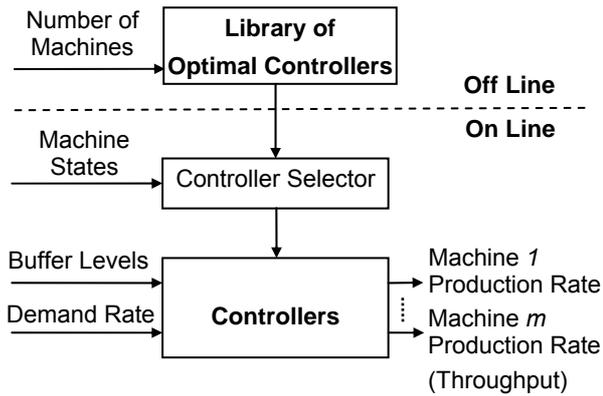


Figure 2: Schematics of the control policy.

For example, in a two-machine-one-buffer line, the machine states change among the four possible combinations (i.e., (1 1), (0 1), (1 0), (0 0)). For the machine state combination (0 0), in which both machines fail, there is no controller selected. For each of the other three combinations, an optimal controller is selected from the library. Each optimal controller determines the machines' production rates. Detailed description of the control policy for the general case of the  $m$ -machine line will be provided in the next section.

## 2.2 Control Policy

Consider a serial line with  $m$  machines and  $m-1$  buffers as depicted in Figure 3. The line produces only one part type; a constant production demand rate  $d$  is given.

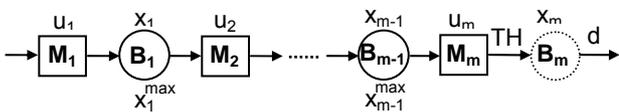


Figure 3: A serial line with  $m$  machines.

In our approach we are using a virtual buffer  $B_m$  at the end of the line. The level of this buffer is the cumulative difference between the actual production of machine  $m$  (which is the line throughput) and the target demand.

Let the level of buffer  $i$  be the continuous state variable  $x_i$  and the production rate of machine  $i$  the control variable  $u_i$ . The dynamics of the system are

$$\begin{aligned} \dot{x}_i &= u_i - u_{i+1} \quad (i = 1 \dots m-1); \\ \dot{x}_m &= u_m - d \end{aligned} \quad (1)$$

where  $x_m$  is the level of the virtual buffer  $B_m$ . This variable can be either positive or negative, and is zero indicating the demand is fulfilled, while the level of the physical buffer,  $x_i$ , is bounded by 0 and a given buffer size  $x_i^{\max}$ :

$$0 \leq x_i \leq x_i^{\max} \quad (i = 1 \dots m-1) \quad (2)$$

Machines have production capacity constraints. The production rate of machine  $i$  cannot exceed a given maximum value  $u_i^{\max}$  when the machine is operational

$$0 \leq u_i \leq u_i^{\max} \alpha_i \quad (i = 1 \dots m) \quad (3)$$

where  $\alpha_i$  is the machine state (0 or 1).

In the duration of each machine state combination, the part flow is obstructed by the failed machine(s), and thus the line can be considered as the union of sub-lines, where the machines within a sub-line are all operational. Sub-lines are separated by the failed machine(s) and have no stochastic disturbances. The controllers in the optimal controller library are designed for these sub-lines. The overall objective of the control policy is to control the levels of the system buffers while keeping the virtual buffer level close to zero. This objective is realized through the optimal controllers for the sub-lines.

For each sub-line, there is a corresponding optimal controller. The objective of the controller is to track given constant references of states (including  $x_m$ ) and inputs. Since the dynamics in the sub-lines are linear, linear quadratic optimal control problems are formulated. The objective function to be minimized is represented as

$$J = \frac{1}{2} \int_{t_0}^{t_f} [(x - x^r)^T Q (x - x^r) + (u - u^r)^T R (u - u^r)] dt \quad (4)$$

where  $\mathbf{x}$  and  $\mathbf{u}$  are the vector forms of the states and inputs, respectively, and  $\mathbf{x}^r$  and  $\mathbf{u}^r$  are the vector forms of the references of the states and inputs, respectively. The matrices  $\mathbf{Q}$  and  $\mathbf{R}$  are diagonal. The machine state combination starts at  $t_0$  and ends at  $t_f$ .

The sub-lines have different attributes when they contain the most upstream machine (machine 1) and the most downstream machine (machine  $m$ ). There is no part flow coming into the sub-lines that do not include machine 1, and there is no part flow leaving the sub-lines that do not include machine  $m$ . Therefore, the sub-lines and their corresponding controllers can be classified into the following four categories, based on their different attributes, and the references ( $\mathbf{x}^r$  and  $\mathbf{u}^r$ ) are determined accordingly. Each category may contain multiple controller types. Note that the sub-lines may occur in many machine state combinations, so the number of the controllers can be greatly reduced compared to the number of machine state combinations when the manufacturing line is large.

**Category I:** Working line with all the machines operational ( $\alpha_i = 1, i = 1 \dots m$ )

The reference of the virtual buffer level is zero to meet the required production. For the physical buffer, the desired level is chosen to be the half of its size. The half inventory is to protect the adjacent downstream machine from starvation, and the half space is to protect the adjacent upstream machine from blockage, since the adjacent machines at the both ends are operational.

$$x_i^r = \frac{x_i^{\max}}{2} \quad (i = 1 \dots m-1); \quad x_m^r = 0 \quad (5)$$

The desired machines' production rates are the demand rate to maintain the desired production flow in steady state.

$$u_i^r = d \quad (i = 1 \dots m) \quad (6)$$

Note that there is only one type of the controller in this category.

**Category II:** Working sub-lines that include machine  $m$  and exclude machine 1

$$(\alpha_i = 1, i = j \dots m, 2 \leq j \leq m)$$

Here machine  $j$  is the most upstream machine in the sub-line. If machine  $j-1$  is assumed to be repaired at the end of the duration, all the inventory levels (including buffer  $j-1$ ) are preferred to be consumed to zero before that time. The reference of the virtual buffer level is still zero for demand.

$$x_i^r = 0 \quad (i = j-1 \dots m) \quad (7)$$

The reference of the machine's production rate is zero to reduce the utilization of the machine if possible.

$$u_i^r = 0 \quad (i = j \dots m) \quad (8)$$

There are  $m-1$  types ( $j = 2 \dots m$ ) of controllers in this category.

**Category III:** Working sub-lines including machine 1 and excluding the machine  $m$

$$(\alpha_i = 1, i = 1 \dots k, 1 \leq k \leq m-1)$$

Here machine  $k$  is the most downstream machine in the sub-line. Since fulfilling the demand is not possible in this category, buffer  $k$  is desired to be full at the end of the duration such that the inventory can be used when machine  $k+1$  is repaired. For the other buffers, the references are chosen to be half full for the same reason in Category I.

$$x_i^r = \frac{x_i^{\max}}{2} \quad (i = 1 \dots k-1); \quad x_k^r = x_k^{\max} \quad (9)$$

The references of the machines' production rates are the same as in Category II.

$$u_i^r = 0 \quad (i = 1 \dots k) \quad (10)$$

There are  $m-1$  types ( $k = 1 \dots m-1$ ) of the controllers in this category.

**Category IV:** Working sub-lines that do not include both machine 1 and machine  $m$

$$(\alpha_i = 1, i = j \dots k, 2 \leq j \leq k \leq m-1)$$

Here the most upstream and downstream machines in the sub-line are machine  $j$  and  $k$ , respectively. Since the entire inventory level will not reduce in the duration, the references of the buffer levels are chosen such that the inventory in the upstream buffers is moved to the most downstream buffers as much as possible. The reference of the machine's production rate is the same as in Category II.

$$u_i^r = 0 \quad (i = j \dots k) \quad (11)$$

Note that there are  $(m-1)*(m-2)/2$  possible types of the controllers in this category.

Table 1 illustrates all the possible combinations of the machine states and their corresponding controller types in a four-machine line.

In summary, for a serial line with  $m$  machines, the total number of the controller types is the summation of the ones in the four categories.

Table 1: Controller types in a four-machine line.

Machine State Combination				Controller Type			
Machine 1 State	Machine 2 State	Machine 3 State	Machine 4 State	I	II	III	IV
1	1	1	1	1			
0	1	1	1		3		
1	0	1	1		2	1	
1	1	0	1		1	2	
1	1	1	0			3	
0	0	1	1		2		
0	1	0	1		1		1
0	1	1	0				3
1	0	0	1		1	1	
1	0	1	0			1	2
1	1	0	0			1	1
1	0	0	0			1	
0	1	0	0				1
0	0	1	0				2
0	0	0	1		1		
0	0	0	0				N/A

Figure 4 displays the relation between the number of machines versus the number of controller types. Note that when  $m = 20$ , there are 210 types of controllers, while the total possible machine state combinations are  $2^{20}$ , which is over one million. This shows the control policy is feasible for a large system.

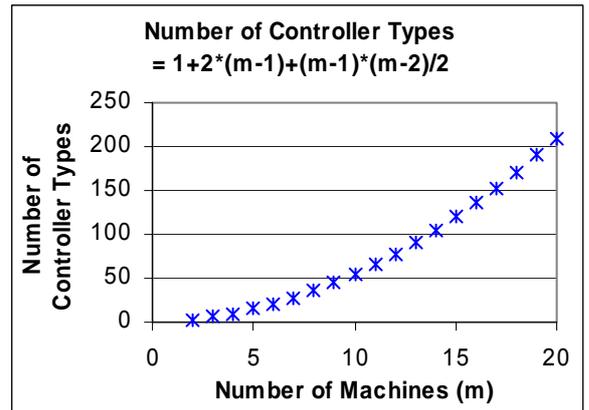


Figure 4: Number of controller types vs. number of machines in the line.

### 2.3 Linear Quadratic Optimal Control with Input and State Constraints

The optimal controllers for the sub-lines have been formulated in Section 2.2. However, because of the constraints on the inputs and the states in (2) and (3), the conventional linear quadratic regulation (LQR) techniques can not be applied to obtain the optimal solutions. Model predictive control (MPC) is one of the popular solution techniques for complex constrained control problems in the literature. In this paper, an improved approach of MPC, utilizing multi-parametric quadratic programming (mp-QP), is employed to solve our linear quadratic optimal control problems with constraints on inputs and states in discrete-time systems [9]. Unlike conventional MPC which requires on-line computation, the mp-QP approach moves the computation efforts off-line, and provides control inputs in an explicit state feedback form. This approach divides the state space into multiple partitions. For any partition  $i$ , two associated constant matrices  $F_i$  and  $G_i$  are

calculated. When the state  $\mathbf{x}$  is located in partition  $i$  at any time step, the feedback control law is given by

$$u = F_i x + G_i \quad (12)$$

Prediction horizon is a choice for the discrete-time finite horizon problems. When a greater prediction horizon is selected, the complexity of state partitions increases and the off-line computation takes more time to complete.

### 3 EXAMPLES AND SIMULATION RESULTS

In this section, the control policy will be implemented on a two-machine line and a five-machine line through simulations performed in a commercial package called WITNESS. These two lines were analyzed in [8]. The machines have exponential failure and repair distributions. Simulation results with 95% confidence intervals are obtained through 20 replications, each with a different random number seed. Each replication runs 10,000 time-units with a warm-up period of 3,000 time-units. The values of simulation parameters, as well as units, are selected from [8] in order to compare the results in the same conditions. Note that the matrix  $R$  in the objective function is chosen to be small compared to  $Q$ , so reducing machine usage is not considered for the current study. The prediction horizon is chosen as 1 to reduce computation time. Numerical results indicate that they are not sensitive to the choice of prediction horizon.

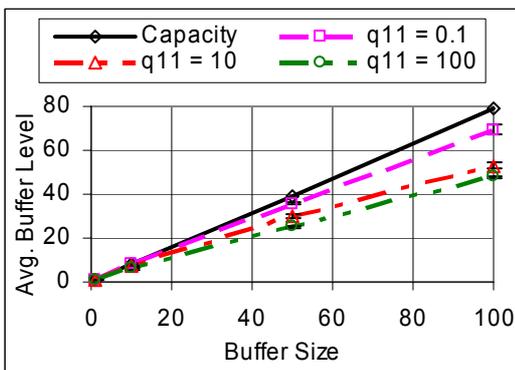
#### 3.1 Two-machine-one-buffer Line

Machines' parameters are shown in Table 2. Since the machine's MTF is greater than MTTR, the sub-line in Category I occurs most frequently. Thus, varying the weights in Category I has the most influence on performance.

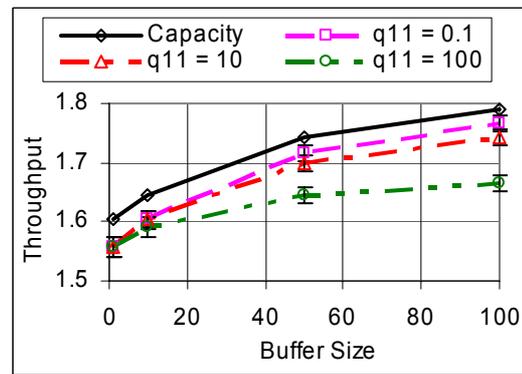
Table 2: Machines' parameters used in simulation.

	Mean Time To Failure (MTTF)	Mean Time To Repair (MTTR)	Maximum Production Rate
<b>Machine 1</b>	100 (day)	20 (day)	2.5 (lots / day)
<b>Machine 2</b>	10 (day)	2 (day)	2.2 (lots / day)

Figure 5 shows the graphs of the average buffer level and the throughput versus the buffer size, while varying the weight of the physical buffer in Category I ( $q_{11}$ ). In "Capacity," all the machines produce at their maximum production rates. It produces an upper bound on throughput and also results in the largest average buffer level. The results are generated from a software package called PAMS [10], a production analysis tool created in our center. Note that here the throughputs in Capacity are used as the demand rates for the control policy.



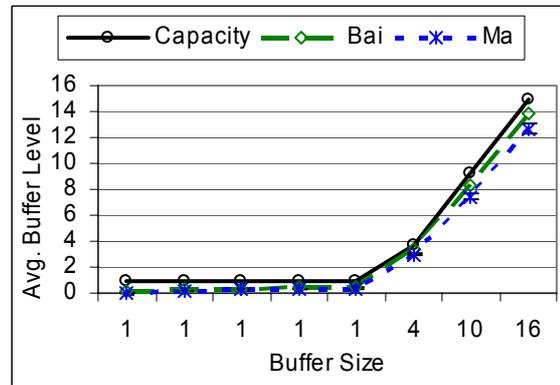
(a)



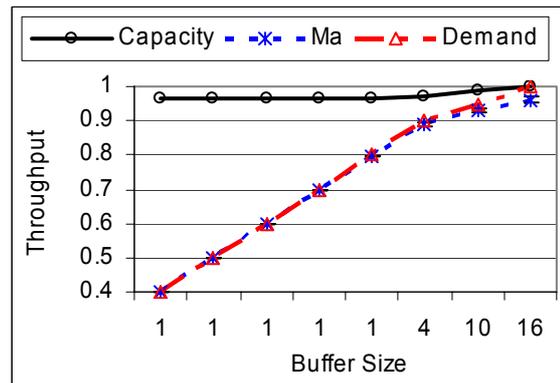
(b)

Figure 5: Effect of various weights.

When the weight increases, the average buffer level reduces closer to the half of the buffer size, but the throughput also suffers. In general, with a choice of the moderate weight, the average buffer level can reduce up to 30%, compared with Capacity, while the loss of the throughput is less than 3%.



(a)



(b)

Figure 6: Comparison with [8] for a two-machine line.

Figure 6 compares the simulation results with the ones in Bai and Gershwin [8], with various demand rates. The demand rates are listed in Table 3, so are the buffer sizes, which are determined by the demand rates in [8].

Table 3: Demand rates and buffer sizes from [8].

<b>Demand Rate</b>	0.4	0.5	0.6	0.7	0.8	0.9	0.95	1
<b>Buffer Size</b>	1	1	1	1	1	4	10	16

Machines' parameters are the same in Table 2 except machine 2's maximum production rate is 1.25 lots/day here. The weights in Category I and III are adjusted to obtain best results. The information of the throughput is not available in Bai's simulation results. Note that the demand rates in Table 3 are not used for calculations in Capacity; the results change because of the different buffer sizes. When the demand rate is close to the capacity of the line, the throughput suffers a loss about 4% of the demand rate, and the average buffer level has a 8% reduction compared with Bai's result. When the demand rate is not close to the capacity, the difference between the throughput and the demand rate is less than 1%, while the reduction of the average buffer level is up to 47%. In general, the proposed control policy always produces lower average buffer levels than Bai's approach.

### 3.2 Five-machine Line

In the simulation of a five machine line, all the machines are identical: MTTF = 10, MTTR = 2 and  $u^{max} = 2$ , chosen from the case 5 in [8]. Figure 7 compares the results when the demand rate is 1 and all the buffer sizes are all 1.

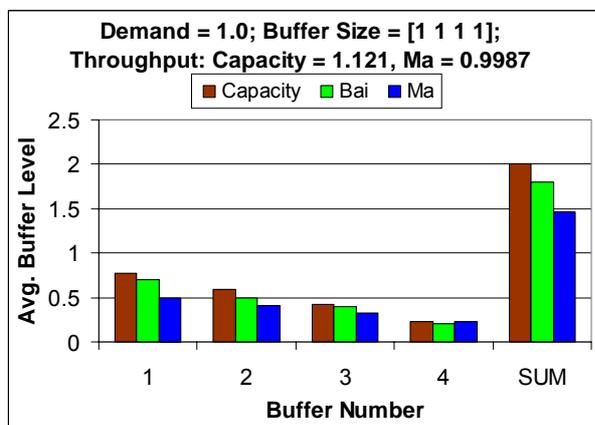


Figure 7: Comparison with [8] for a five- machine line.

Compared with Bai's result, the control policy reduces 19% of the total average buffer levels, and the loss of the throughput is only 0.13%. In this example, the demand rate (= 1.0) here is not very close to the upper bound in Capacity (= 1.121). Note that varying weights does not change the simulation results distinctly in the five-machine line example.

## 4 SUMMARY

In this paper, a novel control policy for manufacturing system operation has been presented. It is based on modelling an  $m$ -machine line as an  $m$ -order state-space system and applying optimal control theory to adjust the WIP while keeping the production demand. For a serial line with random machine failures, the policy divides the stochastic system into multiple deterministic sub-lines, each operating optimally for the duration in which the machine state combination does not change. The simulation results demonstrate that the proposed policy successfully generates low WIP while the demand is still fulfilled. The policy shows better performance than the one presented by Bai and Gershwin in [8], and is capable of being easily applied to large manufacturing systems.

## 5 ACKNOWLEDGMENTS

The authors are pleased to acknowledge the support by the National Science Foundation Engineering Research Center for Reconfigurable Manufacturing Systems under Grant No. NSF-Sub-EEC-9529125, and would like to thank Pascal Grieder for his assistance on the mp-QP programs.

## 6 REFERENCES

- [1] Porter, B., Zadeh, N. N., 1997, Evolutionary design of fuzzy-logic controllers for manufacturing systems, *Annals of the CIRP*, 46/1:425-428.
- [2] Duffie, N., Falu, I., 2002, Control-theoretic analysis of a closed-loop PPC system, *Annals of the CIRP*, 51/1:379-382.
- [3] Ratering, A. M., Duffie, N. A., 2003, Design and analysis of a closed-loop single-workstation PPC system, *Annals of the CIRP*, 52/1:355-358.
- [4] Kimemia, J. G., Gershwin, S. B., 1983, An algorithm for the computer control of production in flexible manufacturing systems, *IIE Transactions*, 15/4:353-362.
- [5] Gershwin, S. B., 1994, *Manufacturing systems engineering*, Prentice Hall, Englewood Cliffs, NJ.
- [6] Brandimarte, P., Sharifnia, A., Von Turkovich, B. F., 1996, Continuous flow models of manufacturing systems: a review, *Annals of the CIRP*, 45/1:441-444.
- [7] Van Ryzin, G., Lou, S. X. C., Gershwin, S. B., 1993, Production control for a tandem two-machine system, *IIE Transactions*, 25/5:5-20.
- [8] Bai, S. X., Gershwin, S. B., 1995, Scheduling manufacturing systems with work-in-process inventory control: single-part-type systems, *IIE Transactions*, 27/5:599-617.
- [9] Bemporad, A., Morari, M., Dua, V., Pistikopoulos, E. N., 2002, The explicit linear quadratic regulator for constrained systems, *Automatica*, 38/1:3-20.
- [10] Koren, Y., Hu S. J., Weber, T. W., 1998, Impact of Manufacturing system configuration on performance, *Annals of the CIRP*, 47/1:369-372.