Paper 446d

A Model Predictive Control Approach for Managing Semiconductor Manufacturing Supply Chains under Uncertainty

Wenlin Wang[†], Junyhung Ryu[†], Daniel E. Rivera[†], Karl G. Kempf[††] and Kirk D. Smith[†††]

[†]Department of Chemical and Materials Engineering
Control Systems Engineering Laboratory,
Arizona State University, Tempe, Arizona 85287-6006

[††]Decision Technologies,
[†††] Components Automation Systems,
Intel Corporation,
5000 W. Chandler Blvd., Chandler, AZ, USA 85226

(480) 965-9476
email: daniel.rivera@asu.edu

Prepared for presentation at the Annual AIChE Meeting, San Francisco, CA
November 16–21, 2003
Session: [446] Enterprise-Wide Optimization I

November, 2003

UNPUBLISHED

A Model Predictive Control Approach for Managing Semiconductor Manufacturing Supply Chains under Uncertainty

Wenlin Wang[1], Junyhung Ryu[1], D. E. Rivera[1,*], K. G. Kempf[2] and Kirk D. Smith[3]

[1]Department of Chemical and Materials Engineering
Control Systems Engineering Laboratory,
Arizona State University, Tempe, Arizona 85287-6006

[2]Decision Technologies,
[3]Components Automation Systems,
Intel Corporation,
5000 W. Chandler Blvd., Chandler, AZ, USA 85226

November 14, 2003

*To whom all correspondence should be addressed. phone: (480) 965-9476 fax: (480) 965-0037; e-mail: daniel.rivera@asu.edu

Abstract

A two level architecture using Model Predictive Control as a tactical decision module is presented for supply chain management in the semiconductor manufacturing industries. The strategic and inventory planning steps in the outer loop provide the inventory targets and capacity limits by solving an optimization problem that maximizes profits. These decisions are usually weekly or monthly. The MPC-based tactical decision module takes advantage of these targets, capacity limits and demand forecasts to make daily decisions on starts at the various manufacturing nodes. Fluid analogies are used to model the supply chain dynamics in semiconductor manufacturing which facilitates the application of Model Predictive Control. Several benchmark problems which contain distinguishing features of semiconductor manufacturing, such as nonlinear and stochastic throughput times and customer demands, are examined. All of these problems involve two types of manufacturing nodes, Fab/Test1 and Assembly/Test2, and three types of inventories, Assembly-Die Inventory, Semi-Finished Goods Inventory and Finished Components Inventory. Both supply side uncertainty, including varying throughput times and yields, and demand side uncertainty are addressed. The nonlinear relationship between the throughput time and load is considered in each case. The effects of judiciously picking tuning and model parameters to achieve performance, robustness and improved customer satisfaction are studied by comparing the variance in starts, inventories, and load as well as the percentage of unfilled orders. Increasing move suppression and choosing the nominal throughput times at average values usually gives better performance with lower variance and less backlog. The flexibility provided by the choice of tuning and model parameters in MPC to achieve more effective supply chain management in semiconductor manufacturing is demonstrated in each case study.

1 Introduction

Successful supply chain management is an important consideration in today's manufacturing industries because of the vital role it plays in distributing resources and generating profits. More effective operation of supply chains for manufactured goods is worth billions of dollars to our national economy (ASCET, 2003). Reports of savings achieved by best-in-class companies as a result of effective supply chain management amount to 5 - 6% of sales (SimulationDynamics, 2003). For example, in a diversified American corporation with $25 billion in annual sales (such as Intel), eliminating supply chain inefficiencies across the enterprise represents over one billion dollars in potential profits. While a generic supply chain stretches from suppliers through manufacturing to customers, one of the most promising areas for improvement is in the generation and execution of the plans for the factories, warehouses, and transport links that form the backbone of all such supply chains.

There are many approaches to building strategic plans to effectively operate complex supply chains (Simchi-Levi *et al.*, 2000). The most sophisticated utilize some form of mathematical optimization, often linear programming (LP) (Hopp and Spearman, 1996; Chopra and Meindl, 2001). The resulting plan specifies material release from warehouses into factories and transport links over multiple future weeks and months. While such strategic planning systems are very useful, they are deterministic and the techniques to include the unavoidable stochasticity of supply and demand directly into LP-based optimizers are very difficult at best. However, recent progress in multi-echelon inventory theory can act as adjuncts to the LP optimizers used for strategic plan construction (Kapuscinski and Tayur, 1999; Graves and Willems, 2000). Given target service levels, estimates of future supply and demand uncertainty, and historical forecast bias and error, these inventory algorithms compute safety stock positions and targets to be used as input to the LP optimizers. This safety stock is intended to buffer the expected variability in both supply and demand while executing the LP-generated multi-period plan.

In common supply chain practice, such planning-with-safety-stock hybrids can be utilized on a weekly basis (Kempf, 2003*b*). Unfortunately, the stochastic processes driving the supply and demand variability operate minute to minute, hour by hour, day after day. In order to reduce the safety stock level with lower supply chain costs and improved levels of delivery performance generating higher revenues, we present an approach of developing decision policies based on control-theoretic concepts as shown in Figure 1. In this architecture, the standard Strategic Planning and Inventory Planning modules form the outer loop of the controller providing long term weekly and monthly goals. These goals are passed to the Tactical Execution module that we present as the inner loop of controller. This paper focuses on a formulation based on Model Predictive Control to implement the inner loop controller.

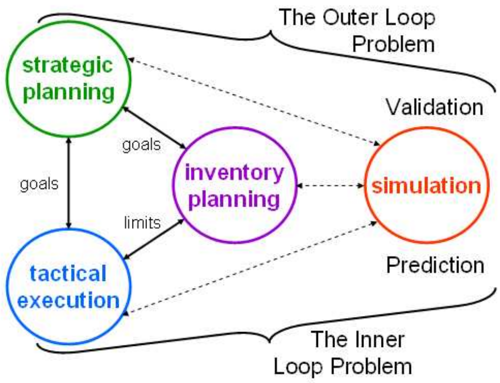


Figure 1: Proposed Approach to Supply Chain Management

Model Predictive Control has been used as a decision-making tool for handling complex integrated production planning problems within a stochastic environment (Tzafestas *et al.*, 1997). Recent work using Model Predictive Control has shown it as attractive method for inventory control in supply chain(Braun *et al.*, to appear, 2003*a*; Braun *et al.*, 2003*b*; Braun *et al.*, 2002). These papers show the effectiveness of a partially decentralized MPC structure under model mismatch and demand forecasted bias in a deterministic environment. A centralized MPC strategy is successfully used in semiconductor manufacturing supply chain management to track inventory targets and satisfy the customer demands as much as possible (Wang *et al.*, 2003). An MPC strategy has also been used to find the optimal decision variables to maximize profit in supply chains with multi-product, multiechelon distribution networks with multiproduct batch plants (Perea *et al.*, 2003). The centralized MPC coordinator is shown to have better performance due to the knowledge of all possible conflicts and restrictions in system. In all these papers, the appeal of MPC for dynamic inventory management in supply chains can be summarized as follows: as an optimizer, MPC can minimize or maximize an objective function that represents a suitable measure for supply chain performance. As a controller, MPC can be tuned to achieve stability, robustness, and performance in the presence of plant/model mismatch, failure, disturbance and uncertainty which affect the system.

This paper presents a two level approach in semiconductor manufacturing supply chain management with the focus on the development of the inner loop controller using a centralized MPC strategy. Section 2 presents a process flow representation for semiconductor manufacturing supply chain dynamics. In Section 3, Model Predictive Control formulation and application to supply chain are discussed. In Sections 4-6, three typical problems involving different features of semiconductor manufacturing are studied. Simulation results showing proof of concept for the MPC approach are discussed in each individual problem. This paper concludes with a discussion of the flexibility and advantages of using MPC in supply chain management, and a description of future research.

2 Process Control in Semiconductor Manufacturing Supply Chain Management

Process control systems are widely used in the chemical industries to adjust flows to maintain level and product compositions at desired values. Material flows in supply chains can be modeled using fluid analogy, and as a result one can expect that decision policies based on process control principles can have a large impact on supply chain management. In particular, advanced process control techniques such as Model Predictive Control (MPC) offer a decision framework that can provide improved performance (García *et al.*, 1989). As control-oriented frameworks, these schemes have the advantage that they can be tuned to provide acceptable performance in the presence of significant supply and demand variability and forecasting error as well as constraints on inventory levels, production and shipping capacity. Figure 2 shows the basics of semiconductor manufacturing process. Wafers are first processed in Fab/Test1 and die is created on them. Die is then tested and put into different packages with different speeds. This is done in Assembly/Test2. The semi-finished goods will be configured, packed and shipped to customers in the finishing manufacturing stage. A fluid representation of a three-node semiconductor manufacturing supply chain (consisting of one Fab/Test1, one Assembly/Test2, and one finish node) and its corresponding inventory locations is shown in Figure 3. In a very general sense, the manufacturing nodes are represented as "pipes", while the inventory locations are represented as "tanks"; material in these pipes and tanks correspond to Work-in-Progress (WIP) and inventory, respectively. In the example per Figure 3 the problem is simplified in that die-only inventories are considered (that is, there are enough packages for every die).

The application of advanced process control methods to supply chains associated with discrete-parts manufacture (such as semiconductor manufacturing) contains features in common with chemical processing as well as novel and distinct features. The distinguishing features of these discrete manufacturing problems include:

- The throughput times associated with the manufacturing process are typically longer than those found in the chemical process industry. For example, a semiconductor product may not be ready for sale for as long as three months from the time of initial material release.

- The supply process is highly stochastic as a consequence of the features of the discrete manufacturing process, the large number of process steps and associated processing machinery, and the re-entrant nature of parts of the process. Stochasticity is present in throughput time, overall yields of functional products, and resulting product performance distributions in every stage of manufacturing.

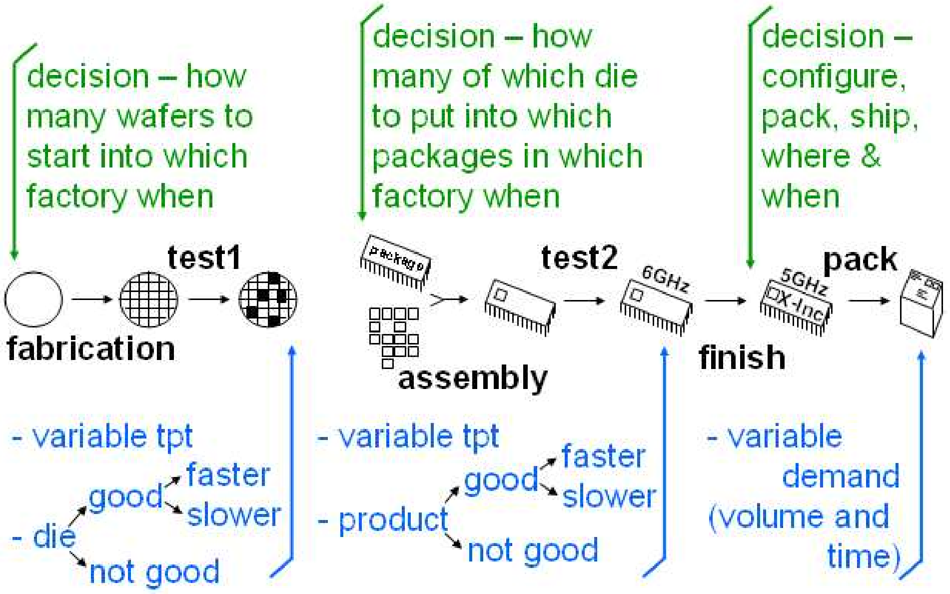


Figure 2: Sequence of steps associated with semiconductor manufacturing

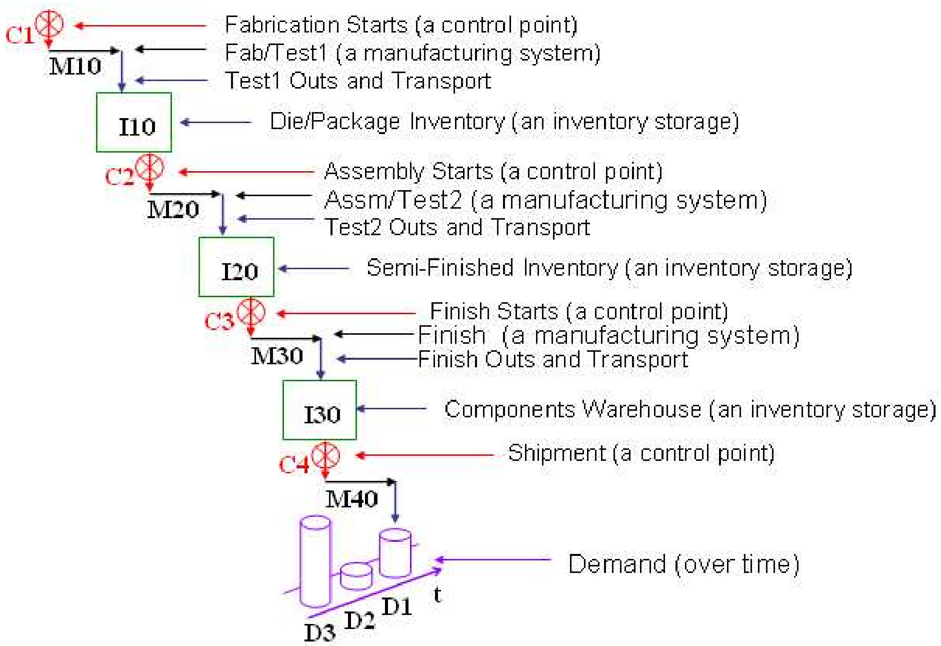


Figure 3: Fluid representation of a three node semiconductor mfg. supply chain

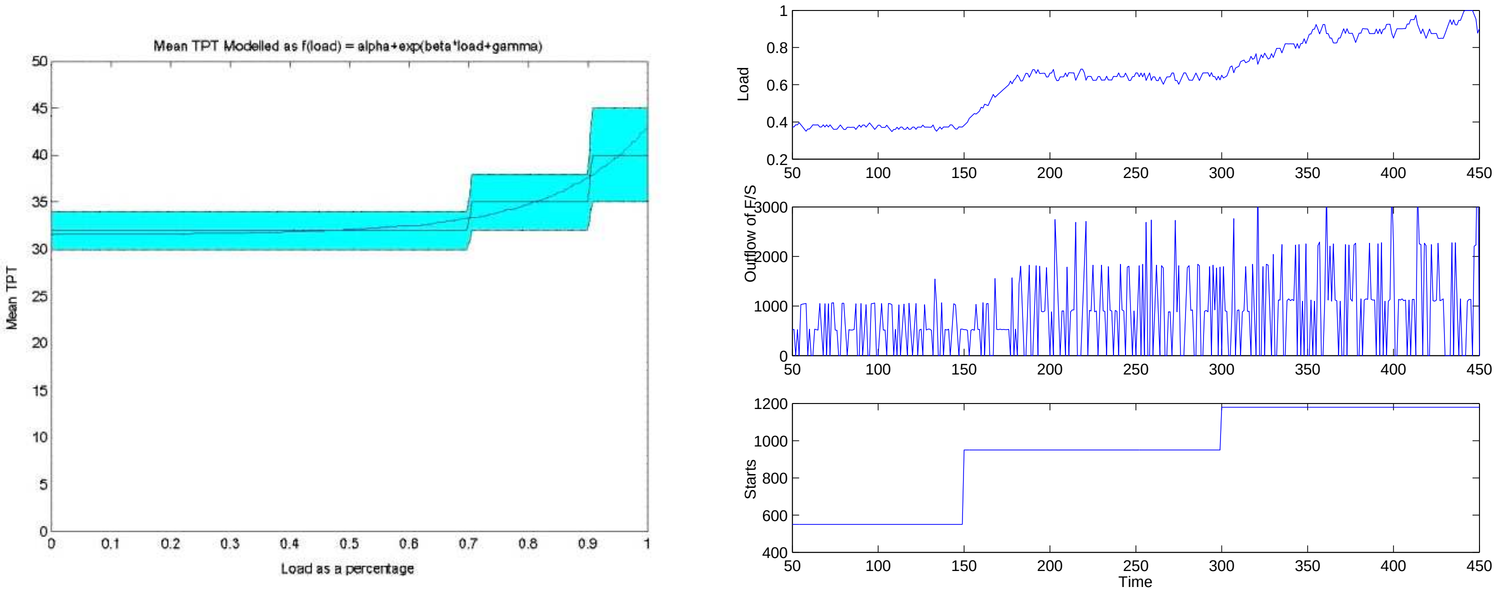
- The throughput times vary nonlinearly with load on the manufacturing nodes; likewise, the magnitude of the stochasticity increases with load. This can be seen in Figure 4.
- Demand forecasting plays a critical role in making judicious decisions, particularly since demand can change dramatically within the time horizon denoted by supply lead times. Demand forecasts must be treated as anticipatory feedforward signals in the controller in order to achieve a high performance solution. While feedforward control is commonplace in the process industries, disturbance anticipation is not a typical feature of chemical process control systems.
- The demand process is highly stochastic as a consequence of the large number of semiconductor products that are produced, the large number of applications for which these products are used, and the large number of competitors and customers in the marketplace.

Some of the challenges can be demonstrated in Figure 4 and 5. Figure 5 shows a typical example of the stochasticity and uncertainty in demand. The actual demand can be different from and much more noisy than the forecast of future customer demand. Figure 4 shows the input/output relation for one Fab/Test1 node. In our simulation model, the throughput time of this node is nonlinearly dependent on the load or Work-In-Progress (WIP). It varies uniformly between three different load ranges (30 to 32 days at 0 to 70% load, 32 to 38 days at 70 to 90% load, and 35 to 45 days at 90 to 100 % load). For other factories, the throughput time is only a uniformly distributed number varying from 5 to 7 days for Assembly/Test2 node and 1 to 3 days for Finish node. Yield rates also vary uniformly for each manufacturing node.

Following the example in Figure 3, Intel has developed a series of fluid representations for problems that contain the distinguishing features of semiconductor manufacturing listed previously (Kempf, 2003*a*). Besides the basic problem shown in Figure 3, these extensions include the assumption of perishable demand, management of both die and package inventories (Figure 6(a)), stochastic splits in manufacturing and configuring of products with varying performance characteristics (Figure 6 (b)), multi-factory problems with cross-shipments (Figure 6 (c)), multi-product problems with shared capacity (Figure 6 (d)), and problems involving multiple products with independent capacity but correlated demand. All the features of stochasticity and nonlinearity in the manufacturing processes require a sophisticated control strategy to improve the decision making and performance of the supply chain.

3 Model Predictive Control

Model Predictive Control is an optimization-based control scheme. Its formulation integrates optimal control, stochastic control, control of processes with dead time and multivariable control. It is perhaps the most general way of posing the process control problem in the time domain (García *et al.*, 1989). One of the advantages of using MPC is that it can easily handle constraints on both



(a) (b)

Figure 4: (a) Nonlinear and stochastic relationship between throughput time and load a in manufacturing node; (b) Manufacturing node response to changes in starts.

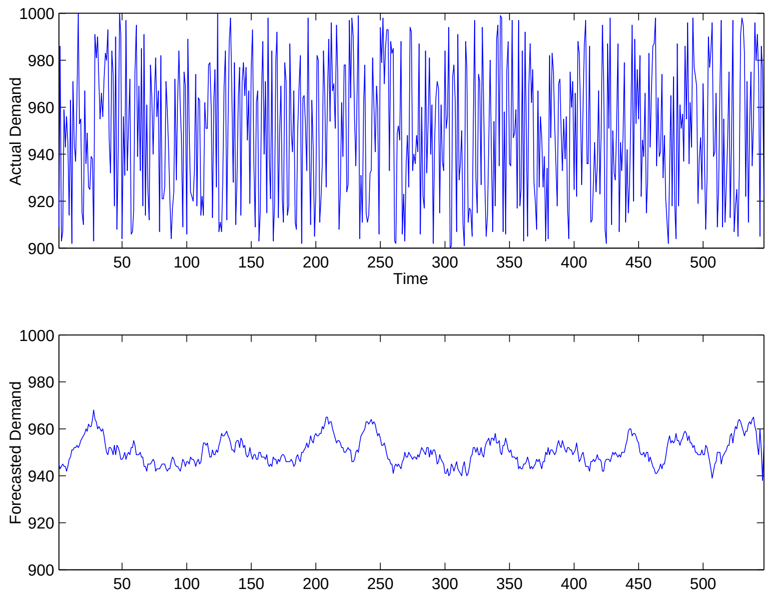
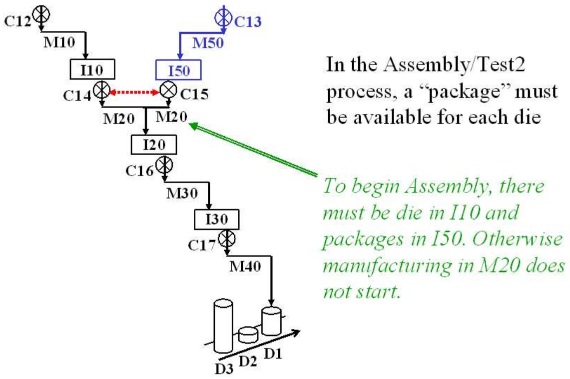
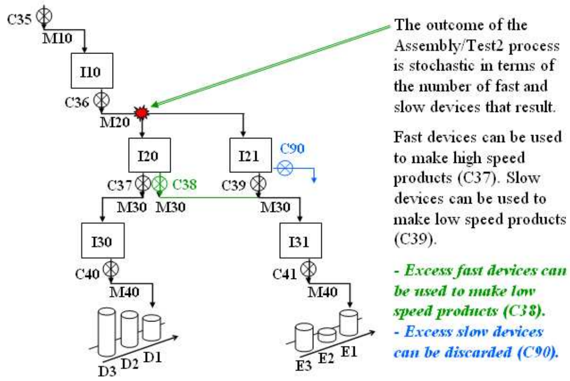


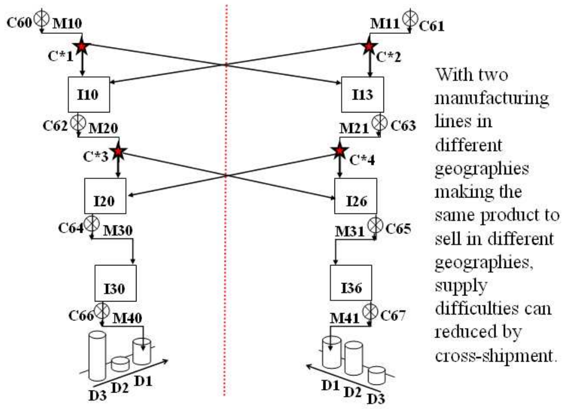
Figure 5: An example of forecasted (bottom) vs actual customer demand (top)



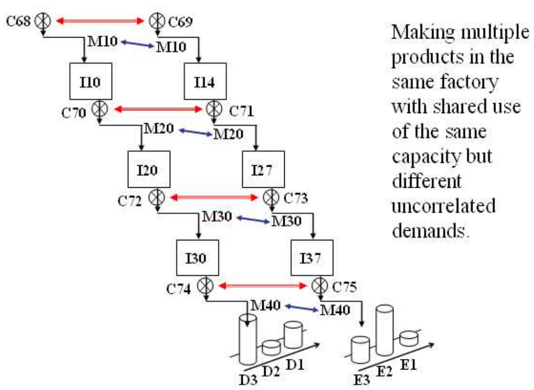
(a) the Die-Packaging Problem



(b) The test2 stochastic split problem



(c) Multi-factory problem with cross-shipments



(d) multi-product problem with shared capacity

Figure 6: Flow representation problem set

manipulated and control variables. The MPC controllers considered in this paper are based on the linear state-space model:

$$\begin{aligned} x(k+1) &= Ax(k) + B_u u(k) + B_v v(k) + B_d d(k) \\ y(k) &= Cx(k) + D_v v(k) + D_d d(k) \end{aligned} \tag{1}$$

where $x(k)$ represents the states of the system, $u(k)$ is the manipulated variables (MV) or command inputs, $v(k)$ is the vector of measured disturbances (MD), $d(k)$ is the vector of unmeasured disturbances (UD), and $y(k)$ is the output vector, which is composed of measured outputs (MY) $y_m(k)$ and unmeasured outputs (UY) $y_u(k)$. In our formulation, u corresponds to the start of each manufacturing node, y contains the inventory levels and WIP, v is the forecasted demand and d is the unforecasted demand.

The basic principle of MPC can be shown Figure 7. At each time instant t, the controller considers

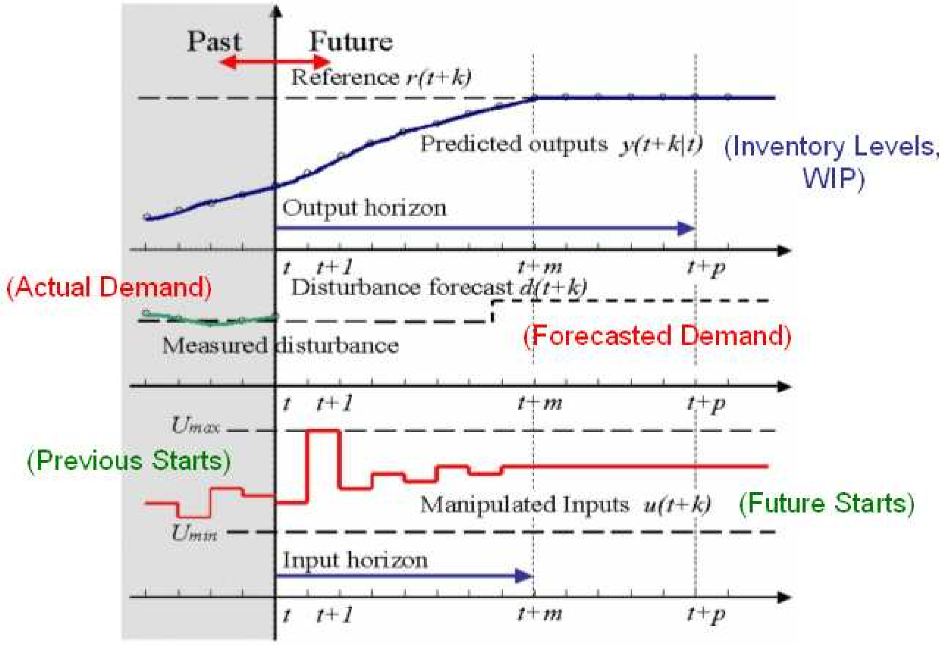


Figure 7: Moving horizon representation for Model Predictive Control

the previous information on inventory levels, actual customer demands, starts and future information on inventory targets, forecasted customer demand to calculate a sequence of future starts by solving the following optimization problem.

$$\min_{\Delta u(k|k) \ldots \Delta u(k+m-1|k)} J \tag{2}$$

where the individual terms of J correspond to:

$$\begin{aligned} J \;=\; & \overbrace{\sum_{\ell=1}^{p} Q_e(\ell)(\hat{y}(k+\ell|k) - r(k+\ell))^2}^{\text{Keep Inventories at Inventory Planning Setpoints}} + \overbrace{\sum_{\ell=1}^{m} Q_{\Delta u}(\ell)(\Delta u(k+\ell-1|k))^2}^{\text{Penalize Changes in Starts}} \\ & + \overbrace{\sum_{\ell=1}^{m} Q_u(\ell)(u(k+\ell-1|k) - u_{target}(k+\ell-1|k))^2}^{\text{Maintain Starts at Strategic Planning Targets}} + \epsilon^T \rho \epsilon \end{aligned} \tag{3}$$

s.t.

$$u_i^{min} \leq u(k+i \mid k) \leq u_i^{max} \quad (4)$$

$$\triangle u_i^{min} \leq \triangle u(k+i \mid k) \leq \triangle u_i^{max} \quad (5)$$

$$i = 0, \ldots, p-1$$

$$y_i^{min} - \epsilon \leq \hat{y}(k+j+1 \mid k) \leq y_i^{max} + \epsilon \quad (6)$$

$$\triangle u(k+j \mid k) = 0 \quad (7)$$

$$j = m, \ldots, p$$

$$\epsilon \geq 0 \quad (8)$$

$$r = [x_k], k = 1, \ldots, t_{final}$$

Here p is the prediction horizon and m is the control horizon. t_{final} is the end time for a whole simulation. r is the references of the outputs; x_k is the anticipated reference value at time k. $Q_u, Q_{\triangle u}, Q_e$ are penalty weights on the control signal, move size and control error, respectively. Manipulated variables have high, low and move size constraints. Output constraints are soft which is done by introducing a slack variable vector ϵ to prevent infeasible solutions. There is one slack variable coefficient for each output over the time horizon. The weight ρ is arbitrarily set by the user. This problem can be solved by standard quadratic program algorithms. Clearly, the objectives of solving this optimization problem include:

1. A *setpoint tracking* term that is intended to maintain inventory levels at user-specified targets over time. These targets need not be constant and can change over the prediction horizon p.

2. A *move suppression* term that penalizes changes (also referred to as *moves*) in the starts. While penalizing changes in the starts may be desirable to manufacturing node personnel, this term serves a more important control-theoretic purpose, as the primary means for achieving robustness in the controller in the face of uncertainty (García *et al.*, 1989),

3. An *input target* term that is meant to maintain the starts close to daily (or per shift) target values, based on the weekly targets calculated at the strategic level.

Following the receding horizon principle, MPC applies the first element of the calculated control action to the system. After new measurements are available, a new optimization problem is solved. The use of future setpoint and disturbance changes in MPC is referred to as anticipative action when the value of these signals is known in advance. Making use of anticipation in the controller is a significant contributor to improved performance. With proper tuning, an MPC approach relying on linear models with fixed parameters can achieve good performance in spite of the very nonlinear and stochastic characteristics in manufacturing process. This will be shown in the following sections.

4 The basic problem with backlog

As shown in Figure 3, this basic problem represents the critical dynamics of a semiconductor manufacturing network. It contains one Fab/Test1 (F/T1) node, one Assembly/Test2 (A/T2) node, one Finish/Pack (F/P) node, one Assembly-Die Inventory (ADI), one Semi-Finished Goods Inventory (SFGI) and one components warehouse inventory. There are two kinds of inputs to the MPC controller. One is the starts of manufacturing nodes which can be manipulated by controller to achieve different performance. The other one is the customer demand which can be measured but can not be manipulated by controller. The customer demand comes in at the end of the chain. In order to achieve better performance, the anticipation or forecast of future customer demand is used by MPC to make a prediction over the horizon. The outputs of this problem are classified into two categories in terms of *controlled variables* and *associated variables*. Controlled variables need to track some targets generated from the outer loop. In our formulation, these are the inventory levels for ADI, SFGI and CW. The MPC controller will generate starts to keep these three inventory levels as close to the targets as possible. The discrete time model for inventory based on material balances is as follows:

$$I_{10}(k+1) = I_{10}(k) + Y_1 C_1(k-\theta_1) - C_2(k) \tag{9}$$

$$I_{20}(k+1) = I_{20}(k) + Y_2 C_2(k-\theta_2) - C_3(k) \tag{10}$$

$$I_{30}(k+1) = I_{30}(k) + Y_3 C_3(k-\theta_3) - C_4(k) \tag{11}$$

where I_{10}, I_{20} and I_{30} are the ADI, SFGI and components warehouse inventories. C_1, C_2 and C_3 are the starts for F/T1, A/T2 and F/P respectively. C_4 is the customer demand which we treat as measured disturbance with anticipation. θ_1, θ_2 and θ_3 are the throughput time for F/T1, A/T2 and F/P, while Y_1, Y_2 and Y_3 are yields for F/T1, A/T2 and F/P respectively.

Associated variables do not have specific targets, but they have to stay within some range. For instance, the Work-In-Progress (WIP) can not exceed the capacity or go negative. The controller has to insure that these variables stay within high and low limits of the entire horizon. For each manufacturing node, the WIP can be described as

$$WIP_{10}(k+1) = WIP_{10}(k) + C_1(k) - C_1(k-\theta_1) \tag{12}$$

$$WIP_{20}(k+1) = WIP_{20}(k) + C_2(k) - C_2(k-\theta_2) \tag{13}$$

$$WIP_{30}(k+1) = WIP_{30}(k) + C_3(k) - C_3(k-\theta_3) \tag{14}$$

One of the advantages of MPC is constraint handling. It is easy to implement constraints on both inputs and outputs of the controller. In this problem, the constraints exist on inventory levels and capacity of each factory as high and low limits

$$I_{10}^{min} \leq \quad I_{10}(k+i|k) \quad \leq I_{10}^{max} \tag{15}$$

$$I_{20}^{min} \leq \quad I_{20}(k+i|k) \quad \leq I_{20}^{max} \tag{16}$$

$$I_{30}^{min} \leq \quad I_{30}(k+i|k) \quad \leq I_{30}^{max} \tag{17}$$

$$0 \leq \quad WIP_{10}(k+i|k) \quad \leq WIP_{10}^{max} \tag{18}$$

$$0 \leq \quad WIP_{20}(k+i|k) \quad \leq WIP_{20}^{max} \tag{19}$$

$$0 \leq \quad WIP_{30}(k+i|k) \quad \leq WIP_{30}^{max} \tag{20}$$

$$i = \quad 1, 2, \ldots, p$$

It is also necessary to impose constraints on the starts and the move size of starts

$$0 \leq \quad C_8(k+j|k) \quad \leq C_8^{max} \tag{21}$$

$$0 \leq \quad C_9(k+j|k) \quad \leq C_9^{max} \tag{22}$$

$$0 \leq \quad C_{10}(k+j|k) \quad \leq C_{10}^{max} \tag{23}$$

$$\triangle C_8^{min} \leq \quad \triangle C_8(k+j|k) \quad \leq \triangle C_8^{max} \tag{24}$$

$$\triangle C_9^{min} \leq \quad \triangle C_9(k+j|k) \quad \leq \triangle C_9^{max} \tag{25}$$

$$\triangle C_{10}^{min} \leq \quad \triangle C_{10}(k+j|k) \quad \leq \triangle C_{10}^{max} \tag{26}$$

$$j = \quad 1, 2, \ldots, m$$

where p is prediction horizon and m is control horizon.

In this formulation, the nominal model for the controller is deterministic in nature. The stochasticity is present in the simulation model on TPT, yield and demand as discussed in previous section. Only $M10$ has the nonlinear relationship between load and throughput time as shown in Figure 4. Parameters that define in the simulation model are shown in Table 1. Additional data on inventory targets and constraints are shown in Table 2.

Stochasticity also occurs on demand side as well. The demand generated in this problem is a random sequence with an average of 950 items per day and a variance of 150 items per day. The forecast has the same average but has reduced variance compared to the actual demand. This is shown in Figure 5.

In this simulation, the choice of prediction horizon p is 70 days and control horizon m is 60 days. When there is no stochasticity in simulation model, i.e. TPT, yield and demand are all deterministic and the same as those in the controller model, the controller can manipulate the starts to meet the inventory targets and the customer demand perfectly in the presence of sufficient capacity. This is because we have perfect model to describe the dynamics of the system and future events.

				M10	M20	M30	M40
load=0% TO 70%Max	TPT	Min	days	30.0	5.0	1.0	1.0
	TPT	Ave	days	32.0	6.0	2.0	1.0
	TPT	Max	days	34.0	7.0	3.0	1.0
		Distribution		Unif	Unif	Unif	
load=70% TO 90%Max	TPT	Min	days	32.0	5.0	1.0	1.0
	TPT	Ave	days	35.0	6.0	2.0	1.0
	TPT	Max	days	38.0	7.0	3.0	1.0
		Distribution		Unif	Unif	Unif	
load=90% TO 100%Max	TPT	Min	days	35.0	5.0	1.0	1.0
	TPT	Ave	days	40.0	6.0	2.0	1.0
	TPT	Max	days	45.0	7.0	3.0	1.0
		Distribution		Unif	Unif	Unif	
Applied at Output	Yield	Min	%	93.0	98.0	98.5	100.0
	Yield	Ave	%	95.0	98.5	99.0	100.0
	Yield	Max	%	97.0	99.0	99.5	100.0
		Distribution		Unif	Unif	Unif	
Capacity	Load	Max	Items	45000	7500	2500	2500
	Load	Initial	Items	33285	5706	1902	951

Table 1: Manufacturing nodes data for basic problem with backlog: TPT- throughput time; Unif- uniform distribution

.

We consider the stochasticity introduced on both the supply side (TPT and yields) and the demand side (the forecasted and actual demand). This system is highly stochastic and each TPT is assumed to have a uniform distribution. In practice, the nominal model may be different from the actual model in process due to the variation in process or inaccurate knowledge. So first we use the low end of the distribution of TPT in nominal model, i.e. 30 days for F/T1, 5 days for A/T2 and 1 day for F/P, while in the process, the TPT has property as described in Table 1. With the same p, m and zero move suppression on all starts, the simulation is shown in Figure 8. Because no move suppression is used in the manipulated variables, the starts are very aggressive. The inventory levels are influenced too, but there is no offset. $M20$ and $M30$ run out of the capacity frequently and the load variation is very large. These aggressive responses take $I30$ to zero on occasion and the warehouse sometimes does not have any products when customers make orders, which generates backlog.

Instead of using low end of the range, we can use the average TPT in the nominal controller model. Still using zero move suppression on all the starts, we can improve the performance by reducing the

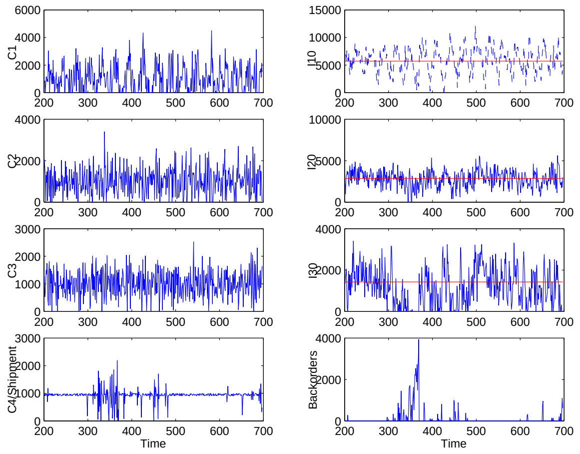
			I10	I20	I30
Inventory	UCL	Item	12000	6000	3000
	TAR	Item	5706	2853	1427
	LCL	Item	1000	1000	1000
	Level	Max	20000	10000	10000
	Level	Initial	3000	3000	1500

Table 2: Inventory setting in basic problem with backlog:UCL-Upper Control Limits; TAR-Target, LCL-Lower Control Limits

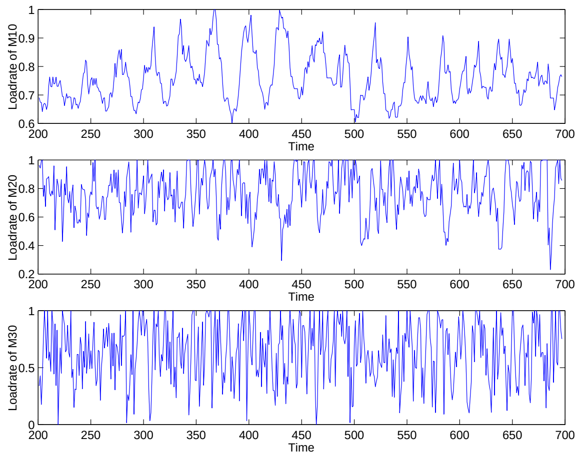
variance of the controlled and manipulated variables. Backorders are reduced as shown in Table 3. However these responses are still very aggressive.

In order to smooth the responses and decrease the variance, the move suppression penalty is set to 10 for C_1, C_2, and C_3. If TPT in the nominal model is at the low end of the distribution, compared with previous two sets of parameters, this biased TPT with zero suppression gives better performance with zero backorders and smaller variance on inventory levels and starts as shown in Table 3. If we use the average of TPT in the nominal model and the same increased move suppression, the results are shown in Figure 9. The customer demand is met perfectly without any backorder, and the starts are smoother than before. The inventory levels still have oscillations. But these are mainly caused by the stochastic manufacturing process as described in Figure 4 (b). The inventory levels for $I10$,$I20$ and $I30$ are all greater than zero all over the entire horizon. Especially in $I10$, we have more than enough to stock. This simulation results suggest that with this set of parameters, the safety stock level for $I10$ can be decreased.

Table 3 shows the average and variance of inventory levels, WIP and starts. From the comparison, one can find that a larger move suppression gives better performance, i.e. zero backorder and smaller variance. The larger move suppression helps the system to achieve robustness and reduce the variance. The mean of each inventory level is also closer to the target with larger move suppression. For instance, when there is no bias on TPT, the variance of ADI using large move suppression is reduced by 28.7% compared with the one using zero move suppression. In biased TPT case, it is reduced by 41% compared with the one using zero move suppression. With average TPT, the error between the mean of ADI and its target is 281 using zero move suppression, and 61.8 using larger move suppression. The choice of model parameters also influences the system performance. Using the average of TPT in nominal model clearly creates less backorders and smaller variance. However, the choice of move suppression has more influence on system performance than the choice of parameters in the nominal model. When larger move suppression is used, no backorders are observed in both biased and non-biased TPT cases. Although there is inherent stochasticity in

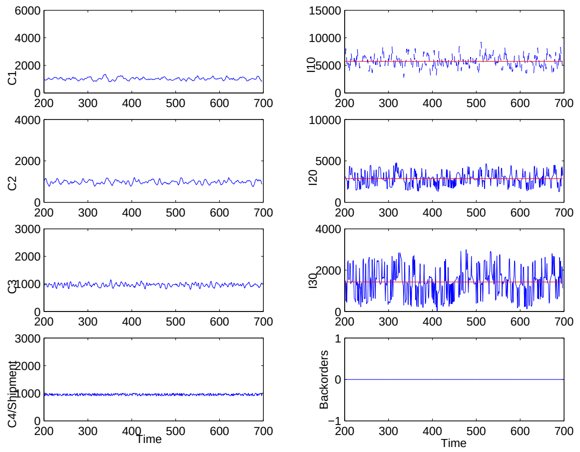


(a) Starts and Inventories

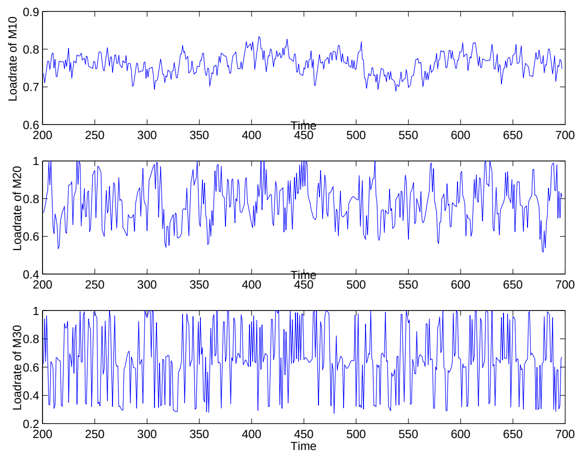


(b) Factory Load

Figure 8: Basic problem with backlog: Move Suppression [0 0 0], nominal controller TPT at the low end



(a) Starts and Inventories



(b) Factory Load

Figure 9: Basic problem with backlog: Move suppression [10 10 10], nominal controller TPT at average

Mean	MS 0/TPT Low End	MS 0/TPT Average	MS 10/TPT Low End	Ms 10/TPT Average
I10	5997.3	5987.0	5752.3	5767.8
I20	2859.2	2832.0	2881.1	2883.5
I30	1144.5	1140.6	1442.2	1425.9
C1	1024.5	1022.2	1024.8	1024.3
C2	1023.5	1018.6	1024.8	976.3
C3	1021.7	1016.6	959.98	963.2
M10 Load	0.7468	0.7523	0.7579	0.7583
M20 Load	0.7692	0.7723	0.7708	0.7721
M30 Load	0.6501	0.6212	0.6466	0.6421
Variance	MS 0/TPT Low End	MS 0/TPT Average	MS 10/TPT Low End	MS 10/TPT Average
I10	2.6901e+6	1.8575e+6	1.5853e+6	1.3239e+6
I20	9.9192e+6	8.8792e+5	6.1217e+5	5.5917e+5
I30	6.0722e+5	6.0790e+5	4.7038e+5	4.6520e+5
C1	7.2749e+5	5.9264e+5	1.7487e+4	7.5520e+3
C2	4.2602e+5	3.9084e+5	6.9425e+3	7.8223e+3
C3	2.3854e+5	2.5764e+5	3.8832e+3	4.5739e+3
M10 Load	0.0029	0.0021	7.2842e-4	8.8787e-4
M20 Load	0.0252	0.0242	0.0109	0.0105
M30 Load	0.0665	0.0697	0.0459	0.0486
Order Unfilled	2.66%	1.46%	0%	0%

Table 3: Comparison of the effect of using different tuning parameters in the basic problem with backlog

the manufacturing process, proper tuning strategies and choice of model parameters can help to improve system performance by reducing variance and backlogs.

5 The Die-Packaging Problem

In this problem, we expand the basic problem to include the packaging process. The fluid analogy for this problem is shown in Figure 6 (a). Compared with the basic problem with backlog, here in the Assembly/Test2 process, ($M20$), one package must be combined with one die. In order to begin assembly, there must be enough dies in I10 and enough packages in I50. Otherwise, manufacturing in Assembly/Test2 does not start. The two control points $C14$ and $C15$ are manipulated by the same signal to make the same amount of dies and packages to start the manufacturing process if there are enough inventory for both dies and packages. This approach avoids the involvement of solving an integer problem in the decision policy. The nominal model for controlled variables (inventory levels) can be described via following equations.

$$I_{10}(k+1) = I_{10}(k) + Y_1 C_{12}(k-\theta_1) - C_{15}(k) \quad (27)$$
$$I_{50}(k+1) = I_{50}(k) + Y_5 C_{13}(k-\theta_5) - C_{15}(k) \quad (28)$$
$$I_{20}(k+1) = I_{20}(k) + Y_2 C_{15}(k-\theta_2) - C_{16}(k) \quad (29)$$
$$I_{30}(k+1) = I_{30}(k) + Y_3 C_{16}(k-\theta_3) - C_{17}(k) \quad (30)$$

For all manufacturing nodes, the associated variables (WIP) can be described as

$$WIP_{10}(k+1) = WIP_{10}(k) + C_{12}(k) - C_{12}(k-\theta_1) \quad (31)$$
$$WIP_{50}(k+1) = WIP_{50}(k) + C_{13}(k) - C_{13}(k-\theta_5) \quad (32)$$
$$WIP_{20}(k+1) = WIP_{20}(k) + 2C_{15}(k) - 2C_{15}(k-\theta_2) \quad (33)$$
$$WIP_{30}(k+1) = WIP_{30}(k) + C_{16}(k) - C_{16}(k-\theta_3) \quad (34)$$

We can set high and low limits on inventory levels and capacity as shown below,

$$I_{10}^{min} \leq I_{10}(k+i|k) \leq I_{10}^{max} \quad (35)$$
$$I_{50}^{min} \leq I_{50}(k+i|k) \leq I_{50}^{max} \quad (36)$$
$$I_{20}^{min} \leq I_{20}(k+i|k) \leq I_{20}^{max} \quad (37)$$
$$I_{30}^{min} \leq I_{30}(k+i|k) \leq I_{30}^{max} \quad (38)$$
$$WIP_{10}^{min} \leq WIP_{10}(k+i|k) \leq WIP_{10}^{max} \quad (39)$$
$$WIP_{50}^{min} \leq WIP_{50}(k+i|k) \leq WIP_{50}^{max} \quad (40)$$
$$WIP_{20}^{min} \leq WIP_{20}(k+i|k) \leq WIP_{20}^{max} \quad (41)$$
$$WIP_{30}^{min} \leq WIP_{30}(k+i|k) \leq WIP_{30}^{max} \quad (42)$$
$$i = 1, 2, \ldots, p$$

As well as impose constraints on starts and change of starts.

$$
\begin{aligned}
0 \leq \quad & C_{12}(k+j|k) & \leq C_{12}^{max} \quad (43)\\
0 \leq \quad & C_{15}(k+j|k) & \leq C_{15}^{max} \quad (44)\\
0 \leq \quad & C_{16}(k+j|k) & \leq C_{16}^{max} \quad (45)\\
0 \leq \quad & C_{17}(k+j|k) & \leq C_{17}^{max} \quad (46)\\
\triangle C_{12}^{min} \leq \quad & \triangle C_{12}(k+j|k) & \leq \triangle C_{12}^{max} \quad (47)\\
\triangle C_{15}^{min} \leq \quad & \triangle C_{15}(k+j|k) & \leq \triangle C_{15}^{max} \quad (48)\\
\triangle C_{16}^{min} \leq \quad & \triangle C_{16}(k+j|k) & \leq \triangle C_{16}^{max} \quad (49)\\
\triangle C_{17}^{min} \leq \quad & \triangle C_{17}(k+j|k) & \leq \triangle C_{17}^{max} \quad (50)\\
j = \quad & 1, 2, \ldots, m &
\end{aligned}
$$

In order to insure there are enough dies and packages available for processing in $M20$, we have to enforce that inventories of $I10$ and $I50$ be no less than the amount required by C_{15}.

$$
\begin{aligned}
C_{15}(k+j|k) \quad & \leq I_{10}(k-1+j|k) \quad (51)\\
C_{15}(k+j|k) \quad & \leq I_{50}(k-1+j|k) \quad (52)\\
j = \quad & 1, 2, \ldots, m
\end{aligned}
$$

The uncertainty occurs on TPT and yields of $M10, M50, M20$, and $M30$. The nonlinear relationship between the load and TPT shown in Figure 4 still holds for $M10$ and $M50$. The prediction horizon p is 70 days and control horizon m is 60 days. The data for all manufacturing nodes implemented in this simulation is shown in Table 4. The inventory targets and constraints are in Table 5.

As before, we try to use the low end of TPT distribution in the nominal model and zero move suppression. As shown in Figure 10, the responses are very aggressive and some backorders are created. If we change the TPT in the nominal model to the average value, the variance of most variables is reduced while the backorder is increased as shown in Table 6.

In order to smooth the responses and achieve robustness, a larger move suppression is tried with the biased TPT. As shown in Table 6, the variance of each variable is reduced and no backorder is created. If TPT in the nominal controller model is changed to the average value and still use larger move suppression, the responses are shown in Figure 11. The variance for each variable is further reduced without any backlog. There are no offsets between the inventory levels and the targets. The loads of factories are within the capacity limits.

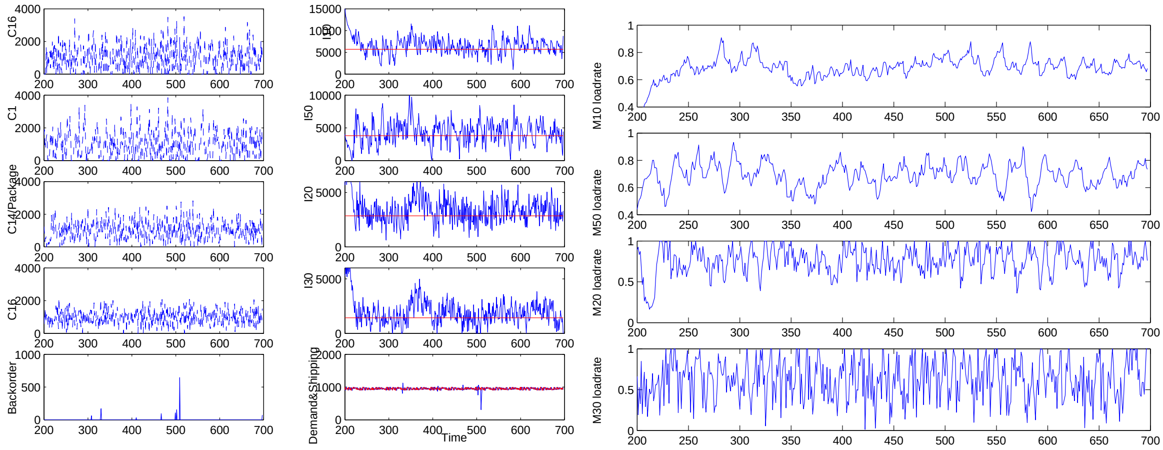
From Table 6 we can compare the effects of different move suppression and the choice of TPT. It is clear to see the same variance change as in the basic problem with backlog. In both controller TPT

				M10	M50	M20	M30	M40
load=0% TO 70%Max	TPT	Min	days	30.0	20.0	5.0	1.0	1.0
	TPT	Ave	days	32.0	21	6.0	2.0	1.0
	TPT	Max	days	34.0	23	7.0	3.0	1.0
		Distribution		Unif	Unif	Unif	Unif	
load=70% TO 90%Max	TPT	Min	days	32.0	21.0	5.0	1.0	1.0
	TPT	Ave	days	35.0	23.0	6.0	2.0	1.0
	TPT	Max	days	38.0	25.0	7.0	3.0	1.0
		Distribution		Unif	Unif	Unif	Unif	
load=90% TO 100%Max	TPT	Min	days	35.0	23.0	5.0	1.0	1.0
	TPT	Ave	days	40.0	27.0	6.0	2.0	1.0
	TPT	Max	days	45.0	30.0	7.0	3.0	1.0
		Distribution		Unif	Unif	Unif	Unif	
Applied at Output	Yield	Min	%	93.0	93.0	98.0	98.5	100.0
	Yield	Ave	%	95.0	95.0	98.5	99.0	100.0
	Yield	Max	%	97.0	97.0	99.0	99.5	100.0
		Distribution		Unif	Unif	Unif	Unif	
Capacity	Load	Max	Items	45000	30000	7500	2500	2500
	Load	Initial	Items	33285	22190	5706	1902	951

Table 4: Manufacturing nodes data for the die-packaging problem: TPT-throughput time; Unif-uniform distribution

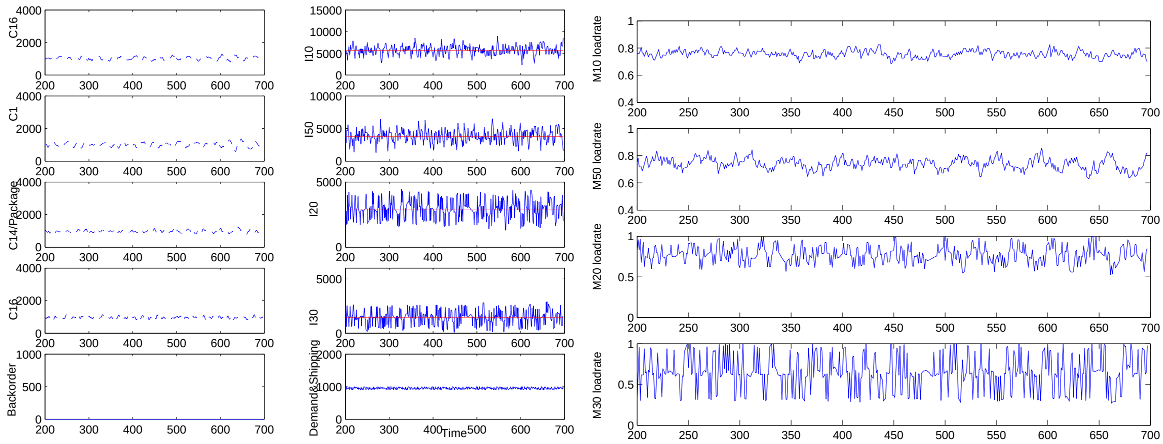
			I10	I50	I20	I30
Inventory	UCL	Item	12000	8000	6000	3000
	TAR	Item	5706	3804	2853	1427
	LCL	Item	1000	650	1000	1000
	Level	Max	20000	13500	10000	10000
	Level	Initial	3000	2000	3000	1500

Table 5: Inventory setting in die-packaging problem:UCL-Upper Control Limits; TAR-Target ;LCL-Lower Control Limits



(a) Starts and Inventories (b) Factory Load

Figure 10: Die-Packaging problem: Move Suppression [0 0 0 0], nominal controller TPT at the low end



(a) Starts and Inventories (b) Factory Load

Figure 11: Die-Packaging problem: Move Suppression [10 10 10 10],nominal controller TPT at average

cases, larger move suppression gives better performance which means closer agreement to targets, less backorder and low variance on all manipulated and controlled variables. For instance, in both cases the unfilled order percentage is 0 with move suppression 10 for each manipulated variable. For ADI, the errors between the mean and target values are 841.8 and 48.5 with no and large move suppression, respectively, when there is no bias on TPT. The ADI variances are reduced 52% and 54.1% for the cases with no bias on TPT and bias on TPT respectively. With the same large move suppression, the variance on ADI with no bias on TPT is 33.5% smaller than that with biased TPT. Again, although both the move suppression and the choice of model parameters can affect the performance of the system, the move suppression has more influence based on the results of this study.

6 The Assembly/Test2 Stochastic Split Problem

In this problem, a stochastic split is considered in Assembly/Test2 process as shown in Figure 6(b). Items coming into $M20$ through $C36$ will be split into two bins, one is made up of fast speed components, while the other one has slow speed devices. The number of items in each bin is determined by a split factor in A/T2 which is stochastic and can have different average and variance values. Besides meeting the fast device demand D, fast devices in $I20$ can also be used to make slow devices to meet the demand E. Devices in $I21$ can only be used to meet the customer demand E. In other words, if there are more than enough products available to meet customer demand D while no enough to meet E, through $C38$ some fast devices will be transferred from $I20$ to $I31$ to meet the demand E. The opposite direction is not allowed. Excess devices in $I21$ will be discarded through $C90$ if the inventory level reaches a maximum.

The nominal controller model used for inventories in this problem is as follows:

$$I_{10}(k+1) = I_{10}(k) + Y_1 C_{35}(k-\theta_1) - C_{36}(k) \tag{53}$$

$$I_{20}(k+1) = I_{20}(k) + Y_2 C_{36}(k-\theta_2)\cdot\alpha - C_{37}(k) - C_{38}(k) \tag{54}$$

$$I_{21}(k+1) = I_{21}(k) + Y_2 C_{36}(k-\theta_2)\cdot(1-\alpha) - C_{39}(k) \tag{55}$$

$$I_{30}(k+1) = I_{30}(k) + Y_3 C_{37}(k-\theta_3) - C_{40}(k) \tag{56}$$

$$I_{31}(k+1) = I_{31}(k) + Y_3 C_{38}(k-\theta_3) + Y_3 C_{39}(k-\theta_3) - C_{41}(k) \tag{57}$$

Here α is the split factor which is assumed in simulation as a random number with a uniform

Mean	MS 0/TPT Low End	MS 0/TPT Average	MS 10/TPT Low End	MS 10/TPT Average
I10	6575.6	6547.8	5775.9	5754.5
I50	4264.3	4343.4	3796.8	3865.5
I20	3394.1	3423.8	2897.3	2927.9
I30	1930.9	1712.7	1427.9	1462.5
C12	976.75	968.33	1018.0	1022.1
C14	1000.8	1030.0	973.49	973.66
C1	979.69	971.42	1026.7	1022.1
C16	982.83	1017.8	962.83	963.25
M10 Load	0.6871	0.6927	0.7539	0.7574
M50 Load	0.6927	0.6956	0.7361	0.7408
M20 Load	0.7569	0.7829	0.7667	0.7710
M30 Load	0.6256	0.6498	0.6351	0.6420
Variance	MS 0/TPT Low End	MS 0/TPT Average	MS 10/TPT Low End	MS 10/TPT Average
I10	3.6169e+6	2.3044e+6	1.6603e+6	1.1037e+66
I50	2.9037e+6	1.7652e+6	1.0076e+6	8.7395e+5
I20	1.4690e+6	8.4040e+5	5.8403e+5	5.2230e+5
I30	1.2804e+6	7.6449e+5	4.6855e+5	4.2726e+5
C12	6.6034e+5	5.9987e+5	2.1230e+4	8.0775e+3
C14	3.4938e+5	4.0087e+5	6.6943e+3	6.1335e+3
C1	6.5619e+5	4.9654e+5	3.2235e+4	1.5995e+4
C16	2.1009e+5	2.5695e+5	4.1272e+3	3.8709e+3
M10 Load	0.0060	0.0028	6.4937e-4	6.3675e-4
M50 Load	0.0078	0.0049	0.0020	0.0017
M20 Load	0.0278	0.0224	0.0108	0.0101
M30 Load	0.0609	0.0672	0.0477	0.0434
Order Unfilled	0.22%	0.43%	0%	0%

Table 6: Comparison of the effect of using different tuning parameters in the die-packaging problem

distribution. The WIPs of manufacturing nodes can be described in the following equations.

$$
\begin{aligned}
WIP_{10}(k+1) &= WIP_{10}(k) + C_{35}(k) - C_{35}(k-\theta_1) && (58)\\
WIP_{20}(k+1) &= WIP_{20}(k) + C_{36}(k) - C_{36}(k-\theta_2) && (59)\\
WIP_{30}^{high}(k+1) &= WIP_{30}^{high}(k) + C_{37}(k) - C_{37}(k-\theta_3) && (60)\\
WIP_{30}^{low}(k+1) &= WIP_{30}^{low}(k) + C_{38}(k) + C_{39}(k) - C_{38}(k-\theta_3) - C_{39}(k-\theta_3) && (61)
\end{aligned}
$$

The following constraints are used to keep high and low limits on inventory levels and capacities.

$$
\begin{aligned}
I_{10}^{min} \leq \; & I_{10}(k+i|k) \; \leq I_{10}^{max} && (62)\\
I_{20}^{min} \leq \; & I_{20}(k+i|k) \; \leq I_{20}^{max} && (63)\\
I_{21}^{min} \leq \; & I_{21}(k+i|k) \; \leq I_{21}^{max} && (64)\\
I_{30}^{min} \leq \; & I_{30}(k+i|k) \; \leq I_{30}^{max} && (65)\\
I_{31}^{min} \leq \; & I_{31}(k+i|k) \; \leq I_{31}^{max} && (66)\\
WIP_{10}^{min} \leq \; & WIP_{10}(k+i|k) \; \leq WIP_{10}^{max} && (67)\\
WIP_{20}^{min} \leq \; & WIP_{20}(k+i|k) \; \leq WIP_{20}^{max} && (68)\\
WIP_{21}^{min} \leq \; & WIP_{21}(k+i|k) \; \leq WIP_{21}^{max} && (69)\\
WIP_{30}^{min} \leq \; & WIP_{30}(k+i|k) \; \leq WIP_{30}^{max} && (70)\\
i = \; & 1, 2, \ldots, p
\end{aligned}
$$

The starts and the change of starts can also have high and low limits.

$$
\begin{aligned}
0 \leq \; & C_{35}(k+j|k) \; \leq C_{35}^{max} && (71)\\
0 \leq \; & C_{36}(k+j|k) \; \leq C_{36}^{max} && (72)\\
0 \leq \; & C_{37}(k+j|k) \; \leq C_{37}^{max} && (73)\\
0 \leq \; & C_{38}(k+j|k) \; \leq C_{38}^{max} && (74)\\
0 \leq \; & C_{39}(k+j|k) \; \leq C_{39}^{max} && (75)\\
\triangle C_{35}^{min} \leq \; & \triangle C_{35}(k+j|k) \; \leq \triangle C_{35}^{max} && (76)\\
\triangle C_{36}^{min} \leq \; & \triangle C_{36}(k+j|k) \; \leq \triangle C_{36}^{max} && (77)\\
\triangle C_{37}^{min} \leq \; & \triangle C_{37}(k+j|k) \; \leq \triangle C_{37}^{max} && (78)\\
\triangle C_{38}^{min} \leq \; & \triangle C_{38}(k+j|k) \; \leq \triangle C_{38}^{max} && (79)\\
\triangle C_{39}^{min} \leq \; & \triangle C_{39}(k+j|k) \; \leq \triangle C_{39}^{max} && (80)\\
& j = 1, 2, \ldots, m && (81)
\end{aligned}
$$

Here the prediction horizon p is 70 days and control horizon m is 60 days. The manufacturing nodes data implemented in this problem are listed in Table 7. The inventory targets and constraints are presented in Table 8. The two customer demands for fast and slow devices are stochastic with

				M10	M20	M30	M40
load=0% TO 70%Max	TPT	Min	days	30.0	5.0	1.0	1.0
	TPT	Ave	days	32.0	6.0	2.0	1.0
	TPT	Max	days	34.0	7.0	3.0	1.0
		Distribution		Unif	Unif	Unif	
load=70% TO 90%Max	TPT	Min	days	32.0	5.0	1.0	1.0
	TPT	Ave	days	35.0	6.0	2.0	1.0
	TPT	Max	days	38.0	7.0	3.0	1.0
		Distribution		Unif	Unif	Unif	
load=90% TO 100%Max	TPT	Min	days	35.0	5.0	1.0	1.0
	TPT	Ave	days	40.0	6.0	2.0	1.0
	TPT	Max	days	45.0	7.0	3.0	1.0
		Distribution		Unif	Unif	Unif	
Applied at Output	Yield	Min	%	93.0	98.0	98.5	100.0
	Yield	Ave	%	95.0	98.5	99.0	100.0
	Yield	Max	%	97.0	99.0	99.5	100.0
		Distribution		Unif	Unif	Unif	
Capacity	Load	Max	Items	45000	7500	2500	2500
	Load	Initial	Items	15785	2706	902	451

Table 7: Manufacturing nodes data in the test2 stochastic split problem: TPT-throughput time; Unif-uniform distribution

			I10	I20	I30	I31	I31
Inventory	UCL	Item	12000	4000	2000	2000	1000
	TAR	Item	3306	1102	551	351	176
	LCL	Item	1000	667	667	333	333
	Level	Max	20000	6667	6667	3333	3333
	Level	Initial	3000	2000	1000	1000	500

Table 8: Inventory setting in test2 stochastic split problem

different means and variances. Usually the demand for the slow device is higher than that for the fast device. So the average of the slow device demand is larger than that of the fast device demand. Depending on the split in Assembly/Test2, we will have different amounts of products to meet different customer demands. Based on the split factor in A/T2 node, we developed three cases for study. The average and variance of the split are shown in Table 9. These three cases are studied by using different tuning parameters to achieve robustness and better customer service. In each case, the results are compared based on the variance of each variable and the backorders created by the process.

6.1 Case 1 (Balanced)

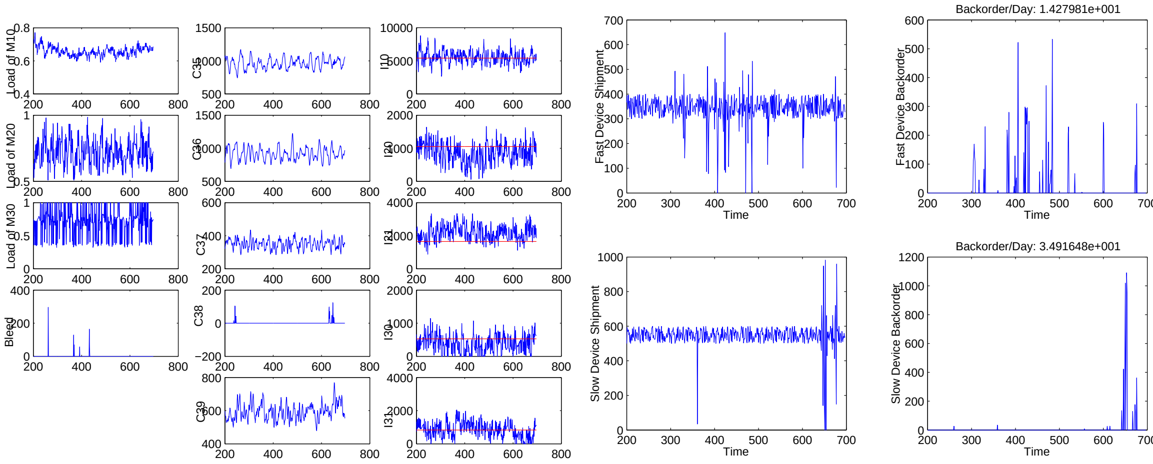
The average of the split in Case 1 is the same as the proportion of the two customer demands $\frac{D}{E}$ (balanced). That means the split will make the amount of the fast devices and the slow devices the same as the demands for the fast and the slow devices if no uncertainty occurs. The deterministic simulation proves all targets can be met without any backorder.

When uncertainty is introduced to TPT, yields, demands, split and small move suppression is used on all starts, many backorders are created because we run out of the $I30$ and $I31$ frequently. The responses are very aggressive and the variation of each variable is large as shown in Table 6.3. The load of $M30$ hits 100% many times. There are only a small amount of items shipped from $I20$ to meet the demand E through $C38$.

Although the split is balanced with respect to the customer demands, there are many backorders created for both demands. This is due to the stochasticity and aggressive responses in the system. If larger move suppression is implemented, the responses are shown in Figure 12. The number of backorders is decreased, all factories have loads within the capacity limits. There are still some backorders created for both demand D and E, because $I30$ and $I31$ are depleted sometimes. The unfilled orders for D exceed those for E. But compared with those using small move suppression, the number of unfilled orders are reduced. The variance of each variable is also decreased as shown, Table 6.3. Clearly, larger move suppression achieves improved robustness and customer service.

6.2 Case 2 (Reconfigure High)

In this case, the split average is larger than the average of $\frac{D}{E}$. So there are more products to meet the demand for D than to meet the demand for E at steady state. In a deterministic simulation, all of the demands and targets can be met. The difference compared to Case 1 is that $C38$ is not zero. Because the split in $M20$ does not make enough products to meet the demand for E, some of the fast devices in $I20$ are used to make the slow devices through $M20$ to meet the demand for E.



(a) Load, Starts and Inventories

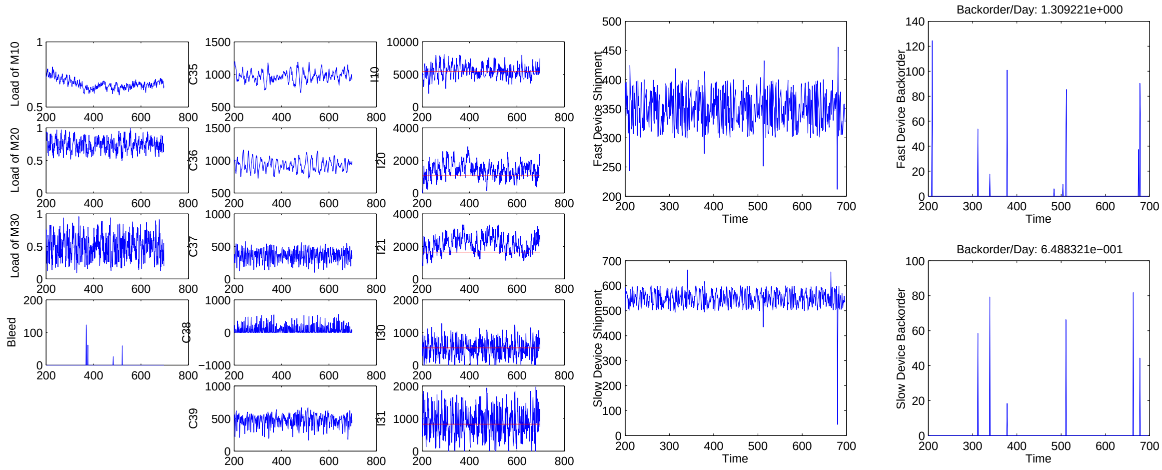
(b) Shipments and Backorders

Figure 12: Assembly/Test2 stochastic split problem Case 1: Move Suppression [10 10 10 0.1 20]

If small move suppression is used when uncertainty is introduced on the split, TPT, yield and demand, as shown in Table 6.3, the variation in each variable is large. Also due to the high uncertainty in the system and very aggressive responses, backorders occur on both demands. Although some items are transferred from $I20$ through $C38$, this is still not enough to meet the demand for E. The capacities of both $M20$ and $M30$ are exceeded. If larger move suppression is used for all manipulated variables except $C38$ (we need this manipulated variable to react as fast as possible to meet the request from demand E), the results are shown in Figure 13. The responses are much smoother than before. Robustness is achieved with much less backorders for both demands. The inventory levels are high enough to meet the customer demands in most cases. Because we have flexibility in using $C38$, we only have a little bit of backorders for the demand E although the split is not balanced. The load of $M20$ and $M30$ are decreased and smoothed. The oscillations observed are mainly due to the stochasticity within the manufacturing processes.

6.3 Case 3 (Discard low)

In Case 3, the average split is smaller than the average of $\frac{D}{E}$. So at steady state, there will be not enough items to meet the demand for D if demand for E is met without any excessive products left in $I21$ or $I31$. Even in a deterministic simulation, in order to meet the demand D, the total amount of items processed in $M10$ and $M20$ must be increased. So enough items will be shipped to $I20$ and $I30$ to meet the demand D, but there will be more than enough products to satisfy the demand for E. Because the capacity of $I21$ is limited, the inventory will be held until it hits the high limit of capacity and excessive products will be scrapped. Even in a deterministic environment, the capacities of F/T1 and A/T2 are more than 95%. Once stochasticity is introduced, the



(a) Load, Starts and Inventories

(b) Shipments and Backorders

Figure 13: Test2 stochastic split problem Case 2: Move Suppression [10 10 10 0.1 20]

original capacity will not be sufficient.

It is not surprising to have many backorders with small move suppression when stochasticity is introduced, because of the aggressive responses as shown in Table 6.3. Figure 14 shows the results using larger move suppression with full uncertainty in system. Even if larger move suppression is used for all manipulated variables except for $C38$, three factories' capacities are still exceeded. Not all items needed can be processed on time, and there is still not enough product for the demand of E. Many backorders are observed for both demands. This is generated by limited capacity and is a problem that can not be overcome simply by controller tuning.

If we increase the capacities of $M10$ and $M30$ by 20% and $M20$ by 30% and try different tuning parameters, the results are significantly improved. Table 6.3 shows the results for using both small and larger move suppression. Even with small move suppression, the backorders are much less than before. If larger move suppression is implemented, the backorders decrease dramatically as shown Figure 15. One can find here we have enough capacity to process items as many as they are needed and the inventory levels can track the targets without being depleted so often. Because we lose the flexibility of using $C38$ to make some fast devices from $I21$ and the split can not make enough fast devices, the backorders of the demand for D are more than that for E. Significant amount of scrap product is created from $I21$ in order to increase the amount of items to meet demand D.

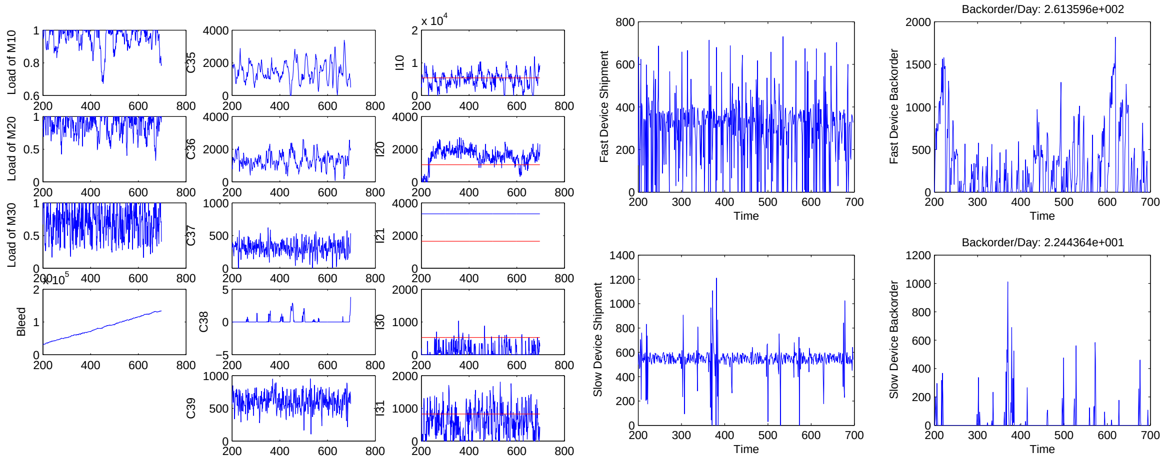
Table 6.3 shows all of the means and variances for signals involved in this problem. One can still come to the same conclusion we have made in the previous section. A larger move suppression is helpful to achieve robustness and improved customer service. In Case 3, in order to discard the

	I20	Hi	spd	I21	Low	spd
Bin Splits	Max	Average	Min	Max	Average	Min
Case1	0.49	0.39	0.29	0.71	0.61	0.51
Case2	0.59	0.49	0.39	0.61	0.51	0.41
Case3	0.39	0.29	0.19	0.81	0.71	0.61

Table 9: Three different split factor

	Case 1		Case 2		Case 3(Lim. Cap.)		Case 3(Exp. Cap.)	
Mean	No MS	Large MS	No MS	Large MS	Small MS	Large MS	Small MS	Large MS
I10	5493.6	5536.6	5676.0	5484.5	6882.8	4975.5	6042.1	5194.3
I20	1073.6	870.61	1267.9	1411.7	1917.9	1562.4	1332.7	1016.2
I21	1911.9	2184.2	1875.5	2147.8	3333	3333	3333	3333
I30	560.74	393.0	479.53	520.55	134.87	116.37	422.28	482.18
I31	870.38	847.5	482.03	884.60	614.71	600.99	723.40	818.15
C35	966.75	964.4	967.52	971.72	1466.0	1461.0	1339.2	1337.0
C36	913.05	919.6	973.04	926.22	1881.0	1333.4	1444.7	1265.7
C37	350.72	350.4	637.60	354.53	291.16	310.26	350.82	354.67
C38	0.5363	1.6380	68.98	86.31	0.9348	0.1514	0.6661	4.0982
C39	587.39	594.42	591.19	467.49	604.95	600.33	578.65	553.26
Load M10	0.5044	0.6585	0.6794	0.6760	0.8546	0.9517	0.7843	0.8183
Load M20	0.5366	0.7200	0.7247	0.7390	0.8199	0.8645	0.7552	0.7586
Load M30	0.6079	0.7133	0.6995	0.4967	0.6534	0.6979	0.5982	0.5971
Variance	No MS	Large MS	No MS	large MS	Small MS	Large	Small MS	Large MS MS
I10	1.1342e+6	9.1851e+5	1.7512e+6	9.4645e+5	6.2861e+6	4.9735e+6	5.1266e+6	2.8744e+6
I20	1.6262e+5	9.0873e+4	2.5645e+5	2.3683e+5	9.9933e+5	2.8015e+5	2.4247e+5	1.6516e+5
I21	3.3318e+5	2.0793e+5	2.9658e+5	2.5593e+5	0	0	0	0
I30	1.4782e+5	6.4252e+4	1.2848e+5	7.3027e+4	6.1321e+4	3.4869e+4	1.4979e+5	8.0216e+4
I31	3.0788e+5	2.0095e+5	1.9648e+5	1.8721e+5	2.0075e+5	1.7829e+5	2.0344e+5	1.6264e+5
C35	3.9726e+5	6.6934e+3	6.3544e+5	6.8067e+3	2.4655e+6	3.4924e+5	1.9016e+6	2.3937e+4
C36	3.5857e+5	9.0169e+3	5.0917e+5	7.2991e+3	2.4771e+6	1.8523e+5	1.8276e+6	2.0845e+4
C37	3.7024e+4	847.1811	6.6307e+4	8.3897e+3	7.2686e+4	1.0747e+4	7.8268e+4	399.8460
C38	1.0753	116.4988	7.0161e+3	1.6511e+4	2.4588	0.2302	0.8689	41.0133
C39	7.9210e+4	2.3592e+3	9.7238e+4	8.1256e+3	7.1503e+4	1.6417e+4	7.3031e+4	1630.4
Load M10	7.7746e-4	8.0797e-4	0.0020	0.0014	0.0190	0.0039	0.0062	0.0020
Load M20	0.0196	0.0092	0.0326	0.01	0.0474	0.0212	0.0435	0.0141
Load M30	0.0704	0.0414	0.0716	0.0347	0.0668	0.0505	0.0613	0.0433
"D" Not Filled	2.29%	2.12%	1.94%	0%	25.4%	16.22%	5.44%	0.58%
"E" Not Filled	1.18%	0.8%	5.59%	0.2%	2.37%	1.57%	1.0%	0%

Table 10: Comparison of different cases with different parameters in test2 stochastic split problem



(a) Load, Starts and Inventories

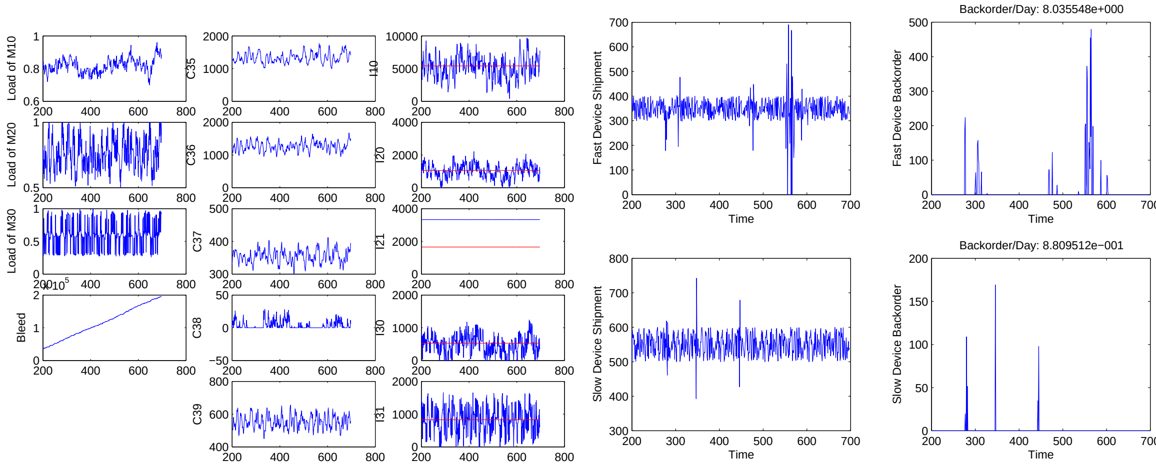
(b) Shipments and Backorders

Figure 14: Assembly/Test2 stochastic split problem Case 3 Limited Capacity: Move Suppression [100 100 1000 0.1 100]; Output Weight: [10 10 0 10 10 0 0 0]

excessive slow devices in $I21$, the output weight for $I21$ is set to zero. So $I21$ begins to scrap when it reaches the maximum and does not track any target. This shows the effect of the choice of output weights to select the position to hold the inventory. The capacity is the bottleneck in Case 3. Given enough capacity, better performance can be expected by tuning the move suppression and output weights at the cost of throwing away many items from $I21$.

7 Conclusions

Our analysis confirms the importance of the interplay between outer loop and inner loop per Figure 1. A Model Predictive Control formulation as an inner loop tactical controller in semiconductor manufacturing is presented and validated via simulation. Although relying on a linear model with fixed throughput times and yields, MPC can have satisfactory performance for systems with high stochasticity and uncertainty. The flexibility to handle constraints and the choice of tuning parameters makes it possible to track inventory targets while meeting customer demand as much as possible. The inventory targets and demand forecast will come from the outer loop by solving the optimization problem to maximize the profits. If good inventory targets and demand forecasts are provided, MPC can determine the manufacturing starts to get few or no backorders. The actions of the MPC controller can show when and where the capacities are depleted and give a reasonable justification for expanding capacity. Move suppression in MPC plays an important role in achieving robustness under uncertainty in systems. Increasing move suppression can help make the responses smooth with less backorders. It can also influence WIP in factories to not change



(a) Load, Starts and Inventories (b) Shipments and Backorders

Figure 15: Assembly/Test2 stochastic split problem Case 3 Increased Capacity: Move Suppression [100 100 1000 0.1 100]; Output Weight: [10 50 0 50 10 0 0 0]

sharply, which is desirable in practice. Output weights selection can be used to keep inventory levels at certain targets. The choice of output weights can determine where to hold the inventory and give us flexibility to track some if not all of the targets. Future research will focus on some more complex problems (such as shown in Figure 6 (c), 6 (d)) and combination problems that include configuration, packaging, shared capacity and correlated demand.

8 Acknowledgement

Support from the Intel Research Council and Institute for Manufacturing Enterprise System (IMES) in ASU is gratefully acknowledged.

References

ASCET (2003). Achieving supply chain excellence through technology. `http://www.ascet.com`.

Braun, M. W., D. E. Rivera, M. E. Flores, W. M. Carlyle and K. G. Kempf (to appear, 2003*a*). A Model Predictive Control framework for robust management of multi-product multi-echelon demand networks. *Annual Reviews in Control, Special Issue on Enterprise Integration and Networking.*

Braun, M. W., D. E. Rivera, W. M. Carlyle and K. G. Kempf (2002). A Model Predictive Control framework for robust mangement of multi-product, multi-echelon demand networks. In: *15th IFAC World Congress*. Barcelona, Spain.

Braun, M. W., D. E. Rivera, W. M. Carlyle and K. G. Kempf (2003*b*). Application of Model Predictive Control to robust management of multi-echelon demand networks. *Simulation: Transactions of the Society for Modeling and Simulation International* **79**(3), 139–156.

Chopra, S. and P. Meindl (2001). *Supply Chain Management: Strategy, Planning, and Operation.* Prentice-Hall, Upper Saddle River. New Jersey.

García, C. E., D. M. Prett and M. Morari (1989). Model Predictive Control: Theory and practice - a survey. *Automatica* **25**(3), 335–348.

Graves, S. C. and S. P. Willems (2000). Optimizing strategic safety stock placement in supply chains. *Manufacturing and Service Operations Management* **2**(1), 68–83.

Hopp, W. J. and M. L. Spearman (1996). *Factory Physics: Foundations of Manufacturing Management.* McGraw Hill. New York.

Kapuscinski, R. and S. Tayur (1999). Optimal policies and simulation-based optimization for capacitated production inventory systems. In: *Quantitative Models for Supply Chain Mangement* (S. Tayur, R. Ganeshan and M. Magazine, Eds.). Kluwer Academic.

Kempf, K. G. (2003*a*). Models needed for supply-demand network management. In: *Workshop on Transport in Supply Chains, Traffic, and Biology.* Arizona State University, `http://math.asu.edu/~dieter/workshop/`.

Kempf, K. G. (2003*b*). Optimization of a semiconductor supply chain. In: *Proc. INFORMS Conf. on OR/OS Practice*. Phoenix, AZ. pp. 43–49.

Perea, Edgar, Erik Ydstie and Ignacio Grossmann (2003). A model predictive control strategy for supply chain optimization. *Computers and Chemical Engineering* **27**, 1201–1218.

Simchi-Levi, D., P. Kaminsky and E. Simchi-Levi (2000). *Designing and Managing the Supply Chain.* McGraw Hill. New York.

SimulationDynamics (2003). Importance of supply chain management. http://www.simulationdynamics.com/Sc/SupplyChainImportance.htm.

Tzafestas, S., G. Kapsiotis and E. Kyriannakis (1997). Model-based predictive control for generalized production planning problems. *Computers in Industry* **34**(2), 201–210.

Wang, W., D. E. Rivera and K. G. Kempf (2003). Centralized Model Predictive Control strategies for inventory management in semiconductor manufacturing supply chains. In: *American Control Conference*. Denver, CO. pp. 585–590.