To be presented at the 2003 American Control Conference,
Denver, Colorado, June 4-6, 2003
Session: Control of Industrial Processes

Centralized Model Predictive Control Strategies for Inventory Management in Semiconductor Manufacturing Supply Chains

Wenlin Wang, Daniel E. Rivera[1]
Department of Chemical and Materials Engineering
Arizona State University, Tempe, Arizona 85287-6006

Karl G. Kempf
Decision Technologies, Intel Corporation
5000 W. Chandler Blvd., Chandler, AZ 85226

Abstract

Centralized strategies based on Model Predictive Control (MPC) are applied to inventory management problems associated with semiconductor supply chains. Specifically, two benchmark problems of relevance to semiconductor manufacturing are examined. The first is a single product, two node problem consisting of a Fab/Sort and an Assembly/Test facility controlled with a predictive controller using anticipation. The performance of the control scheme under conditions of plant-model mismatch and unforecasted demand are evaluated. The insights gained from this problem are used in the design of a centralized MPC controller for a four node problem involving two interconnected Fab/Sort and Assembly/Test facilities. In this latter problem, inventory management of wafer, die, and package inventories are considered.

1 Introduction

Supply chains (also known as demand networks) consist of several nodes or components. Both physical flows and decision flows are transformed along the chain. The overall goal of supply chain management is to maximize the four customer service factors - the right products, in the right quantity, in the right place, at the right time - while minimizing the four major costs - materials, production, storage, transport [1]. Improving the management of supply chains is one way in which modern enterprises can reduce their costs by millions, if not billions of dollars. In semiconductor manufacturing, lead times can range from weeks to months. This usually requires keeping safety stocks at very high levels, sometimes, as much as a whole year's worth of demand [2]. Operating in this fashion is clearly undesirable. The goal of this work, therefore, is to apply Model Predictive Control-based strategies to robustly manage inventory levels in supply chains despite inaccurate lead times and random disturbances.

Recent work using Model Predictive Control has shown it as an attractive method for inventory control [3] and supply chain management [4]. These approaches are conceptually different and require less detailed knowledge in comparison with cost-optimal stochastic programming solutions which require many "what-if" cases to be run and examined by highly skilled professionals [6]. Yet MPC offers the same flexibility in terms of the information sharing, network topology, and constraints that can be handled. 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, failures and disturbances which affect the system.

Previous work focused on partially decentralized control strategies [4]. A two product, six node three echelon supply chain was controlled by using three MPC controllers, one per echelon. This work focuses on fully centralized strategies, which are feasible for problems where all nodes belong to one enterprise.

The paper is organized as follows. In Section 2, the basic MPC algorithm is presented. A 2-node benchmark problem is examined under varying conditions using centralized strategy in Section 3. The benefits of a centralized strategy motivate an application to a four node problem that includes managing wafer and packaging inventories; this is described in Section 4. The

[1]To whom all correspondence should be addressed; phone:(480) 965-9476;email:daniel.rivera@asu.edu

paper ends with summary and conclusions in Section 5.

2 Model Predictive Control

Model Predictive Control is an optimization-based control scheme that can be tuned for good performance and robustness properties. 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 [5]. Another advantage is that it can easily handle constraints on manipulated and control variables. The MPC controllers considered in this paper are based on the state-space model below:

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

where $x(k)$ represents the state of the system, $u(k)$ are manipulated variables (MV) or command inputs, $v(k)$ is a vector of measured disturbances (MD), $d(k)$ are unmeasured disturbances (UD), and $y(k)$ is the output vector, which is composed of measured outputs (MY) $y_m(k)$ and unmeasured outputs (UY) $y_u(k)$.

The MPC controller selects the input $u(k)$ by solving the following optimization problem

$$\begin{aligned} J = & \sum_{i=0}^{p-1} \parallel w_i^u [u(k+i \mid k) - u_{target}(k)] \parallel^2 \\ & + \parallel w_i^{\triangle u} \triangle u(k+i \mid k) \parallel^2 \\ & + \parallel w_{i+1}^y [y(k+i+1) \mid k) - r(k+i+1)] \parallel^2 \end{aligned} \quad (2)$$

$$\min_{u(k|k),\ldots,\triangle u(m-1+k|k)} J \quad (3)$$

s.t.

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

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

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

$$y_i^{min} \leq y(k+i+1 \mid k) \leq y_i^{max} \quad (6)$$

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

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

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

Here p is the prediction horizon, m is the control horizon. $w_i^u, w_i^{\triangle u}, w_i^y$ are penalties on the control signal, move size and control error, respectively. This problem can be solved by standard quadratic program algorithms. Only the first control action is applied to system; after new measurements are available, a new optimization problem is solved. This is referred to as receding horizon principle. r is the references of the outputs; x_k is the anticipated reference value at time k. The use of future references in MPC is referred to as anticipative action when the value of reference is known in advance. A similar anticipative action can be performed with respect to measured disturbance $v(k)$. t_{final} is the end time for a whole simulation. Taking use of anticipation in the controller is a significant contributor to improved performance.

3 Inventory Management of a Two-Node Demand Network

In this section, a benchmark two node supply network problem is analyzed from a process control perspective. If all the facilities are owned by the same company, inventoy level and demand information can be shared easily. So a centralized control strategy is appropriate. All available information is fed to the controller and decisions can be made by taking into account all customer and node demands. A basic two-node system relevant to the semiconductor manufacturing industry is illustrated in Figure 1; it involves Fab/Sort (F/S) and Assembly/Test (A/T) facilities.

Figure 1: Material flow for 2 nodes:F/S stands for Fab/Sort, A/T stands for Assembly/Test

Clean wafers are fed to the Fab/Sort node. After processing for 6 to 8 weeks, wafers with die are shipped out from the Fab/Sort to the Assembly/Test node for testing and packaging. After approximately 2 weeks, chips with packages are shipped to a components warehouse to meet customers' demand. In order to buffer the variance of the material flow, an Assembly-Die-Inventory (ADI) is held before the A/T node and a Semi-Finished-Goods-Inventory (SFGI) is held after the A/T node.

Figure 2 represents a "fluid" analogy to this two node system which allows one to translate information in the supply chain to corresponding process control variables. Here the liquid levels correspond to the ADI and SFGI, respectively. The piping transportation lags denote the throughput time in the F/S and A/T factories. A linear model for this two tank system can be built based on the principle of conservation of total mass:

$$I_{ADI} = \frac{e^{-\theta_1 s}}{s} S_{F/S} - \frac{1}{s} S_{A/T} \quad (8)$$

$$I_{SFGI} = \frac{e^{-\theta_2 s}}{s} S_{A/T} - \frac{1}{s} D' \quad (9)$$

where I_{ADI} and I_{SFGI} are the deviation in ADI and SFGI, respectively. For simplicity, Δ is dropped in the

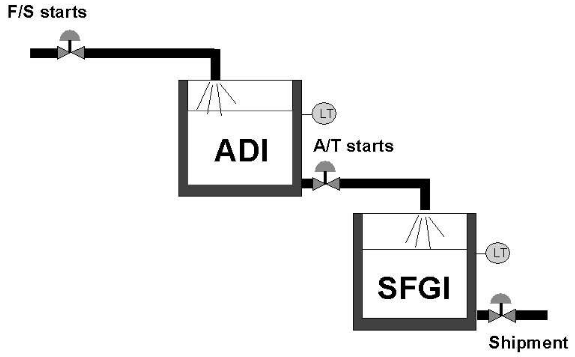

Figure 2: Fluid analogy of a two-node network

equations; $S_{F/S}$ and $S_{A/T}$ are the changes in the incoming and outgoing streams of ADI, respectively; $S_{A/T}$ is the change in the incoming stream of SFGI, while D' is the change in the outgoing stream of SFGI. θ_1 and θ_2 represent the throughput time in the F/S node and A/T node.

3.1 Centralized MPC controller structure

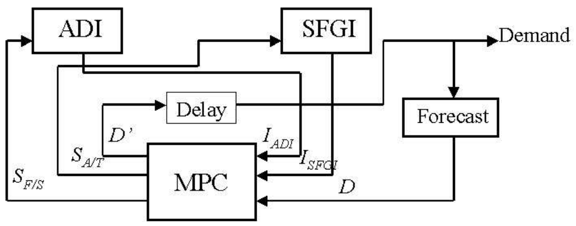

Figure 3: Centralized MPC controller information flow

In this case, a single MPC controller with anticipation is used as shown in Figure 3. Here it is assumed that the customer desired demand that is fed directly to the SFGI is known. The safety stock level is taken as the setpoint for the inventory level. The whole system is a three input/three output system, with the supply of the SFGI acting as a measured disturbance. The resulting control system is thus representative of a combined feedback and feedforward control system. The system can be described by following equations:

$$\begin{bmatrix} I_{ADI}(s) \\ I_{SFGI}(s) \\ D(s) \end{bmatrix} \begin{bmatrix} \frac{e^{-\theta_1 s}}{s} & -\frac{1}{s} & 0 \\ 0 & \frac{e^{-\theta_2 s}}{s} & -\frac{1}{s} \\ 0 & 0 & e^{-\theta_3 s} \end{bmatrix} \begin{bmatrix} S_{F/S} \\ S_{A/T} \\ D' \end{bmatrix} \tag{10}$$

The outputs/controlled variables here are the levels of the ADI and SFGI. The outflow of the SFGI represents the demand D from the customer that arrives θ_3 time units after an order is placed. The forecast of customer demand is used to anticipate changes in D by the controller, which will manipulate $S_{F/S}$ and $S_{A/T}$ to insure that D is satisfied while keeping the inventory level in each node at setpoint. Because the controller contains models for both nodes and the anticipation of future demand, good performance can be expected under conditions of no plant-model mismatch. Figure 4 shows the simulation results when perfect models for the F/S and A/T nodes are used in the MPC controller, where $w_i^y = 1$ for all outputs, $p = 50$, $m = 40$ and no move suppression is added. Here θ_1 is 8 weeks; θ_2 is 2 weeks. The transportation delay θ_3 is 1 week. At $t = 30$, the ADI is built up, followed by the SFGI at $t = 60$. At $t = 90$ customer demand is introduced and SFGI shipments begin to meet the demand. From the plots, it is clearly that all control variables track the setpoints perfectly without any overshoot or oscillation. Because no move suppression or constraints are imposed, the manipulated variables move sharply.

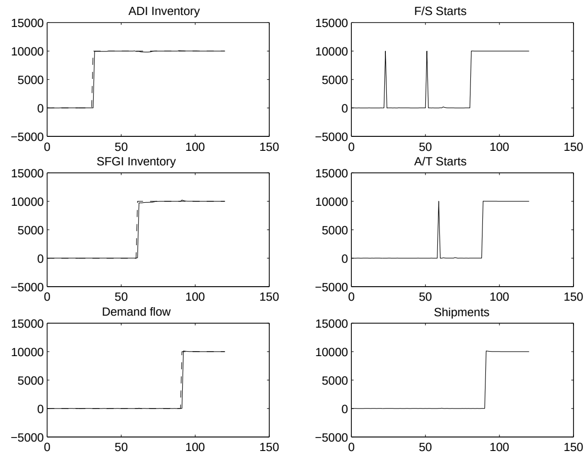

Figure 4: Simulation results with no plant model mismatch for Centralized MPC. Dashed: Setpoint for ADI and SFGI, forecast for demand flow; Solid: controlled and manipulated variables

When model mismatch is introduced between the control and simulation models in both the F/S and A/T nodes, robustness and stability are still possible with the proper selection of MPC tuning parameters. Figure 5 shows one such simulation result. Here the actual F/S delay is 10 weeks (25% greater than θ_1) and the A/T delay is 3 weeks (50% greater than θ_2). At first, no constraints are imposed by the controller. Move suppression weights of [200 200] are used for $S_{F/S}$ and $S_{A/T}$ respectively. Demand is satisfied and all responses are stable. However, the levels of ADI and SFGI are very high. In order to lower the level of inventories, output constraint limits in the MPC controller are set to 20,000 units for each node. Some inventory target tracking ability is sacrificed but as shown in Figure 5 the inventory levels decrease, the responses are stable, and demand is satisfied with available safety stock.

In practice, there is always some variability in the demand which is not captured by a forecast. This variability will be transferred upstream in the supply chain and can lead to the detrimental “bullwhip” phenomenon [2] if the controller is not properly tuned. In order to analyze this phenomenon in the centralized control scheme, autoregressive noise with variance of 411.4 is introduced to the outflow of the SFGI (D'). Figure 6 shows the results. From the plot it shows the variance is amplified due to the integration. But it is still within reasonable level as noted in Table 1. The variance on the ADI is 954.8 and on the SFGI is 971.1.

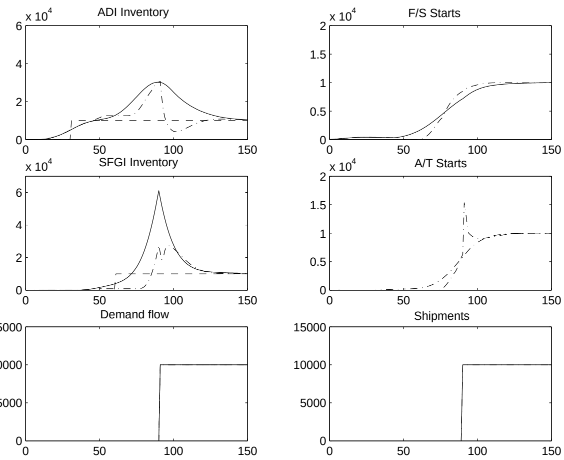

Figure 5: Simulation results with plant model mismatch for Centralized MPC. Dashed: Setpoint for ADI and SFGI, forecast for demand flow; Solid: unconstrained MPC solution; Dashed and dot: constrained MPC solution.

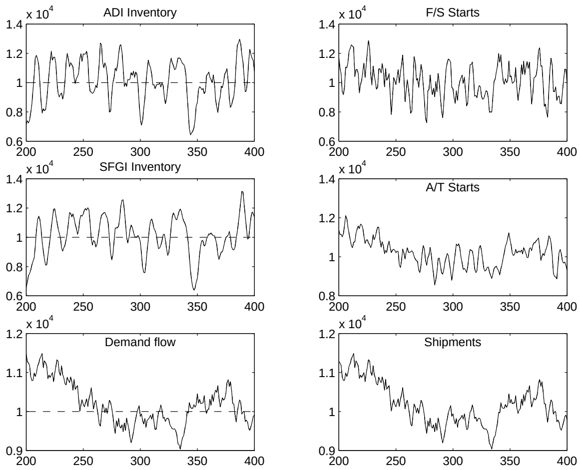

Figure 6: Simulation results with random disturbance for Centralized MPC. Dashed: Setpoint for ADI and SFGI, forecast for demand flow; Solid: controlled and manipulated variables

One of the advanges for MPC is the flexibility to tune the parameters in controller to meet different requirements. Figure 6 shows the results with the same output weights on both ADI and SFGI. But if different output weights are assigned to ADI and SFGI, the variance amplitude can be transferred to other parts of the system as shown in Figures 7 and 8.

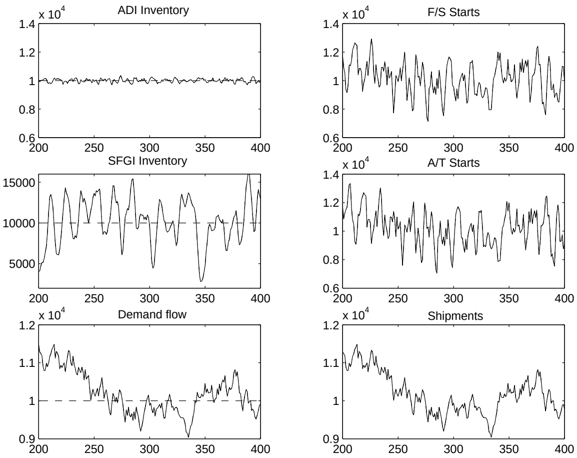

Figure 7: Simulation results with output weights [100 1]. Dashed: Setpoint for ADI and SFGI, forecast for demand flow; Solid: controlled and manipulated variables

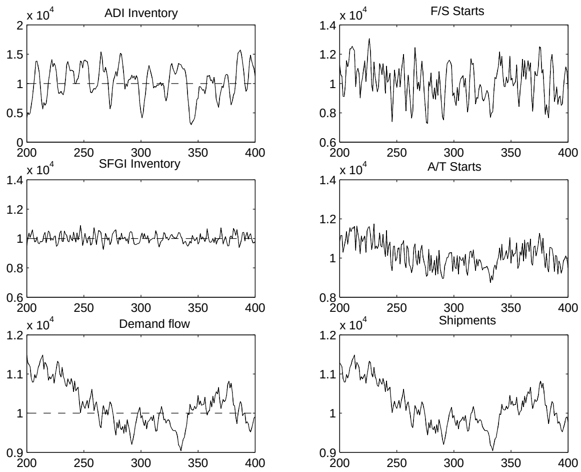

Figure 8: Simulation results with output weights [1 100]. Dashed: Setpoint for ADI and SFGI, forecast for demand flow; Solid: controlled and manipulated variables

In Figure 7, the output weight for ADI is 100 which is larger than that for SFGI which is 1. Clearly, the magnitude of variance on ADI is reduced, while that of SGFI is amplified. But the amplitude on ADI is much smaller than that with same output weights. If the output weight on SFGI is larger than that on ADI, we will get opposite results as shown in Figure 8. The numerical results of variance on ADI and SFGI under different output weights are shown in Table 1. These

OutputWeight	ADI	SFGI
1 1	954.8	971.1
100 1	91.6364	2035.4
1 100	1982.6	221.7234

Table 1: Variance for different output weight

results show the flexibility available to users through MPC to put different emphasis on different inventories. If variance reduction is needed, larger output weight should be put on this controlled variable.

4 Four-node case study with centralized MPC

The goal of this section is to show the possibility to apply the centralized control strategy discussed previously to a more realistic system with two echelons/four nodes/one product proposed by Intel corporation. Figure 9 shows the material flows of this problem.

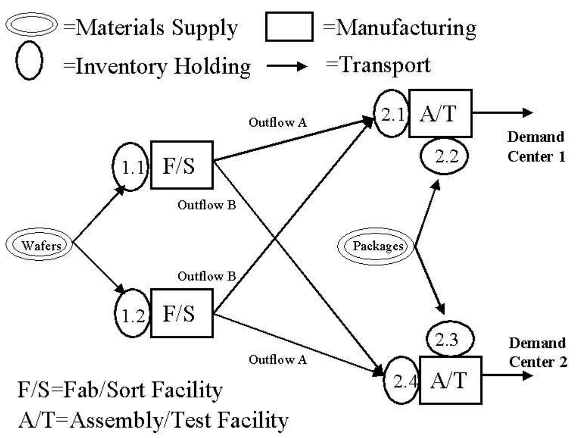

Figure 9: Material flows of four nodes supply chain

Wafers are shipped to the two F/S nodes and each of the F/S nodes has two outflows connected to each A/T nodes. Here outflow A has higher priority than outflow B. This means the demand for outflow A will always be filled first. If there are some wafers left, outflow B will be considered. For the A/T node, it has three inputs. Two of them are the outflows from both F/S nodes. Another one is the package supply. Each wafer combines with a package to make the final product for delivery. The wafer and package supply of each A/T node share the same customer demand. However they may be different due to noise or delay. So the output of A/T node will be the smaller value between wafers and packages which are ready for delivery at that day. Two customers' demands have to be met separately. There is no priority for these two demands.

One MPC controller is used. The overall dynamic model used in controller can be described by the following transfer function

$$\begin{bmatrix} I_{F/S1W} \\ I_{A/T1W} \\ I_{A/T1pa} \\ I_{F/S2W} \\ I_{A/T2W} \\ I_{A/T2pa} \\ Y_{A/T1} \\ Y_{A/T2} \end{bmatrix} = M \begin{bmatrix} P_{1W} \\ S_{F/S1A} \\ S_{F/S1B} \\ P_{2W} \\ S_{F/S2B} \\ S_{F/S2A} \\ P_{1pa} \\ S_{A/T1} \\ P_{2pa} \\ S_{A/T2} \end{bmatrix} \tag{11}$$

I stands for inventory, S for supply or outflow from each node, P stands for production, w stands for wafer and pa for package. M is a transfer function matrix composed of integrators with delays with similar structure to Equation 10.

For this centralized controller, the controlled variables are the inventories of each node I_*. The manipulated variables are the inflows and outflows of each node, except for the outflows of the two A/T nodes which treated as measured, anticipated disturbances. As before, there is one unit time delay between the outflow of A/T and the demand.

Simulation of the centralized MPC performance follows as before. Performance under plant-model mismatch is tested by setting the actual delays of the F/S and A/T nodes to 8 and 5 weeks, respectively, while the controller models delays are set to 6 weeks for the F/S nodes and 3 weeks for the A/T nodes. Step changes represent the demands for inventories and customer demands. Output weights are 1 for all outputs. Move suppressions for all manipulated variables are set to 15 to provide some control system robustness; constraint limits of 25,000 units are set on all inventory levels to keep these at a reasonable value. The prediction horizon is 50 and move horizon is 20. At first, all inventories have initial value which is 10,000 and the shipment is 2500. When t=150, the system experiences simultaneous setpoint and demand changes. The demand increases from 2500 to 5000. And the setpoints for all inventories of F/S and A/T are changed from 10000 to 20,000.

The simulation results are shown in Figure 10 and Figure 11. These indicate that the inventories decrease before the demand arises. This is partly due to anticipation in the MPC controller. When the controller anticipates the setpoint change, it will make the manipulated variables change before the setpoint change takes place to meet the demand. When the increase occurs, inventories are drained below their setpoint to fill the next node or customer demands. Then they will jump up and at last, all responses are stable and properly return to the setpoints. In any case, the demand is perfectly met and no backorder occurs.

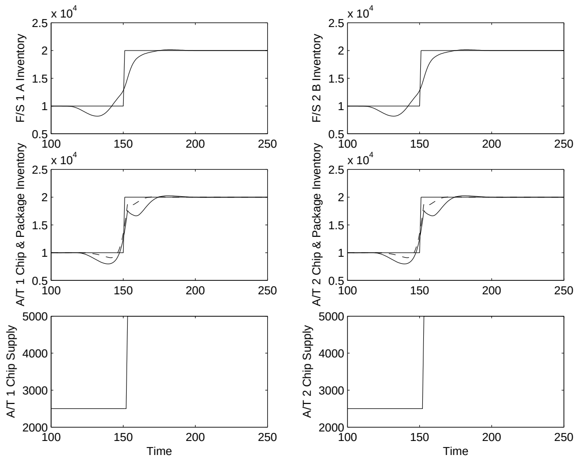

Figure 10: Simulation results for the two echelon/four node supply chain with Centralized MPC controller under plant-model mismatch; Solid: wafer controlled variables; Dash-dot: package controlled variables.

5 Summary and Conclusions

It has been shown that centralized MPC control strategies can be developed for inventory management in supply chains, and that these can be tuned for varying levels of performance even under conditions of plant-model mismatch and random disturbances. The centralized structure performs well because it has complete process knowledge and signal information which allows it to coordinate the decisions in the supply chains. The insights gained from the two node problem are used to develop a more sophisticate centralized control strategy for a four node network of interest to Intel Corp. A more formal analysis of the robustness and performance properties of these different structures and evaluation under condition of throughput time and yield uncertainty represents future directions in this research.

6 Acknowledgments

Support from the Intel Research Council and ASU's Institute for Manufacturing Enterprise Systems (IMES) is gratefully acknowledged.

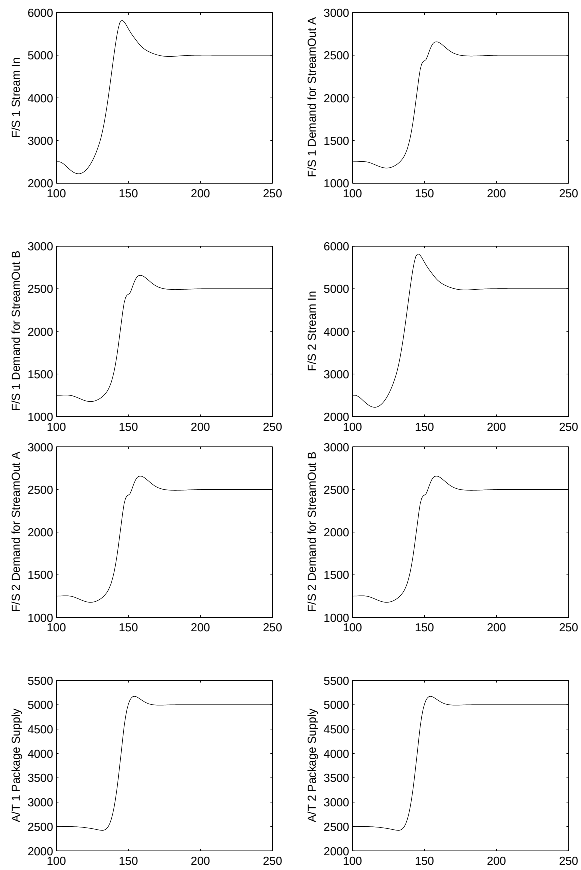

Figure 11: Simulation results for all manipulated variables of the two echelon/four node supply chain with Centralized MPC controller under plant-model mismatch.

References

[1] K.G. Kempf, K. Knutson, Fowler, J., B. Armbruster, P. Babu, and B. Duarte. Fast accurate simulation of physical flows in demand networks. *Proc. Int. Conf. on Semiconductor Mfg. Operational Modeling and Simulation '01 Seattle, WA.*, April 24-25, 2001, pp.111-116.

[2] Lee, H.L., V. Padmanabhan and S. Whang. The bullwhip effect in supply chains. *Sloan Management Review*, 1997, pp 93-102.

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

[4] Braun, M.W., D.E. Rivera, W. M. Carlyle, and K. G. Kempf , A Model Predictive Control Framework for Robust Management of Multi-Product, Multi-Echelon Demand Networks, 15th IFAC World Congress, Barcelona, Spain, July 21-26, 2002.

[5] C.E. Garcia, D.M. Prett, and M. Morari. Model predictive control: theory and practice-a survey. *Automatica*, Vol. 25, No. 3, pp 335-348, 1989.

[6] Kafoglis, C. C. Maximize competitive advantage with a supply chain vision. *Hydrocarbon Processing* pp. 47–50, 1999.