# Impact of Partial Manufacturing Flexibility on Production Variability

Ana Muriel
Anand Somasundaram
Yongmei Zhang
University of Massachusetts Amherst

## Abstract

As manufacturers in various industries evolve toward predominantly make-to-order production to better serve their customers' needs, increasing product mix flexibility emerges as a necessary strategy to provide adequate market responsiveness. However, the implications of increased flexibility on overall system performance are widely unknown. We develop analytical models and an optimization-based simulation tool to study the impact of flexibility on shortages, production variability, component inventories and order variability induced at upstream suppliers in general multi-plant multi-product make-to-order manufacturing systems. Our results show that:

1. Partial flexibility leads to a considerable increase in production variability, and consequently in component inventory levels and upstream order variability. Although a modest increase in flexibility yields most of the sales benefits, production variability is reduced as more flexibility is added to the system. Consequently, investments in additional flexibility will be justified when component inventories are expensive.
2. The performance of flexible systems is highly dependent on the capacity allocation policies implemented. Policies that evenly distribute capacity to product demands lead to consistently better performance, since they avoid the misplacement of inventories by replicating the performance of a single-plant system.

These insights and the simulation tool can be used by practitioners to guide the design of their flexible production systems, trading off the initial capital outlay versus the sales benefits and the expected system operational costs.

## 1. Introduction

A modern manufacturing firm operating in today's dynamic business environment needs to orchestrate its supply chain proactively, to respond effectively and efficiently to consumer demands in an increasingly competitive marketplace. Recent developments in communications and production technologies are making it possible for companies to offer high product variety at low cost while providing high service levels and on-time delivery performance. As the pressure mounts to give consumers exactly what they want, where they want it, when they want it and at a competitive price, many firms, whose production has traditionally been driven by demand forecasts, are looking for ways to move from predominantly make-to-stock to make-to-order systems (Bish et al. (2001)).

In a pure make-to-order system, the firm holds no final goods inventory. However, it does need to keep inventory of components with long lead times to cope with the variability in consumer demands and ensure timely delivery. In addition, the firm can invest in flexible capacity to accommodate this variability. We focus on product mix flexibility[1] which, in a general multi-product multi-plant manufacturing system, enables each plant to produce a variety of products and some products to be built in different plants to better match supply and demand. The benefits of flexibility in increased sales and capacity utilization in such systems have been thoroughly studied by Jordan and Graves (1995) considering a single production period. They introduce the concept of *chaining* to achieve maximum benefits from limited flexibility configurations where each plant produces only a few of the products. Furthermore, this strategic analysis has been extended recently to multi-stage supply chains by Graves and Tomlin (2000). The question remains, however, as to what the impact of the additional flexibility is on other system performance measures, such as component safety stocks and order variability propagating upstream. Note that in the make-to-order setting, flexibility is used to hedge not only against the

[1] Manufacturing flexibility has been extensively studied in the literature. For brevity, we refer the reader to the recent reviews of Beach et al. (2000), De Toni and Tonchia (1998), de Groote (1994), Kouvelis (1992) and Sethi and Sethi (1990).

long-term forecast uncertainty faced at the investment stage, but also to accommodate the variability in customer orders every period.

Bish et al. (2001) study the impact of flexibility and various capacity allocation policies on supply chain performance in a 2-product, 2-plant make-to-order setting through stylized analytical models. They focus on the management of flexible capacity at the tactical level. The analysis of complex multi-product multi-plant systems, however, requires more general capacity allocation policies and hence a different approach to analyze the cost-benefit tradeoffs of manufacturing flexibility. We model such a system and develop an optimization-based simulation tool to analyze the impact of different levels of flexible capacity on system performance at the operational level. Our objective is to provide valuable managerial insights and a support tool to guide the decision making process of a firm in planning the extent to which it should invest in flexible capacity and in deciding how flexibility should be managed later.

The remainder of the paper is organized as follows. Section 2 motivates the central research question through a simple example. Section 3 further explores the impact of partial flexibility on component inventory requirements through a stylized analytical model. Section 4 introduces the optimization-based simulation model and the simulation results, which reinforce the previous findings and offer additional insight. Section 5 justifies the selection of a distributed capacity allocation policy in the simulation model. We conclude with a summary of our findings and managerial insights.

## 2. Motivating Example

Consider a 3-product 3-plant production system with plant capacities all equal to C. Each product is built from a unique component and lead times for component replenishment are equal to 1 for all plant-component pairs; that is, component orders are placed at the beginning of each period, before demand is observed, and received immediately. Inventory levels are set such that any shortfall will be caused by capacity and not inventory. The production system is build-to-order and plans production in each period to minimize shortfall. Product demands in any given period come from one of two equally likely scenarios. In scenario 1 product demands are $(d_1,d_2,d_3) = (2C,C/2,C/2)$ and in scenario 2 product demands are $(d_1,d_2,d_3) = (C/2,C/2,2C)$.

In the 3-product, 3-plant scenario we consider three flexibility configurations: Dedicated, Partial Flexibility and Full Flexibility. These are depicted in Figure 1, where circles denote products, squares denote plants and links denote plants that can build products.

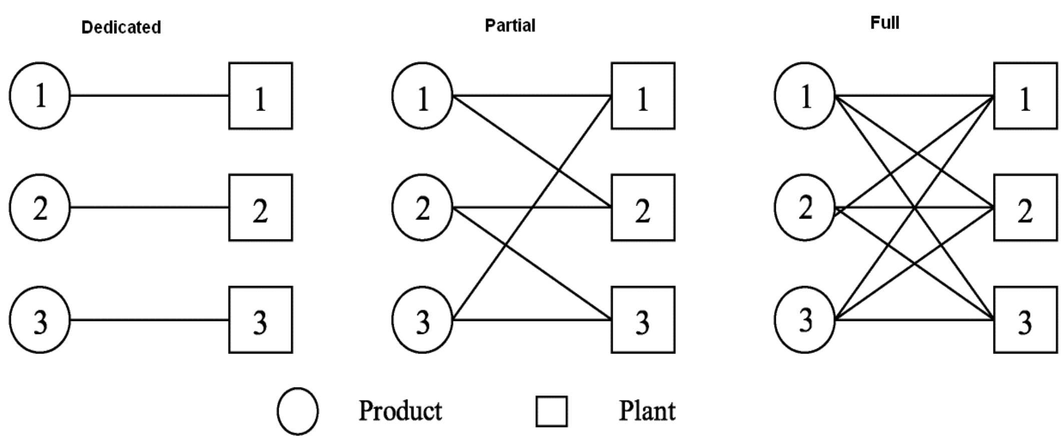


**Figure 1: No flexibility, partial flexibility and full flexibility configurations of 3-product 3-plant system.**

The following matrices display the production quantities of product i at each plant j, $x_{ij}$, for $i=1,2,\ldots,m$, and $j=1,2,\ldots,n$, under each flexibility configuration and demand scenario. It should be noted that there are often multiple solutions to the production planning optimization problem. Ties have been broken so that system inventory is minimized.

**Dedicated Configuration**

**Scenario 1** **Scenario 2**

$$\begin{pmatrix} C & 0 & 0 \\ 0 & C/2 & 0 \\ 0 & 0 & C/2 \end{pmatrix} \qquad \begin{pmatrix} C/2 & 0 & 0 \\ 0 & C/2 & 0 \\ 0 & 0 & C \end{pmatrix}$$

**Partial Flexibility Configuration**

**Scenario 1** **Scenario 2**

$$\begin{pmatrix} C & C & 0 \\ 0 & 0 & C/2 \\ 0 & 0 & C/2 \end{pmatrix} \qquad \begin{pmatrix} 0 & C/2 & 0 \\ 0 & C/2 & 0 \\ C & 0 & C \end{pmatrix}$$

**Full Flexibility Configuration**

**Scenario 1** **Scenario 2**

$$\begin{pmatrix} 2C/3 & 2C/3 & 2C/3 \\ C/6 & C/6 & C/6 \\ C/6 & C/6 & C/6 \end{pmatrix} \qquad \begin{pmatrix} C/6 & C/6 & C/6 \\ C/6 & C/6 & C/6 \\ 2C/3 & 2C/3 & 2C/3 \end{pmatrix}$$

If component inventories are never to be the cause of shortage (i.e. inventory is equal to the maximum production of product i in plant j over both scenarios), the system inventory before demand is observed must be as follows:

| Configuration | Expected Shortfall | Comp. 1 Inventory | Comp. 2 Inventory | Comp. 3 Inventory | System Inventory |
|---|---|---|---|---|---|
| **Dedicated** | C | C | C/2 | C | 5C/2 |
| **Partial Flex.** | 0 | 2C | C | 2C | 5C |
| **Full Flex.** | 0 | 2C | C/2 | 2C | 9C/2 |

**Table 1: Shortfall and required system inventory under each configuration.**

Adding flexibility to the dedicated configuration to create a chain where each product can be built in two plants, reduces the expected shortfall to zero but requires doubling the system inventory to 5C. Additional flexibility to create a fully flexible system then reduces the system inventory to 9C/2. In fact, 9C/2 is the minimum possible inventory for any capacity allocation or production plan that results in zero expected shortfall (as the sum of the maximum scenario product demands equals 9C/2).

The partially flexible configuration performs poorly relative to the fully flexible one because minimizing the shortfall results in large swings in the production of the three products at each of the plants between scenario 1 and 2. Additional flexibility enables more stable production irrespective of the scenario.

This stylized example suggests that while a small amount of flexibility configured in a chain can result in substantial reduction of shortfall (as shown by Jordan and Graves), the shortfall reduction may come at the price of substantial production variability leading to increased inventory needs. It also suggests that further increasing the flexibility allows the supply chain to reduce the production variability and thus reduce the inventory requirements. As a result, additional flexibility should be considered in industries with high-value components. These issues are explored further in the next two sections, analytically for a stylized model and through simulation for more general settings.

## 3. Stylized Analytical Model and Results

Consider an N-product N-plant production system. As in the previous section, assume that plant capacities are all equal to C, each product is built from a unique component with a supply lead time of 1, and the production system builds to order and plans production in each period to minimize shortfall. Let the flexibility configuration be specified by the set A of product-plant links; that is, product i can be built in plant j if and only if $(i,j)\in A$. Let $I_{ij}$ be the component inventory level of product i in plant j for each $(i,j)\in A$. Consider the production-planning problem where inventory is ignored and capacity and flexibility are the only constraints. Let **X(D)=** $\{x_{ij}(D), \forall(i,j)\in A\}$ be an optimal solution to this problem when demand is **D**$=(d_1,d_2,\ldots d_N)$.

Adding flexibility corresponds to adding new product-plant links. Following the notation in Graves and Tomlin (2000), "we define an h-type chain as a complete chain in which every product can be produced in h plants and every plant produces h products." This definition builds upon the concept of chaining of Jordan and Graves: "A chain is a group of products and plants that are all connected, directly or indirectly, by product assignment decisions. In terms of graph theory, a chain is a connected graph. Within a chain, a path can be traced from any product or plant to any other product or plant via product assignment links. No product in a chain is built by a plant from outside the chain; no plant in a chain builds a product from outside the chain." Observe that there exist different h-type chain configurations possible depending on the chosen assignment of products to plants. Throughout the paper, we add flexibility by increasing h.

Let us draw upon the notion of bounded demand from Simpson (1958). In Simpson, managers set a bound on the demand that they are willing to meet from inventory. Inventory is set to guarantee that demand within the bound is met. Demand outside the bound is lost. Looser bounds are associated with higher customer service levels. In the production system described above, the inventory in question is the component inventory. The ability to fulfill an order is limited by inventory, capacity and the flexibility configuration. As such, the Simpson demand bound is not directly applicable. However, the concept is still useful. Managers can set a demand bound **B** such that if $\mathbf{D} \in \mathbf{B}$, then they guarantee component inventory will not cause shortfall. Capacity and flexibility might still cause shortfall but the inventory will not. If $\mathbf{D} \notin \mathbf{B}$, then component inventory may cause shortfall. The demand bound approach allows us to specify inventory levels much more readily than a service level approach would. Assuming that customer demand is unbounded, component inventories are then set as $I_{ij} = \max_{D \in B}\{x_{ij}(D)\} \quad \forall (i,j) \in A$ and 0 otherwise. Note that if there exists multiple solutions to the production planning problem then the **X**(**D**) that minimizes the system inventory is chosen.

**Result A:** *Let $B = \Re^N$, i.e. management commits to having sufficient component inventory never to cause shortfall. In essence, they commit to providing a 100% service level to the plants. Then*

*(a) System inventory is increasing in the level of flexibility.*

*(b) The system inventory for a chain in which each plant can build h products equals NhC.*

*Proof:*

$\mathrm{I}_{\mathrm{ij}} = \max_{\mathrm{D} \in \Re^{\mathrm{N}}}\{\mathrm{x}_{\mathrm{ij}}(\mathrm{D})\} = \mathrm{C} \quad \forall (\mathrm{i}, \mathrm{j}) \in \mathrm{A}$ and $\mathrm{I}_{\mathrm{ij}}=0$ for all other product-plant pairings. Therefore the system inventory equals MC, where M is the number of product-plant links. System inventory is therefore increasing in supply chain flexibility. □

The higher inventory levels are required so that the available capacity and flexibility can always be used to the fullest to minimize shortages. What if managers are not committed to ensuring shortfall is never inventory driven?

**Result B:** *Let the demand bound B be specified by*

$$(i)\ \sum_{k=0}^{n} d_{i+k} \leq \left(\min\{n+2, N\}\right)C \quad n = 0, \ldots, N \quad i = 0, \ldots, N \qquad (ii)\ d_N \leq C$$

*and consider h-type chains such that product i can be produced at plants i, (i+1)modN,…, (i+h-1)modN. Then:*

*(a) For a completely dedicated system, the system inventory is NC but shortfall will be positive* for some $\mathbf{D}\in\mathbf{B}$

(b*) For the h=2 chain, the system inventory is 2NC and shortfall is zero for all* $\boldsymbol{D}\in\boldsymbol{B}$

*(c) For the h=3 chain, the system inventory is 2NC-C and shortfall is zero for all* $\boldsymbol{D}\in\boldsymbol{B}$

*(d) The minimum system inventory needed is 2NC-C for any flexibility configuration that delivers zero shortfall for all* $\boldsymbol{D}\in\boldsymbol{B}$

Observe that condition (i) actually specifies the complete set of demand realizations for which the h=2 chain results in zero shortfall. So if **B** were defined by (i) only, then it would be equivalent to managers never wanting component inventory to cause shortfall if flexibility is configured according to this h=2 chain. **B** is actually more restrictive here due to condition (ii) which allows inventory to cause shortfall if demand for product N exceeds C.

*Proof:*

(a) For i=1,..,N, $\mathbf{D}$=(0,0,..,$d_i$=C,0, ..,0)$\in\mathbf{B}$. Therefore, inventory needs to be set to $I_{ii}$=C and $I_{ij}$=0 for $j \neq i$, since product i cannot be made in any plant other than i. As a result, system inventory is NC. For i<N, $\mathbf{D}$=(0,0,..,$d_i$=2C,0, ..,0)$\in\mathbf{B}$ and the dedicated system can only produce C, so there is a shortfall for some elements of **B**.

(b) A shortfall greater than zero in the h=2 chain would imply that $\mathbf{D}\notin\mathbf{B}$; therefore for every $\mathbf{D}\in\mathbf{B}$ there exists a production plan that is capacity and flexibility feasible with no shortfall.

For i<N, **D**=(0,0,..,$d_i$=2C,0, ..,0)$\in$**B**. This implies that the inventory levels must be set to $I_{ij} = C \ \ \forall (i,j) \in A, \ \ i < N$. As production of product i is only possible in plant j if the (i,j) link exists, $I_{ij}$=0 for $(i,j) \notin A$. This still leaves the question of how to set $I_{N1}$ and $I_{NN}$. Consider **D**=(2C,0,0,0,…,0,C)$\in$**B** . In this case the only feasible production plan that results in zero shortfall is given by the following production matrix **X(D)**=($x_{ij}$(**D**)).

$$\begin{pmatrix} C & C & & & \\ & 0 & 0 & & \\ & & \ddots & \ddots & \\ & & & 0 & 0 \\ 0 & & & & C \end{pmatrix} \Rightarrow \text{Therefore } I_{NN} \text{ must be set equal to C.}$$

Consider **D**=(0,0,0,0,…,2C,C)$\in$**B** . In this case the only feasible production plan that results in zero shortfall is given by

$$\begin{pmatrix} 0 & 0 & & & \\ & 0 & 0 & & \\ & & \ddots & \ddots & \\ & & & C & C \\ C & & & & 0 \end{pmatrix} \Rightarrow \text{Therefore } I_{N1} \text{ must be set equal to C.}$$

Total system inventory is then given by $\sum_{(i,j)\in A} I_{ij} = 2NC$

(c) As the flexibility configuration for the h=2 chain is contained within the flexibility configuration for the h=3 chain, there exists a production plan that is capacity and flexibility feasible for every **D**$\in$**B** such that shortfall is 0. Set the component inventories as shown in the production matrix below where $I_{ij}$ = 0 if element (i,j) is not shown.

$$\begin{pmatrix} C & C & 0 & & \\ & C & C & 0 & \\ & & \ddots & \ddots & \ddots \\ C & & & C & 0 \\ 0 & 0 & & & C \end{pmatrix}$$

Any demand $\mathbf{D} \in \mathbf{B}$ can be met with zero shortfall as the (N-1xN-1) sub-matrix obtained by eliminating the Nth column and Nth row is equivalent to the capacity matrix for the h=2 chain for an N-1 plant-product system. Total system inventory is $\sum_{(i,j)\in A} I_{ij} = 2(N-1)C + C = 2NC\text{-}C$.

(d) For i<N, $\mathbf{D}=(0,0,..,d_i=2C,0,\ ..,C) \in \mathbf{B}$. As a result, inventories need to be set such that $\sum_{j=1}^{N} I_{ij} \geq 2C \quad \forall i < N$ and $\sum_{j=1}^{N} I_{Nj} \geq C$ if shortfall is not to be inventory driven. Therefore total system inventory is $\sum_{i=1}^{N}\sum_{j=1}^{N} I_{ij} \geq 2(N-1)C + C = 2NC\text{-}C$ . As the h=3 chain attains this lower bound the minimum system inventory is 2NC-C. □

This result shows that *limited flexibility can require large production swings but increasing flexibility can reduce such swings*. Furthermore, it also suggests that there may be a point (h=3 in the stylized model) after which additional flexibility does not improve the system. To study the effect of partial flexibility on production variability in general settings, we develop a simulation tool.

## 4. Optimization-Based Simulation Model

In this section, we develop a simulation tool to evaluate the impact of the flexibility configuration on expected shortages, inventory levels and order variability that will occur at the operational level. For this purpose, we use the three-stage model proposed by Bish et al. (2001) to capture the relevant strategic, tactical and operational level decisions; see Figure 2.

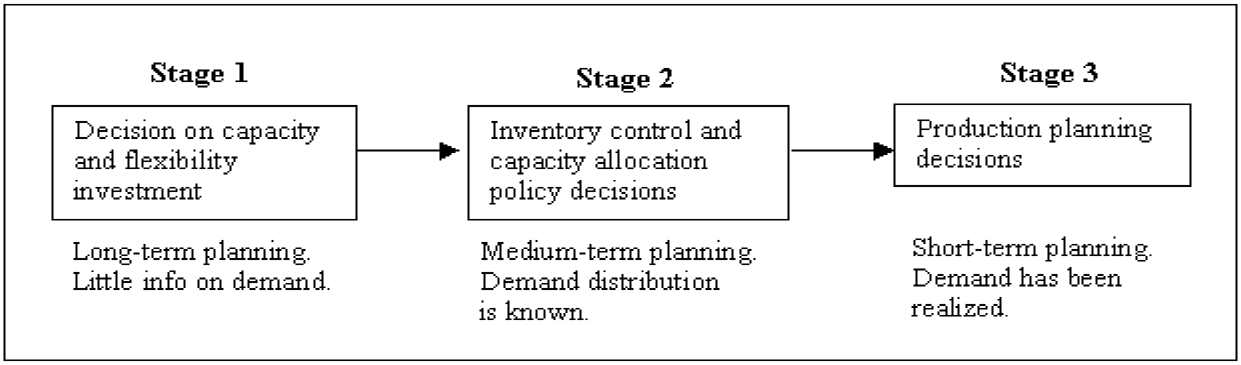


**Figure 2: Three-Stage Decision Framework**

First, the strategic investment decision is made long before production starts under a high level of uncertainty in product demand. In our simulation model, we capture this uncertainty by considering that the variability in the prior distribution of the mean demand is high. For instance, in the automotive industry, which inspired this study, the forecast error is typically around 40% on average at this stage (Jordan and Graves (1995)). At the tactical level, however, the distribution of product demand is fairly well known, but there is still the necessity to hedge against period-to-period variability. We consider a period to be an interval of time in which production is scheduled and fixed (e.g. one week for an automotive manufacturer in a make-to-order environment). In the simulation model we assume that at the tactical level the mean demand is fixed and generate normal period demands with that mean and a given, relatively small, coefficient of variation ($\sigma / \mu = 0.15$). Based on the observed distribution of demand, the firm needs to decide how to allocate system capacity to the different products, setting target levels of production of each product at each plant, and the operational policies to provide the human and material resources necessary to meet the uncertain demand. In particular, it needs to adopt a production planning policy and an inventory control policy to ensure appropriate levels of component inventory, given long lead times ($L>1$). Finally, at the operational level, the firm makes the production plan at the beginning of each period after orders have been received. The production plan is dictated by the capacity allocation policy chosen at the tactical level and the component inventory on-hand. Demand that cannot be assigned to a plant is lost.

The optimization-based simulation tool will follow the three stages described above. Before we proceed to develop the simulator, we need to address the medium/short term decisions: (1) How should production be planned at the operational level? (2) How should target production levels be estimated? (3) What inventory control policy should be used?

The production planning and component inventory control policy decisions are interrelated. The production planning policy induces variability in the amount produced at each plant during each period, and thus drives the need for component safety stocks. On the other hand, the performance of the production planning policy is fully dependent on the availability of component inventories. Ideally, we would want to consider these problems simultaneously. The joint optimization of production allocation and inventory policies would entail solving a multi-period multi-product stochastic inventory problem with lost sales and positive lead times, where each period requires the allocation of the available capacity and components to the realized demand. The resulting inventory/allocation problem is computationally intractable; see Muriel et al. (2001). In fact, even the structure of optimal policies for the single-product case remains an open question (Zipkin (2000)). Under the assumption of immediate component inventory replenishment (i.e. $L=1$), however, interesting structural results have been recently derived in Van Mieghem and Rudi (2002) for a joint capacity, inventory and production allocation problem in a very general setting.

Consequently, we focus on rational production planning and inventory policies that implicitly capture some of the interaction between the two. The next three sections address the necessary decisions. In what follows, the distribution of period demand, its mean and variability are assumed known and stationary at the tactical level.

### 4.1 Production Planning Problem

Given the predetermined target production levels, how should the realized demands be allocated to the available capacity, flexibility and inventory at the operational stage? We would like to obtain an "optimal" production plan; but, what should be the performance measure with respect to which we define optimality? As in Jordan and Graves (1995), the principal objective is to minimize shortages and this can be achieved by solving a simple linear program. We observe, however, that such a production policy could lead to wild swings in period-to-period production of a product at a particular plant. To overcome this drawback, we add a quadratic term to the mathematical program to penalize variability from some prespecified target in the period-to-period production levels.

Let n be the number of plants and m the number of products. In the following we index plants in j=1,2,…,n and products in i=1,2,…,m. Let $C_j$ be the capacity of plant j and A denote a given flexibility configuration, where $(i,j) \in A$ if product i can be built in plant j. Let variable $s_i^t$ represent the shortage of product i at time t and variable $x_{ij}^t$ represent the quantity of product i to be produced at plant j, for all $(i, j) \in A$. Finally, let $\varepsilon$ be a constant and let $\mu_{ij}$ denote the *target production level* (see Section 4.2) of product i at plant j, for all $(i, j) \in A$. Each period t, given each product's demand realization, $d_i^t$, and the available inventory of components for product i at plant j, $I_{ij}^t$, the minimum possible shortage for configuration A can be determined by solving the following *Production Planning Problem*:

$$Min \quad \sum_{i=1}^{m} s_i^t + \varepsilon \sum_{(i,j)\in A} (x_{ij}^t - \mu_{ij})^2$$

$$s.t.$$

$$\sum_{j:\ (i,j)\in A} x_{ij}^t + s_i^t = d_i^t \quad \text{i} = 1,\dots,\text{m}, \qquad (1)$$

$$\sum_{i:\ (i,j)\in A} x_{ij}^t \le C_j \quad \text{j} = 1,\dots,\text{n}, \qquad (2)$$

$$0 \le x_{ij}^t \le I_{ij}^t \quad \forall(\text{i, j}) \in \text{A} \qquad (3)$$

$$s_i^t \ge 0 \quad \text{i} = 1,\dots,\text{m} \qquad (4)$$

Constraints (1) define the shortage of each product as the difference between its demand and overall production. Constraints (2) and (3) ensure that production does not exceed plant capacity or available component inventory. The term $\varepsilon$ is chosen to be small ($10^{-3}$ in our simulation runs) so that the quadratic program minimizes the total shortage, while selecting the production quantities that will result in the minimum deviation from the target production levels, and hence, keep production variability as low as possible. Its value can be varied to reflect the relative importance of variability versus shortages in the system, see Section 4.5.2.

The quadratic program can be efficiently solved with off-the-shelf mathematical programming solvers, such as CPLEX. We use the barrier optimizer in CPLEX version 6.5 in our computational study.

### 4.2 Target Production Levels and Production Variability

For a given flexibility configuration, the firm needs to decide what specific resources, such as component inventories and trained workforce, to make available to each of the products.

These tactical decisions require knowledge of the target level of production of each of the products at each plant and of production variability.

How should the target production levels be chosen? There are a myriad possibilities. This decision not only depends on the expected revenues and inventory costs but also on other tactical and strategic choices of the firm, such as the value placed on economies of scale or the plants' proximity to customers and suppliers of the different products.

We consider a *distributed allocation policy* that aims to evenly distribute product demand among all the plants that can produce it, subject to capacity restrictions at each plant. The distributed policy may be attractive in practice because it enables production closer to the customer when the plants are geographically dispersed, and thus may result in shorter delivery times. Furthermore, it leads to lower shortages, inventories and order variability, see Section 5. It is not clear, however, how capacity should be shared evenly given a general flexibility configuration A. We propose the following method. Let $r_i$ be the number of plants that can produce product i. Given each product's mean demand, $\bar{D}_i$, i=1,2,…,m, the target production levels, $t_{ij}$, for configuration A leading to maximum coverage of the mean demands can be approximated by solving the following *Capacity Allocation Problem*:

$$Min \quad \sum_{i=1}^{m} l_i + \varepsilon \left[ \sum_{(i,j)\in A} (t_{ij} - \frac{\bar{D}_i}{r_i})^2 \right]$$

$$S.t.$$

$$\sum_{j:\ (i,j)\in A} t_{ij} + l_i = \bar{D}_i \quad \forall i = 1,\dots,m,$$

$$\sum_{i:\ (i,j)\in A} t_{ij} \le C_j \quad \forall j = 1,\dots,n,$$

$$l_i \ge 0 \quad \forall i = 1,\dots,m$$

$$t_{ij} \ge 0 \quad \forall i = 1,\dots,m,\ j = 1,2,\dots n$$

As in the quadratic program in Section 4.1, $\varepsilon$ is chosen to be very small so that the quadratic program minimizes total shortage, while selecting production quantities as distributed as possible. The quadratic program ensures that plant capacities and demands will be evenly allocated (this can be shown through the KKT optimality conditions, see Muriel et al. (2001)).

This allocation provides a set of initial target production levels to start the simulation. These levels have been computed using only average demand information, fully disregarding period-to-period demand variability. As a result, they do not properly estimate the actual average production levels that will be observed in the system when using the production planning policy described in the previous section. To get more realistic target production levels, we simulate the system over a number of periods and calculate the actual production solving the Production Planning Problem with the initial target production levels and no inventory constraints; see Phase II of the simulation below. This initial simulation also provides an estimate of the distribution of production levels of each product in each plant, which is essential in determining an appropriate inventory policy.

### 4.3 Inventory Policy

Demand for each component is determined by the production of the final product and thus constrained by plant capacity. As a result, the total inventory (on-hand plus on-order) of components of product i at plant j should never exceed the total plant capacity during the lead-time, since any amount beyond that will never be used. To impose this additional condition, Bish et al. (2001) introduce a *modified base-stock inventory policy* that raises system inventory every period to:

$$L\mu_{ij} + \min\{z\sqrt{L}\sigma_{ij}, L(C - \mu_{ij})\}$$

where $\mu_{ij}$ and $\sigma_{ij}$ are the mean and standard deviation of the production level of product i at plant j, respectively, L is the component supply lead time (typically a few production periods) and z is a safety factor corresponding to a specified stock-out probability for normally distributed demands. Alternatively, the base-stock level could be determined more accurately using the actual distribution of production levels (i.e. demand for components). However, probability plotting shows the normal distribution to be a good approximation in our simulation tests.

The modified base-stock policy reflects the capacity constraint on the safety stock of components, thereby avoiding the wasteful carrying of more inventory than can possibly be used during the lead time. Observe that this inventory policy is not necessarily optimal and is considered mainly for its simplicity. Nonetheless, we believe that the conclusions obtained using the base-stock policy are fairly general because the amount of inventory needed is simply driven by the production variability. We report inventory because it provides a simple overall system performance measure, whereas variability needs to be measured for each plant and product and cannot be aggregated into a single quantity.

### 4.4 Simulation Scheme

The simulation model has three phases, each representing a stage of the decision framework described in Figure 1.

**Phase One** or *Initialization Phase* accepts the strategic decisions on flexibility configuration and plant capacities as inputs. The simulator generates a number (5000 in our study) of possible realizations, $\overline{D}_1, \overline{D}_2, ..., \overline{D}_m$, of the mean demand for each product using forecast demand information. In our study, the forecast of mean product demands is of 100 units for each product and the forecast errors are independent and identically normally distributed with a standard deviation of 40; this is based on an observed average forecast error of 40%. The mean demands are truncated to 2-sigma limits of [20,180].

**For each of the generated mean demand realizations, $\overline{D}_1, \overline{D}_2, ..., \overline{D}_m$,** the following phases determine the production planning and inventory policies at the tactical stage (Phase 2) and then simulate the system in operation, as demands are observed, under those policies (Phase 3).

**Phase Two** or *Policy Determination Phase* computes the target production levels and production variability to determine the production planning and component inventory policies. It first calculates the initial target levels by solving the Capacity Allocation Problem. Then it simulates 1000 periods of system operation assuming a normal distribution of product demands with means, $\overline{\mathrm{D}}_1, \overline{\mathrm{D}}_2, ..., \overline{\mathrm{D}}_{\mathrm{m}}$, and coefficient of variation 0.15, and calculates the associated production levels by solving the Production Planning Problem with the initial target production levels and no inventory constraints, as described in Section 4.2. The average and variability of the production levels of each product at each plant over the 1000 periods simulated provides estimates of $\mu_{ij}$ and $\sigma_{ij}$ for each product i and plant j such that $(i, j) \in A$. These are the only parameters needed to determine the production planning and inventory policies.

**Phase Three** or *Evaluation Phase* simulates the operational stage over a number of periods (1000 in our study). It generates customer demands**,** $d_i^t$, i=1,2,…,m**,** for each period t, and calculates the associated production levels and component orders, using the production planning and inventory policies determined in Phase Two. The outputs of this stage are the system performance metrics: sales, component inventory, and variability induced at the supplier level associated with the mean demand realizations under consideration**.** Inventory is carried over time from period to period and updated as orders are received and demand satisfied each period. We report the average inventory in each period, i.e. (initial period inventory + final period inventory)/2.

This simulation model allows us to quantify system performance for each given capacity and flexibility configuration. The resulting information can then be used in a cost-benefit analysis of manufacturing flexibility at the capacity investment stage that considers not only the capital outlay versus sales benefits, but also the expected operational costs in the supply chain.

### 4.5. Simulation Results and Analysis

In this section, we use the simulation tool described above to analyze the impact of flexibility on production system performance for different product-plant configurations.

As in section 3, we consider production systems in which there are the same number, N, of products as plants, and study the performance of h-type flexibility configurations as h grows. To specify the configuration under consideration we use the following notation: XN-h, where X= {B, PF, FF} to denote whether it is a basic dedicated, a partially flexible or a fully flexible configuration, N is the number of products and plants, and h the number of products built in each plant.

The coefficient of variation of period demand in Phases Two and Three is set to 0.15. In the modified *base-stock* inventory policy, the safety factor is set to 1.64. Each product *i* requires a unique major component, component *i*, that is single sourced and has a lead-time L of 2 periods. The restriction to unique components is of particular interest to the automotive industry because they account for those with higher costs and lead times. The demand that cannot be met due to capacity constraints is considered as lost sales.

The total sales, total inventory and variability induced at suppliers in the system for different flexibility configurations in 3-product 3-plant, the 5-product 5-plant and the 10-product 10-plant systems are presented in Table 1. For more detailed results on the performance of each of the cases simulated, with different mean product demands in the medium term, we refer the reader to Muriel et al. (2001), where we study the effect of flexibility as a function of total system demand and product imbalance.

| Configuration | Total Sales | System Inventory | Order SD |
|---|---|---|---|
| B3-1 | 248.05 | 155.41 | 5.22 |
| PF3-2 | 269.65 | 198.00 | 7.73 |
| FF3-3 | 269.74 | 190.57 | 7.61 |
| B5-1 | 421.84 | 263.94 | 5.31 |
| PF5-2 | 465.52 | 367.90 | 8.58 |
| PF5-3 | 466.20 | 343.81 | 8.58 |
| PF5-4 | 466.19 | 336.60 | 8.35 |
| FF5-5 | 466.15 | 334.03 | 8.32 |
| B10-1 | 843.23 | 529.34 | 5.37 |
| PF10-2 | 941.02 | 817.51 | 9.28 |
| PF10-5 | 947.62 | 713.78 | 9.34 |
| FF10-10 | 947.44 | 697.77 | 9.22 |

**Table 2: Total sales, system inventory, and standard deviation of upstream orders for the different flexibility configurations in the 3-product 3-plant, 5-product 5-plant and 10-product 10-plant cases**

The above simulation results provide the following insights:

1. Partial flexibility provides most of the sales benefits of flexibility, as shown by Jordan and Graves (1995). However, adding more flexibility is beneficial in reducing the overall inventory in the system and the demand variability observed upstream in the supply chain. The more the flexibility in the system, the closer the actual production levels in the final allocation are to the target production levels. Therefore, more flexibility results in lower production variability, leading to better system performance.

2. As the number of plants and products grows, flexibility can be used to build longer chains and reap higher sales benefits. This is accompanied by a corresponding increase in required component inventories. However, as more flexibility is added to the system, the consequent reduction in inventory is more pronounced for these longer chains.

Interestingly, while total sales remain almost the same as more flexibility is added to the system, *production variability is reduced!* The longer the product-plant chain, the greater the reduction. Thus, although a limited flexibility investment following a chain configuration extracts most of the benefits in sales, it may lead to higher operational costs due to the increased production variability in the system. If component inventories are expensive, the cost of additional flexibility may be off-set by the savings in reduced inventories and more stable upstream orders.

### 4.5.1 Sensitivity to Different Levels of Capacity

In the cases studied so far, system capacity was set to the expected mean demand at the investment stage. In what follows, we investigate the sensitivity of these results to the amount of capacity installed relative to expected demand. For that purpose, we study the effect of varying plant capacities over a range of values (50, 75, 90, 100, 110, 125, 150), while demands remain as in our original scenario, in the 3-product 3-plant case.

Figure 3 displays the total sales and total inventory plotted against the capacity utilization for the non-flexible base case and the partially flexible configuration for the different capacities considered. This extends the results of Jordan and Graves (1995) by adding the impact of flexibility on inventory. The trends for the base and the partially flexible configurations are very similar. The dotted lines show the impact of changing flexibility on sales and inventory at a given capacity level and the solid lines show the impact of changing capacity on sales and inventory at a given flexibility level. Table 3 shows the resulting percent increase in sales and inventory as system capacity changes.

If system capacity is either very tight or very loose, i.e. total capacity deviates significantly from the expected system demand, flexibility has limited value in increasing sales. Nonetheless, flexibility is much more beneficial when capacity is tight, since the increase in sales comes at the expense of only a slight increase in inventory. In fact, in the cases where the mean demand realizations are high for some products and low for others, the better distribution of the scarce capacity to the different products in the flexible environment leads to a reduction in production variability. When capacity is loose, however, a modest increase in sales requires a steep increase in component inventories. Observe that the benefits of flexibility in increased sales are the highest when system capacity is slightly above the expected demand, but this increase comes at the expense of a disproportionate increase in the amount of component inventory in the system.

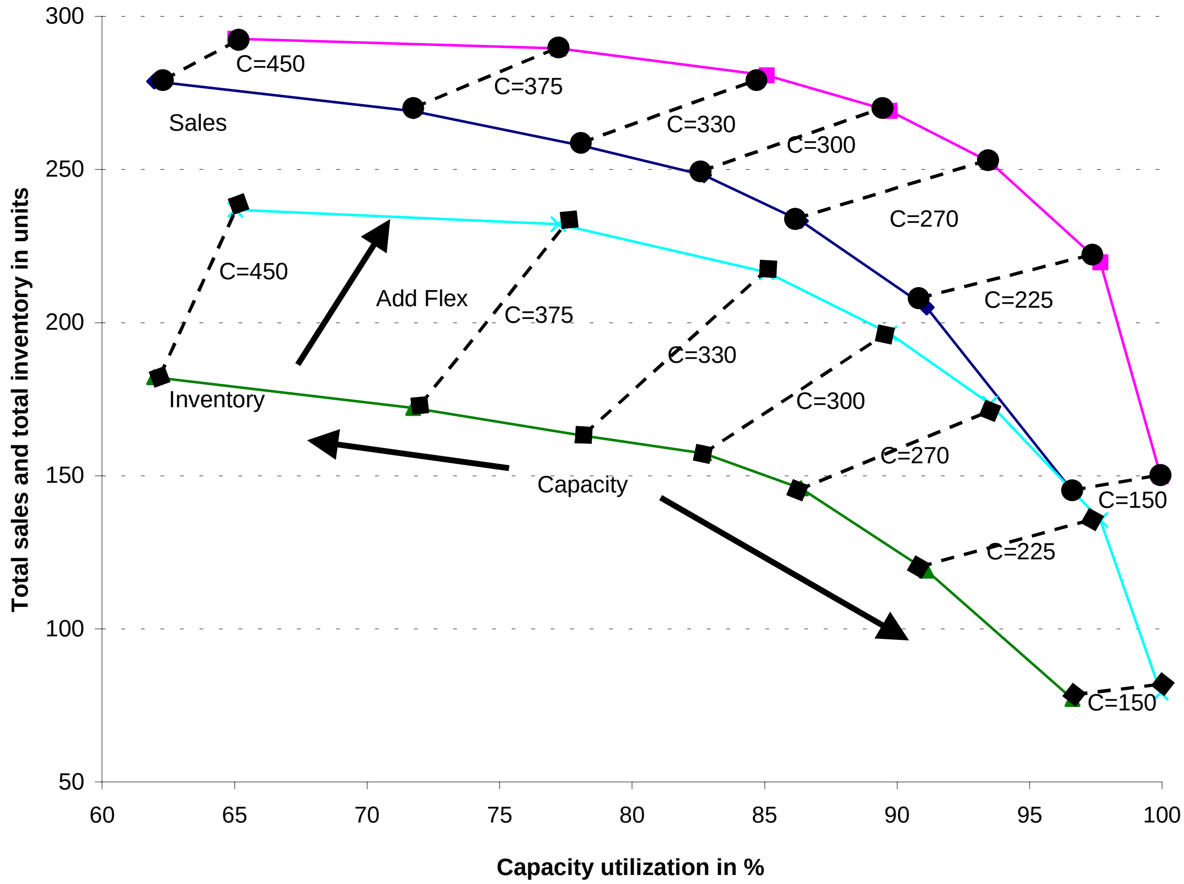


**Figure 3: Impact of Capacity changes on the benefits of flexibility**

| Total Capacity | Percentage increase | |
|---|---|---|
| | Sales | Inventory |
| **150** | 3.46 | 2.74 |
| **225** | 7.20 | 13.85 |
| **270** | 8.26 | 18.88 |
| **300** | 8.71 | 27.38 |
| **330** | 8.91 | 32.67 |
| **375** | 7.63 | 34.84 |
| **450** | 4.96 | 30.01 |

**Table 3: Percentage increase in total sales and inventory for different total capacities with respect to the base (non-flexible) case.**

In order to achieve a certain sales target, flexibility can be used as a substitute to increased capacity. For instance, the total sales in the non-flexible configuration with a total capacity of 270 units is 233 units. In order to increase sales to 248 units, one could either increase capacity to 300 units or incorporate partial flexibility into the manufacturing system, as in PF3-2. The resulting component inventory levels under these two strategies, however, are very different. In our example, the inventory resulting from adding flexibility is 173 units, whereas the inventory associated with the capacity increase is only 157 units. Thus, the cost of increasing capacity can be traded off against the cost of adding flexibility and the associated increase in production variability.

Similar studies have been carried out to test the sensitivity of our results to varying levels of forecast error at the investment stage, different levels of the desired safety factor in the inventory calculations, the requirement of multiple unique components with varying supply lead times[2], and systems with different number of plants and products; see Muriel et al. (2001). These studies support our previous conclusions.

### 4.5.2 Sensitivity to Different Levels of Safety Stock

The previous simulation results show surprisingly high inventory levels in the flexible configurations, as compared to the dedicated ones. The objective of this section is to analyze the effect of lower safety stock levels on the total sales and the total inventory in the system.

---

[2] The percent increase in both sales and inventory in the flexible environment are slightly higher when considering multiple components.

In order to reduce inventories in the system[3], we will increase the weight that we place on deviations from target production. More specifically, we increase the value of ε in the production planning program in Phase 2 of the simulation, which is used to estimate the variability of production levels. Table 4 reports the sensitivity of the percent increase in sales and inventories in the system to the value of the parameter ε. We observe that for values of $\varepsilon \leq 1/100$ the results stabilize.

| | B3-1 | | PF3-2 | | FF3-3 | |
|---|---|---|---|---|---|---|
| **ε** | **Sales** | **Inventory** | **Sales** | **Inventory** | **Sales** | **Inventory** |
| 0.2 | 247.52 | 144.33 | 267.56 | 173.06 | 268.93 | 178.31 |
| 0.1 | 248.06 | 147.95 | 268.53 | 181.33 | 269.54 | 186.42 |
| 0.01 | 248.67 | 155.42 | 269.66 | 197.43 | 269.75 | 190.53 |
| 0.001 | 248.67 | 155.42 | 269.66 | 197.52 | 269.75 | 190.53 |

**Table 4: Sensitivity of system sales and inventory to changes in the variability weight parameter ε.**

These results show that inventory can be reduced without a significant impact on sales. While sales go down by 0.8% in PF3-2, inventory is reduced by over 12% as we increased the weight factor. For the fully flexible configuration, the percent reductions in sales and inventory are 0.3% and 6.4%, respectively. As a result, both sales and inventory are higher in the fully flexible versus partially flexible configuration for low values of ε. We observe, however, that to achieve the same level of sales, the partially flexible allocation requires higher stock levels.

[3] Another technique to reduce inventory in the system, is simply to lower the safety factor, z, used in the computation of safety stocks. This yields similar results.

The analysis above also allows managers to pick the level of ε that best fits their needs given the observed trade-off between sales and inventory.

## 5. Tactical Capacity Allocation Policy Study

Throughout the paper, we have considered that target production levels were determined so as to evenly distribute the production of each product among all the plants that could manufacture it. In this section, we justify this tactical allocation policy and show that it results in better performance.

### 5.1 Example

To illustrate this, consider a situation identical to that in section 2, except that the likely scenarios are as follows: in scenario 1 product demands are $(d_1,d_2,d_3) = (C,C,C)$ and in scenario 2 product demands are $(d_1,d_2,d_3) = (0,C, C)$. These scenarios have the same probability of occurrence. In addition, an occasional spur in demand of product one to 2C units may occur, while the demand for the other two is cannibalized and cut in half; that is, demands of $(d_1,d_2,d_3) = (2C,C/2, C/2)$ occur with very low probability. The system is managed to ensure that no shortages occur in the two likely scenarios only, since the low probability event would require a significant increase in component 1 inventories. Let the lead time for component inventory replenishment be L=2.

Consider the partial flexibility configuration in Figure 1. To ensure no component 1 shortages under scenarios 1 and 2 *over the lead time*, the distributed allocation system must always order to bring the inventory position (on hand plus on order) of component 1 at each of the plants that can build it up to C. Thus, if the spike in demand comes after a low demand period the full 2C units will be satisfied. Any other policy will assign $C/2<X<C$ units of production of

product 1 to one of the plants, say plant 1, and C-X to the second plant to ensure that demand is satisfied in scenario 1. Then, plant 1 will bring its component 1 inventory position up to 2X>C, while plant 2 will bring it up to 2(C-X) each period. As a result, the maximum production of product 1 in any period will be C+2(C-X) < 2C. This is due to inventory misplacement: after a low demand period, 2X units of component 1 will be on hand in plant 1 while the plant can only process C; at the same time, there is capacity available at plant 2, but it will go idle due to the lack of sufficient component 1 inventory there. Consequently, the distributed policy will lead to higher sales and lower component inventories.

### 5.2 Analytical Treatment

In the fully flexible case, it is easy to see that the distributed policy outperforms all others with the same amount of system inventory for each component.

**Result C:** *Consider a fully flexible production system with n plants and m products. Let the on-hand system inventory of component i be $I_i<nC$, for* i=1,2,…m, *then*

*(a) The distributed policy will have $I_i/n$ components of product i in each plant and will be able to produce a total of* $S_1 \equiv \min\{nC, \sum_{i=1}^{m} \min\{d_{i,} I_i\}\}$.

(b) *Any other policy will have $I_{ij} \in [0, I_i]$ components of product i at plant j and thus the amount of demand satisfied is <u>at most</u>* $S_2 \equiv \min\{nC, \sum_{i=1}^{m} \min\{d_{i,} \sum_{j=1}^{n} \min\{C_, I_{ij}\}\} \leq S_1$.

Allocating resources evenly to the different plants is effective in making the multi-plant system perform as a single-plant multi-product system. In a fully flexible multi-plant system where each plant is allocated the same share of demand and the same amount of components, the quantities produced and in inventory are identical to those in a single plant with the same total capacity and component inventories. When using other policies, however, it is possible to forgo sales because of inventory misplacement. As shown in the example above, there may be more component inventory for a particular product than capacity available in its primary plant, while there is available capacity but not enough inventory to satisfy the excess demand in the secondary plants. This leads to lost sales and higher component inventories carried to the next period.

### 5.3 Generality of Results Validated by Simulation

Although the example above may seem contrived, the concept it illustrates is general. We have observed it repeatedly. In fact, in our computational study, a prioritized capacity allocation policy where each product is predominantly assigned to a single plant[4] in the calculation of initial target production level in Phase 2, always resulted in slightly lower sales and significantly higher inventory levels and order variability. See Table 5 for the results for the case of the 3-product 3-plant system studied in Section 4.5. Similar results were found in all our tests.

---

[4] The prioritized allocation may be attractive if there are significant economies of scale and/or learning effects to be gained by concentrating production. These two policies, *prioritized and distributed*, represent extreme cases: pushing as much as possible of the production of an item to a single plant versus distributing it equally among all the plants that can manufacture it. Thus, the study and comparison of their performance provides general insight and understanding of the performance of all policies. Recall that this allocation is used to determine the target production levels $\mu_{ij}$ that are inputs to the Production Planning Problem in Section 4.1.

| Configuration | Allocation Policy | Total Sales | System Inventory | Order SD |
|---|---|---|---|---|
| **PF3-2** | **Prior** | 8.65 | 35.46 | 56.71 |
| | **Dist** | 8.71 | 27.39 | 48.01 |
| **FF3-3** | **Prior** | 8.69 | 27.66 | 52.18 |
| | **Dist** | 8.75 | 22.61 | 45.66 |

**Table 5: Average percent increase in total sales, system inventory, and standard deviation of upstream orders for the different flexibility configurations in the 3-product 3-plant case under both initial capacity allocation policies, prioritized and distributed; the increase is with respect to the base (non-flexible) case.**

## 6. Conclusions

The major conceptual contributions of this paper are: (1) identifying that *partial* flexibility can lead to a significant increase in production variability, and (2) showing that distributed tactical capacity allocation policies, which evenly allocate demand to the plants that can satisfy it, lead to better performance of the flexible system. Stylized examples and analytical models provide a basic understanding of these effects. The optimization-based simulation tool developed allows us to evaluate them in more general settings.

Our results show that while most of the benefits in increased sales due to flexibility can be achieved with a modest increase in flexibility (just having each product built in two plants, see Jordan and Graves 1995), the component inventory required is significantly reduced as more flexibility is added (that is, as each product is built in a larger number of plants). The higher production variability due to insufficient flexibility leads to higher component inventories and induced order variability at the supplier level. Thus, if component inventories are expensive, additional flexibility may be cost-effective. As in Graves and Tomlin (2000), extra flexibility pays off when further system effects are added to the basic Jordan and Graves model.

Our simulation results show that the effect of partial flexibility on production variability is more pronounced for systems with a large number of plants. If system capacity is either very tight or very loose, flexibility has limited value in increasing sales. Nonetheless, flexibility is more beneficial when capacity is tight, since the increase in sales comes at the expense of only a slight increase in inventory. In fact, inventory requirements may actually decrease with flexibility for tightly capacitated scenarios as the scarce capacity can be better matched to the expected product demands (see also Bish et al. (2001)). When capacity is loose, however, a modest increase in sales requires a steep increase in component inventories.

Our study also indicates that system performance is highly dependent on the tactical capacity allocation policy implemented. Distributed allocation policies yield consistently higher sales, lower inventories and reduced order variability observed by suppliers. These policies avoid inventory misplacement, i.e. situations where there is spare capacity but no available component inventory at one plant, while there are available components but no remaining capacity in another plant, for a particular product.

These insights, as well as the simulation tool, can be used by practitioners to guide the design of their production systems, trading off the initial capital outlay on flexible capacity versus the sales benefits and expected operational system costs.

## Acknowledgements

This work has been partially supported by the DMII-MES division of the National Science Foundation. We would like to thank Stephan Biller, Jeff Tew, Ted Costy, and Lynn Truss of the GM R&D Center for their continuous support of this research, feedback, and suggestions. We are also indebted to Professor Ebru Bish for helpful discussions in the early stages of the work, to the editors and the two anonymous referees for their insightful reviews, and to Suzanne Sposato and Carolina Soto for their help with the simulation analysis under an NSF-REU.

## References

BEACH, R., A. P. MUHLEMANN, D. H. R. PRICE, A. PATERSON, and J. A. SHARP. 2000. "A Review of Manufacturing Flexibility". European Journal of Operational Research 122, 41-57.

BISH, E., A. MURIEL, and S. BILLER. 2001. "Managing Flexibility in a Make-to-order Environment". Technical report, University of Massachusetts Amherst.

DE GROOTE, X. 1994. "The Flexibility of Production Processes: A General Framework." Management Science 40:7, 933-945.

DE TONI, A. and S. TONCHIA. 1998. "Manufacturing Flexibility: a literature review". International Journal of Production Research 36:6, 1587-1617.

GRAVES, S. C. and B. T. TOMLIN. 2000. "Process Flexibility in Supply Chains". Working paper, Massachusetts Institute of Technology.

JORDAN, W. C. and S. C. GRAVES. 1995. "Principles and Benefits of Manufacturing Process Flexibility". Management Science 41:4, 577-594.

KOUVELIS, P. 1992. "Design and Planning Problems in Flexible Manufacturing Systems: A Critical Review". Journal of Intelligent Manufacturing 3, 75-99.

MURIEL, A., A. SOMASUNDARAM and Y. ZHANG. 2001. "Optimization-Based Simulation Study of the Impact of Manufacturing Flexibility on Supply-Chain Performance". Technical report, University of Massachusetts.

SETHI, A. K. and S. P. SETHI. 1990. "Flexibility in Manufacturing: A Survey". The International Journal of Flexible Manufacturing Systems 2, 289-328.

SIMPSON, K. F. 1958. "In-Process Inventories." Operations Research 6, 863-873.

VAN MIEGHEM, J. A. and N. RUDI. 2002. Newsvendor Networks: Inventory Management and Capacity Investment with Discretionary Activities. Manufacturing and Service Operations Management 4, 313-335.

ZIPKIN, P. H. 2000. Foundations of Inventory Management. McGraw-Hill.