Design for Process Flexibility: Efficiency of the Long Chain and Sparse Structure

Mabel C. Chou
NUS Business School, National University of Singapore, Singapore, bizchoum@nus.edu.sg

Geoffrey A. Chua
Nanyang Business School, Nanyang Technological University, Singapore, geoff.chua@gmail.com

Chung-Piaw Teo
NUS Business School, National University of Singapore, Singapore, bizteocp@nus.edu.sg

Huan Zheng
Antai College of Economics and Management, Shanghai Jiaotong University, Shanghai, China, zhenghuan@sjtu.edu.cn

The concept of chaining, or in more general terms, sparse process structure, has been extremely influential in the process flexibility area, with many large automakers already making this the cornerstone of their business strategies to remain competitive in the industry. The effectiveness of the process strategy, using chains or other sparse structures, has been validated in numerous empirical studies. However, to the best of our knowledge, there have been relatively few concrete analytical results on the performance of such strategies vis-á-vis the full flexibility system, especially when the system size is large or when the demand and supply are asymmetrical. This paper is an attempt to bridge this gap.

We study the problem from two angles: (1) For the symmetrical system where the (mean) demand and plant capacity are balanced and identical, we utilize the concept of a generalized random walk to evaluate the asymptotic performance of the chaining structure in this environment. We show that a simple chaining structure performs surprisingly well for a variety of realistic demand distributions, even when the system size is large. (2) For the more general problem, we identify a class of conditions under which only a sparse flexible structure is needed so that the expected performance is already within ϵ optimality of the full flexibility system.

Our approach provides a theoretical justification for the widely held maxim: In many practical situations, adding a small number of links to the process flexibility structure can significantly enhance the ability of the system to match (fixed) production capacity with (random) demand.

Subject classifications: random walk; stochastic programming; production: flexible manufacturing; facility planning: design.
Area of review: Optimization.
History: Received October 2007; revisions received March 2008, June 2008; accepted July 2008. Published online in *Articles in Advance* March 30, 2009.

1. Introduction

Recent trends in consumer markets have shown a shift toward more customized products and faster upgrades in technology. This has resulted in more product lines, shorter product life cycles, and higher demand variability. For example, the number of car models offered in the United States market has increased from 195 (in 1984), to 238 (in 1994), to 282 (in 2004), and is projected to reach 330 by 2008. The number of plants, however, has remained stable over those years, bringing down the average annual units sold per nameplate from 106,819 (in 1985) to 48,626 (in 2005) (cf. U.S. Department of Commerce 2006, Van Biesebroeck 2007).

Facing such an increased demand uncertainty as well as heightened market competition, businesses can no longer rely on capacity, pricing, quality, and timeliness alone as competitive strategies. In particular, firms need to employ process flexibility to improve their ability to match supply with uncertain demand. In an interview with the *Wall Street Journal* (Boudette 2006, p. 5), Chrysler Group CEO

Thomas LaSorda disclosed that flexible production "gives us a wider margin of error." With regard to the value of process flexibility, he said, "if the Caliber doesn't sell well, the Jeep Compass and Patriot could take up capacity, and eventually a fourth model will be built, too."

This recent focus on flexibility as a competitive strategy can be observed in major manufacturing industries, such as the above mentioned automobile industry (Wall 2003), the textile/apparel industry (DesMarteau 1999), and the semiconductor/electronics industry (McCutcheon 2004). Moreover, the value of flexibility extends to service industries, where firms have increasingly employed cross-trained workers to provide more flexible services. From here onward, we adopt the term "process flexibility" from the literature (Jordan and Graves 1995, Sethi and Sethi 1990) to denote "a firm's ability to provide varying goods or services, using different facilities or resources."

One of the most important ideas in this literature is the concept of a simple "chaining" strategy. Here, a plant capable of producing a small number of products, but with

proper choice of the process structure (i.e., plant-product linkages), can achieve nearly as much benefit as the full flexibility system (where each plant is equipped to produce all the products). This concept is widely believed to be true and has been applied successfully in many industries. For example, Chrysler CEO LaSorda has repeatedly mentioned the importance of chaining in his interviews and speeches (LaSorda 2006), while VP Frank Ewasyshyn was recently inducted into the Shingo Prize Academy for his contributions to flexibility and efficiency (*Assembly Magazine* 2005).

To illustrate the value of process flexibility and the efficiency of chaining relative to full flexibility, we consider the following n-plant, n-product example. As in existing literature, we use a bipartite graph to represent flexibility structures. On the left is a set A_n of n product nodes, while on the right is a set B_n of n facility nodes. A link connecting product node i to facility node j means that facility j is endowed with the capability to produce product i. We let $\mathcal{G} \subseteq A_n \times B_n$ denote the set of all such links; that is, the edge set of the bipartite graph. Hence, each flexibility configuration can be uniquely represented by a bipartite graph $\mathcal{G}$. Three special configurations we will frequently refer to are:

1. The dedicated system:

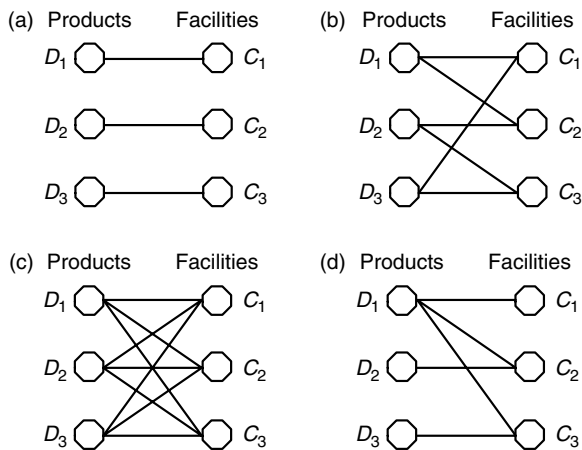$$\mathcal{D}(n) = \left\{ (i, i) \mid i \in \{1, 2, \ldots, n\} \right\}.$$

2. The chaining system[1]:

$$\mathcal{C}(n) = \left\{ (i, i) \mid i = 1, 2, \ldots, n \right\}$$
$$\cup \left\{ (1, 2), (2, 3), \ldots, (n-1, n), (n, 1) \right\}.$$

3. The full flexibility system: $\mathcal{F}(n) = A_n \times B_n$.

Figure 1 shows some examples of flexibility configurations for a three-facility, three-product system. Graphs (a), (b), and (c) are the three respective special configurations as listed above for the case $n = 3$.

Figure 1. Bipartite graph representation of 3×3 flexibility structures.

Assume that each plant has a capacity of $C_j = 100$ units for each j, and each product consumes one unit of capacity and has an expected demand of $D_i = 100$ units for each i. Note that the (mean) demand and supply are balanced and identical in this case. We assume further that the demand is normally distributed with a standard deviation of 33 units (so that the probability of negative demand is negligible).

Let $\mathbf{D} = (D_1, \ldots, D_n)$ denote the random realization of the demand. We simulate the expected system sales for the three special structures. We assume that we first observe the demand realizations, and then we determine how much of each product is produced by each plant to maximize total sales. This boils down to solving the following maximum-flow problem on n supply and n demand nodes, with process structure $\mathcal{G}$:

$$Z_{\mathcal{G}}^*(\mathbf{D}) = \max \ \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij}$$

$$\text{s.t.} \ \sum_{j=1}^{n} x_{ij} \leqslant D_i \quad \forall i = 1, 2, \ldots n, \tag{1}$$

$$\sum_{i=1}^{n} x_{ij} \leqslant C_j \quad \forall j = 1, 2, \ldots n, \tag{2}$$

$$x_{ij} \geqslant 0 \quad \forall \ i, j = 1, \ldots, n, \tag{3}$$

$$x_{ij} = 0 \quad \forall \ (i, j) \notin \mathcal{G}. \tag{4}$$

We solve the above max-flow problem for each random realization of $\mathbf{D}$. For small n (say $n = 10$), our simulation shows that the expected sales in the dedicated, chaining, and full flexibility systems are 864.47, 949.36, and 955.14, respectively. This demonstrates that chaining already achieves most (99.39%) of the benefits of full flexibility in this case. A natural question is how well the chaining structure performs as n increases to infinity. Surprisingly, with the demand distribution given as before, we will show in this paper that even as n approaches infinity, the chaining system can still accrue close to 97% of the expected sales in the full flexibility system! This is important because the chaining system requires many fewer links in its process structure, and thus is much cheaper than the full flexibility system when n is large, yet is able to accrue most of the expected sales in the full flexibility system. This illustrates that, with proper choice of the process structure, one can achieve nearly as much benefit as full flexibility with a sparse process structure.

However, the above results were obtained under the condition that demands are independent. If demands are correlated, the situation can be very different, as illustrated in the following. Consider the other extreme situation, when demands are correlated, with $\sum_{i=1}^{n} D_i = n$, and

$$C_j = 1,$$

$$D_i = \begin{cases} n, & \text{with probability } \dfrac{1}{n} \\ 0, & \text{otherwise} \end{cases} \quad \forall i, j = 1 \ldots, n. \tag{5}$$

In the full flexibility system, it is clear that the system can satisfy all demand using available capacity, and

$E_{\mathbf{D}}(Z^*_{\mathcal{F}(n)}(\mathbf{D})) = n$. On the other hand, suppose that we use a sparse process structure with $O(n)$ arcs, say δ_i arcs, to serve demand for product i, with $\sum_{i=1}^{n} \delta_i = O(n)$. The expected demand served for product i is therefore δ_i/n because each facility has a capacity of 1, and there are δ_i such facilities serving product i. It is thus clear that the sparse system can support an expected flow of

$$\frac{1}{n}\sum_{i=1}^{n}\delta_i = O(1) \ll E_{\mathbf{D}}(Z^*_{\mathcal{F}(n)}(\mathbf{D})).$$

Therefore, unlike the earlier situation, there is a significant loss of optimality if we are restricted to using a sparse process structure (like the chaining structure) in our problem.

Our objective in this paper is to analyze and understand the performance of a sparse process structure, in particular the chaining structure, under different conditions. We address the following problems:

1. What is the performance of the chaining structure as the system size increases, when demand and supply are balanced and identical? We seek a method that can evaluate the efficiency of chaining in this environment.

2. What are the appropriate conditions so that a sparse structure can perform nearly as well as the full flexibility system? An intuitive definition of a good process structure is difficult in the general case. Hence, we focus instead on general structures with a sparse subset of links.[2]

To address the first question, we use the concept of a generalized random walk to evaluate the asymptotic performance of the chaining structure in the environment with identical and balanced demand and supply. Our results show that the performance of the 2-chain is related to properties (overshoots and stopping times) of the random walk model, and we can evaluate the exact asymptotic performance of the 2-chain for a variety of realistic demand distributions. For uniform and normal demands, our method shows that the chaining strategy already reaps a substantial portion (at least 89% and 96%, respectively) of the benefits of the full flexibility system, even when n approaches infinity.

We address the second question in a more general setting. Consider a problem of the form

(P): $\quad Z(b, \{1, \ldots, n\})$

$$= \max\left\{\sum_{i=1}^{n} c_i x_i \colon Ax \leqslant b; \ x_i \geqslant 0, \ i = 1, \ldots, n\right\},$$

where $c_i \geqslant 0$ ($\forall i = 1, \ldots, n$), A is an $m \times n$ nonnegative matrix, b is random and nonnegative, and $n \gg m$. When b is deterministic, it is well known that there is an optimal solution x^* for (P) with support in at most m decision variables (i.e., no more than m variables in x^* are positive while other variables are all zero). There is no loss of optimality if the other variables are discarded from problem (P). Let $S \subseteq \{1, \ldots, n\}$ and

$$Z(b, S) = \max\left\{\sum_{i \in S} c_i x_i \colon A_S x_S \leqslant b; \ x_i \geqslant 0, \ i \in S\right\},$$

where A_S (respectively, x_S) denote the columns (respectively, rows) of A (respectively, x) indexed by the subset S. Our goal is to identify a subset S so that

$$E_b(Z(b, S)) \geqslant (1 - \epsilon)E_b\big(Z(b, \{1, \ldots, n\})\big)$$

and

$$|S| \approx O(m) \ll n.$$

We need to identify and impose further structural conditions on the random parameters b (or $\mathbf{D}$ for the max-flow problem in $Z^*_{\mathcal{G}}(\mathbf{D})$) to ensure the existence of a good sparse support set in our problem (or sparse process structure in the context of the flexibility structure design problem). Note that the optimal solution for $Z^*_{\mathcal{G}}(\mathbf{D})$ in the earlier example (5) has the following property: for any $i, j \in \{1, \ldots, n\}$,

$$x^*_{ij} = \begin{cases} 1, & \text{with probability } \dfrac{1}{n}, \\ 0, & \text{otherwise}, \end{cases}$$

and hence $E_{\mathbf{D}}(x^*_{ij}) = 1/n$. However, $x^*_{ij} = 1$ for some demand realization $\mathbf{D}$, which is much larger than the expected value. It turns out that the difference between the expected optimal flow and the actual optimal flow (for some demand realization) is a crucial property we need to control for this problem. By ruling out such instances, we can prove the existence of a good sparse support set for the optimization problem.

2. Literature Review

In the operations management literature, there are two main streams of research related to flexibility. The first stream examines the trade-off between flexible and dedicated resources. Fine and Freund (1990) characterize the optimal investment in flexibility for a price-setting firm, where demand is modeled by a discrete probability distribution of k possible states that affect demand. Van Mieghem (1998) takes a critical-fractile approach to solving the optimal flexibility investment for a price-taking firm, but for any arbitrary multivariate demand distribution. Bish and Wang (2004) extend van Mieghem's work to a price-setting firm facing different types of correlated demands.

The above studies, though, focus only on full flexibility; that is, all facilities can produce all products. Unfortunately, in practice, the acquisition cost of full flexibility is usually too enormous to permit the recovery of adequate benefits. In response, a second stream of research looks at different degrees of flexibility, and examines the value of these types of process flexibility. The landmark study was by Jordan and Graves (1995), who introduced the concepts of *smart limited flexibility* and *chaining*. They observe, through extensive simulation, that limited flexibility, configured the right way, yields most of the benefits (in terms of expected sales) of full flexibility. Furthermore, they claim that limited flexibility has the greatest benefits when a *chaining* strategy

is used, whereby every facility can produce two products and every product can be produced by two facilities, in a way that "chains" all the facilities and products. Specifically, their numerical results show that for a 10-facility, 10-product example, chaining achieves about 95% of the benefits of full flexibility for the class of demand distribution used in their study. Additionally, in the second part of the paper, they propose a measure (later called the JG index) for evaluating different flexibility structures.

Because the twin ideas of smart limited flexibility and chaining have been well received, significant extensions arose in different directions. For example, Graves and Tomlin (2003) extend this work to multistage supply chains, Gurumurthi and Benjaafar (2004) to queuing systems, and Hopp et al. (2004) to flexible workforce scheduling. Recently, Iravani et al. (2005) proposed a new perspective on flexibility using the concept of "structural flexibility." They introduce new flexibility indices and show their applicability to serial, parallel, open, and closed networks. Bish et al. (2005) go beyond just matching supply and demand as they study the impact of flexibility on the supply chain. They show that in a 2×2 system, certain practices that might seem reasonable in a flexible system can result in greater production swings and higher component inventory levels, which will then increase operational costs and reduce profits. To account for partial flexibility, Muriel et al. (2006) extend the work of Bish et al. (2005) to larger systems. Brusco and Johns (1998) present an integer linear programming model to evaluate different cross-training configurations in a workforce staffing problem. However, these papers present limited concrete analytical results.

To strengthen the analytical aspect, Akşin and Karaesmen (2007) use a network flow model to show that the expected throughput is concave in the degree of flexibility, and provide some results on the interaction between flexibility and capacity. In another study, Chou et al. (2007) analyze the worst-case performance of process flexibility structures for a symmetrical system. They adopt the concept of graph expansion, which is widely used in the area of graph theory and computer science, to prove the existence of a sparse partially flexible structure (not necessarily chaining) that accrues most of the benefits of full flexibility for all demand scenarios.

The process flexibility problem is intimately related to the problem of determining the expected amount of maximum flow in a network with random capacity. Karp et al. (1993) developed an algorithm to find the maximum flow in a random network with high probability, but to the best of our knowledge, the algorithm could not be used to find the expected max-flow value. For the case when the capacities are exponentially distributed, Lyons et al. (1999) used the connection between random walk and electrical network theory to bound the expected max-flow value by the conductance of a related electrical network (where the capacity of each arc is replaced by the expected capacity value). The proof technique relies heavily on the properties of the exponential distribution and hence cannot be utilized for more general distribution.

3. Balanced and Identical Expected Demand and Supply

In this section, we consider the case with n plants and n products, with (fixed) supply and (mean) demand of μ units for each plant and product, respectively. We further assume that all products have independent and identically distributed demand D_i, which follows a symmetric distribution around its mean $E[D_i] = \mu$.[3] Because demand cannot be negative, we assume that $D_i \in [0, 2\mu]$ for all demand realizations. Let $\mathbf{D} = (D_1, \ldots, D_n)$ denote the demand of the n products. Let $MF(\mathcal{G}, \mathbf{D})$ denote the maximum amount of production supported by the structure $\mathcal{G}$ in the system (obtained by solving the max-flow problem $Z_{\mathcal{G}}^*(\mathbf{D})$). For the dedicated and full flexibility system, it is easy to see that

$$MF(\mathcal{D}(n), \mathbf{D}) = \sum_{i=1}^{n} \min(\mu, D_i) = \sum_{i=1}^{n} (\mu - (\mu - D_i)^+)$$

and

$$MF(\mathcal{F}(n), \mathbf{D}) = \min\left(n\mu, \sum_{i=1}^{n} D_i\right)$$

$$= n\mu + \min\left(0, \sum_{i=1}^{n} D_i - n\mu\right).$$

Because demands are independent and bounded, by the Central Limit Theorem,

$$E\left[\min\left(0, \sum_{i=1}^{n} D_i - n\mu\right)\right]$$

$$= \sqrt{n} E\left[\min\left(0, \frac{\sum_{i=1}^{n}(D_i - \mu)}{\sqrt{n}}\right)\right] \sim O(\sqrt{n}).$$

We are interested in comparing the performance of the long chain, vis-á-vis the full flexibility system. In particular, we want to evaluate

$$\lim_{n \to \infty} \frac{E[MF(\mathcal{C}(n), \mathbf{D})]}{E[MF(\mathcal{F}(n), \mathbf{D})]}. \tag{6}$$

Besides tracking the above ratio, we would also like to track the improvement of the chaining structure over the dedicated system. This refinement is useful because it rules out those cases where the dedicated system is already as good as the full flexibility system. In fact, for the dedicated system, it is easy to show that

$$\lim_{n \to \infty} \frac{E[MF(\mathcal{D}(n), \mathbf{D})]}{E[MF(\mathcal{F}(n), \mathbf{D})]} = \frac{\mu - E[(\mu - D_i)^+]}{\mu}. \tag{7}$$

By our assumption, $E[(\mu - D_i)^+] \leqslant \mu/2$, hence we already have

$$\lim_{n \to \infty} \frac{E[MF(\mathcal{D}(n), \mathbf{D})]}{E[MF(\mathcal{F}(n), \mathbf{D})]} \geqslant 1/2.$$

Let

$$CE(n) \triangleq \frac{E[MF(\mathcal{C}(n), \mathbf{D})] - E[MF(\mathcal{D}(n), \mathbf{D})]}{E[MF(\mathcal{F}(n), \mathbf{D})] - E[MF(\mathcal{D}(n), \mathbf{D})]}$$

$$= \frac{E[MF(\mathcal{C}(n), \mathbf{D})] - n\mu + nE[(\mu - D_i)^+]}{nE[(\mu - D_i)^+] - O(\sqrt{n})}. \quad (8)$$

$CE(n)$ measures the extent of the improvement accrued by the chaining structure, vis-á-vis the dedicated system, and normalized by the maximal possible improvement which is attained by the full flexibility system.

Let ACE denote the asymptotic value of the chaining efficiency, where

$$ACE \triangleq \lim_{n \to \infty} CE(n).$$

It follows that

$$ACE = \frac{E[(D_i - \mu)^+] + \lim_{n \to \infty} (1/n)E[MF(\mathcal{C}(n), \mathbf{D})] - \mu}{E[(D_i - \mu)^+]}. \quad (9)$$

Hence, our focus from here onward is to find $(1/n) \cdot E[MF(\mathcal{C}(n), \mathbf{D})]$. Note that

$$\lim_{n \to \infty} \frac{E[MF(\mathcal{C}(n), \mathbf{D})]}{E[MF(\mathcal{F}(n), \mathbf{D})]}$$

$$= ACE + (1 - ACE)\left(\lim_{n \to \infty} \frac{E[MF(\mathcal{D}(n), \mathbf{D})]}{E[MF(\mathcal{F}(n), \mathbf{D})]}\right) \quad (10)$$

$$= ACE + (1 - ACE)\left(\frac{\mu - E[(\mu - D_i)^+]}{\mu}\right). \quad (11)$$

Because the flow on each arc is bounded by μ, we can delete a link from the chain $\mathcal{C}(n)$, to obtain $\mathcal{P}(n)$, without affecting the asymptotic performance of the two structures. In fact, we have the following lemma:

LEMMA 1. $\displaystyle \lim_{n \to \infty} \frac{E[MF(\mathcal{P}(n), \mathbf{D})]}{n} = \lim_{n \to \infty} \frac{E[MF(\mathcal{C}(n), \mathbf{D})]}{n}$.

We thus focus on finding the maximum flow on the path structure $\mathcal{P}(n)$, rather than the chain $\mathcal{C}(n)$.

3.1. Maximum Flow on $\mathcal{P}(n)$

For ease of exposition, we let the arc linking demand node i to supply node i denote the "primary" arc, and the arc linking demand node i to supply node $i+1$ denote the "secondary" arc. We delete the arc from demand node n to supply node 1 from the 2-chain to obtain the path $\mathcal{P}(n)$. The max flow on $\mathcal{P}(n)$ can be determined in a greedy fashion: First, satisfy the demand D_1 using whatever (primary) capacity provided by the primary arc that is available, and then use as much (secondary) capacity provided by the secondary arc as needed. Next, move to the next product, and based on the level of (primary) capacity remaining, satisfy the demand D_i using the primary followed by secondary capacities, with i ranging from 2 to n, in that order. The

amount of max-flow obtained in this greedy fashion is a random variable, depending on the values of D_i.

To present this greedy approach formally and to facilitate our analysis, we let T_i denote the amount of primary capacity left for product i and let S_i denote the amount of secondary capacity consumed by product i, after demands for products 1 to $i-1$ have been satisfied using the greedy method. Therefore, $T_i = \mu - S_{i-1}$ and we set $S_0 = 0$. Let TF denote the total maximum flow. Similarly, let $\text{TE} = \sum_{i=1}^{n} D_i - \text{TF}$ denote the difference between the total demand and the total flow; that is, the total unmet demand. This implies that

$$\frac{1}{n}E[\text{TE}] = \mu - \frac{1}{n}E[\text{TF}]. \quad (12)$$

We account for TF by keeping track of TE as we assign capacity to demand. Consider step i of the greedy approach, wherein T_i is known before D_i is observed. The greedy allocation implies that

$$S_i \leftarrow \min[(D_i - T_i)^+, \mu], \quad T_{i+1} \leftarrow \mu - S_i,$$

$$\text{TE} \leftarrow \text{TE} + [(D_i - T_i)^+ - \mu]^+. \quad (13)$$

We summarize the greedy approach as follows.

ALGORITHM 1. (Greedy Approach)
Step 1. Set $i := 1$, $S_0 := 0$, $T_1 := \mu$, and $\text{TE} := 0$.
Step 2. Observe D_i.
If $D_i > \mu$, then $S_i := \min[S_{i-1} + D_i - \mu, \mu]$, $T_{i+1} = \mu - S_i$, and $\text{TE} := \text{TE} + \max[D_i - T_i - \mu, 0]$.
If $D_i < \mu$, then $S_i := \max[S_{i-1} + D_i - \mu, 0]$, $T_{i+1} = \mu - S_i$, and $\text{TE} := \text{TE}$.
Step 3. If $i = n - 1$, then STOP. $\text{TE} := \text{TE} + \max(D_n - T_n, 0)$. Return TE as the minimum excess.
Otherwise, $i := i + 1$ and go to Step 2.

At this point, note that $\{S_i: i = 0, 1, 2, \ldots\}$ behaves much like a generalized random walk, with random step size $X_i \triangleq D_i - \mu$ and absorbing boundaries 0 and μ. The value TE grows in Step 2 only when $D_i - T_i > \mu$; that is, when $S_i = \min(D_i + S_{i-1} - \mu, \mu) = \mu$. We call this quantity $(X_i - T_i)$ the level of overshoot at the upper boundary. Note that $(X_i - T_i) = D_i - T_i - \mu$.

In Step 2 of the greedy algorithm, when $D_i < \mu$, it is possible that $S_{i-1} + D_i - \mu < 0$. We call this amount $(-S_{i-1} - X_i)$ the level of overshoot at the lower boundary. Note that we do not account for overshoot at the lower boundary while keeping track of TE in the greedy algorithm.

The random walk starts initially at $S_0 = 0$, the lower boundary. It gets trapped at the lower boundary whenever $X_i < 0$, and escapes only when $X_i > 0$. An interesting phenomenon happens when the random walk hits the upper boundary—the walk gets trapped at the upper boundary whenever $X_i > 0$, and it escapes from the upper boundary only when $X_i < 0$.

Let

$$\tau \triangleq \inf\{n: S_n = \mu, \ n \geqslant 1\}$$

denote the stopping time when the walk first hits the upper boundary. We can re-start the random walk from the lower boundary at time τ: interchange the roles of the upper and lower boundaries, and let

$$X_i' \leftarrow -X_i = \mu - D_i \quad \forall i > \tau, \tag{14}$$

$$S_\tau' \leftarrow \mu - S_\tau = 0, \tag{15}$$

$$S_i' = \begin{cases} \min[S_{i-1}' + X_i', \mu] & \text{if } X_i' > 0 \\ \max[S_{i-1}' + X_i', 0] & \text{if } X_i' < 0 \end{cases} \quad \forall i > \tau. \tag{16}$$

Because X_i' is distributed in an identical fashion to X_i by symmetry of demand distribution, the random walk S_i' from $S_\tau' = 0$ onward, under the above change of coordinate, is identical in distribution to the earlier random walk S_i starting at $S_0 = 0$.

Note that the way we account for TE changes under this new model. In the earlier walk, TE changes value only at the upper boundary, whereas in the new random walk, TE changes only when there is overshoot at the lower boundary. We repeat this process whenever the new random walk hits the upper boundary, switching back to the original random walk model. Let $\hat{S}_i$ denote the stochastic process obtained by toggling between S_i and S_i' in the above manner.

EXAMPLE 1. Figure 2 shows an example of a path that the random walk $\{S_i, \ i = 0, 1, 2, \ldots\}$ might traverse. Here, products 1, 3, 4, 9, and 10 have demands lower than μ, while the rest have demands higher than μ. We also see the walk get absorbed in the lower boundary three times and in the upper boundary once. When the walk was absorbed in the upper boundary, some unmet demands for products 6, 7, and 8 were lost. We are interested in the expected amount of such excess quantities.

Suppose we consider another generalized random walk $\{\hat{S}_i, i = 0, 1, 2, \ldots\}$ such that $\hat{S}_0 = S_0$, but $\hat{S}_i$ toggles between $S_i' = \mu - S_i$ and S_i each time $\hat{S}_i$ hits the upper boundary. That is, the first time $\hat{S}_i$ hits the upper boundary, change to $\hat{S}_i = S_i'$; the next time, switch back to $\hat{S}_i = S_i$, and so on. Figure 3 shows the equivalent sample path for the new random walk that corresponds to the sample path for the old random walk in Figure 2.

Note that unmet demand is incurred at the upper boundary when $\hat{S}_i = S_i$, but at the lower boundary when $\hat{S}_i = S_i'$.

Figure 2. Sample path for $\{S_i, \ i = 0, 1, 2, \ldots\}$.
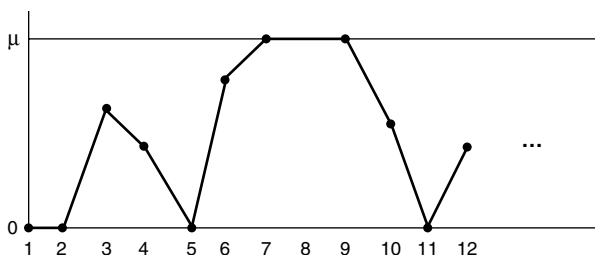
Figure 3. Sample path for $\{\hat{S}_i, \ i = 0, 1, 2, \ldots\}$.
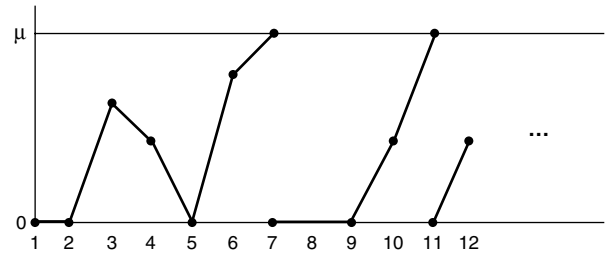
For example, in Figure 3, we easily verify that indeed, unmet demands are incurred for products 6, 7, and 8.

Although it is possible to work on $\{S_i, \ i = 0, 1, 2, \ldots\}$, the transformation to $\{\hat{S}_i, \ i = 0, 1, 2, \ldots\}$ provides a more convenient formulation. In particular, $\{\hat{S}_i, \ i = 0, 1, 2, \ldots\}$ turns out to be a regenerative process whenever the random walk hits the upper boundary—the process regenerates, and its continuation is a probabilistic replica of the original process starting at Step 1 again.

Because all regenerating cycles are probabilistically identical, it suffices to examine the characteristics of one cycle for the purpose of asymptotic analysis. Some of these relevant characteristics are:

• Cycle duration τ: the length of each regenerative cycle. Recall that

$$\tau \triangleq \inf\{n: S_n = \mu, \ n \geqslant 1, \ S_0 = 0\}.$$

• Cycle overshoot ψ: the amount of overshoots at both the lower and upper boundaries in each cycle;

$$\psi \triangleq \sum_{i=1}^\tau \big((S_i - S_{i-1} - X_i)\chi(X_i < 0) \\ + (S_{i-1} + X_i - S_i)\chi(X_i > 0)\big),$$

where $\chi(\cdot)$ denote the indicator function.

Note that ψ can be decomposed into two components, with $\psi = \psi_L + \psi_U$, where

$$\psi_L \triangleq \sum_{i=1}^\tau \big((S_i - S_{i-1} - X_i)\chi(X_i < 0)\big)$$

and

$$\psi_U \triangleq \sum_{i=1}^\tau \big((S_{i-1} + X_i - S_i)\chi(X_i > 0)\big).$$

Consider a renewal process $\{N(t): t \geqslant 0\}$, having i.i.d. interarrival time Y_i with $Y_i \sim \tau$ for all i. The reward R_i obtained at the ith renewal is ψ_L if i is even, and is ψ_U if i is odd. Note that from (12),

$$\sum_{i=1}^n D_i - \sum_{i=1}^{N(n)+1} R_i \leqslant MF(\mathscr{P}_n, \mathbf{D}) \leqslant \sum_{i=1}^n D_i - \sum_{i=1}^{N(n)} R_i. \tag{17}$$

Because $\hat{S}_i$ toggles alternately between S_i and S_i' and by the renewal reward theorem,

$$\lim_{n \to \infty} \frac{E[\sum_{i=1}^{N(n)} R_i]}{n} = \frac{E[\psi_L] + E[\psi_U]}{2} \frac{1}{E[\tau]}.$$

Hence, taking the limit in (17), we obtain:

THEOREM 1. $\lim_{n \leftarrow \infty} \dfrac{E[MF(\mathscr{P}(n), \mathbf{D})]}{n} = \mu - \dfrac{E[\psi]/2}{E[\tau]}$.

For any discrete demand distribution symmetrical around the mean Δ, the parameters $E[\psi]$ and $E[\tau]$ can be obtained by solving a system of linear equations. We represent the distribution as follows:

$$\text{support}\{D_i\} = \{0, 1, \ldots, \Delta, \ldots, 2\Delta - 1, 2\Delta\}.$$

Let

$$P_x = \text{Prob}(D_i = \Delta + x) \quad \forall x = -\Delta, -\Delta + 1, \ldots, \Delta - 1, \Delta,$$

and w.l.o.g.,[4]

$$P_x = P_{-x} > 0, \quad P_0 = 0.$$

Define the stopping time if the random walk started at x,

$$\tau_x \triangleq \inf\{n: S_n = \mu, \ n \geqslant 1, \ S_0 = x\}.$$

Clearly, $\tau = \tau_0$ and $\tau_\Delta = 0$. Conditioning on the next move,

$$E[\tau_x] = 1 + \sum_{j=1}^{\Delta-1} E[\tau_j] P_{j-x} + E[\tau_0] \sum_{j=x}^{\Delta} P_j$$
$$\forall x = 0, 1, \ldots, \Delta - 1. \quad (18)$$

We can obtain $E[\tau] = E[\tau_0]$ by solving the system of Equation (18).

Similarly, given $S_0 = x$, we define the overshoot as

$$\psi_x \triangleq \sum_{i=1}^{\tau_x} \big((S_i - S_{i-1} - X_i)\chi(X_i < 0)$$
$$+ (S_{i-1} + X_i - S_i)\chi(X_i > 0)\big).$$

Obviously, $\psi = \psi_0$ and $\psi_\Delta = 0$. Conditioning on the next move,

$$E[\psi_x] = r_x + \sum_{j=1}^{\Delta-1} E[\psi_j] P_{j-x} + E[\psi_0] \sum_{j=x}^{\Delta} P_j$$
$$\forall x = 0, 1, \ldots, \Delta - 1, \quad (19)$$

where

$$r_x = \sum_{j=\Delta}^{\Delta+x} (j - \Delta) P_{j-x} + \sum_{j=x}^{\Delta} (j - x) P_j \quad \forall x = 0, 1, \ldots, \Delta - 1.$$

We can obtain $E[\psi] = E[\psi_0]$ by solving the system of Equation (19).

By Theorem 1 and the definition of *ACE*, we have:

THEOREM 2. *The asymptotic chaining efficiency can be uniquely obtained as follows*:

$$ACE = 1 - \frac{E[\psi_0]}{2E[\tau_0]E[(D_i - \Delta)^+]},$$

where $E[\psi_0]$ and $E[\tau_0]$ come from the solutions to linear systems (18) *and* (19), *respectively*.

PROOF. We show first the uniqueness of the solutions to (18) and (19). Observe that (18) and (19) have the same homogeneous system. Because $\sum_{j=1}^{\Delta-1} P_{j-x} + \sum_{j=x}^{\Delta} P_j < 1$, the associated matrix is strictly diagonally dominated, hence nonsingular.

Now, from (9), Lemma 1, and Theorem 1,

$$ACE = \frac{E[(D_i - \Delta)^+] + \lim_{n \to \infty}(1/n)E[MF(\mathscr{P}(n), \mathbf{D})] - \Delta}{E[(D_i - \Delta)^+]}$$

$$= 1 - \frac{E[\psi_0]}{2E[\tau_0]E[(D_i - \Delta)^+]}. \quad \square$$

Furthermore, *ACE* is invariant over the scale of the demand.

COROLLARY 1. *Suppose that $D_i' \sim cD_i$, $c > 0$. Then, $ACE' = ACE$.*

PROOF. It is easy to see that $E[\tau_0'] = E[\tau_0]$, $E[\psi_0'] = cE[\psi_0]$, and $E[(D_i' - c\Delta)^+] = cE[(D_i - \Delta)^+]$, from which the result follows. $\square$

This gives rise to an efficient method to determine the asymptotic efficiency of the 2-chain. When demand follows a discrete distribution, Theorem 1 and Corollary 1 can be directly applied. On the other hand, when demand follows a continuous distribution from 0 to 2μ, the above results can still be used to approximate the asymptotic chaining performance. This is done by discretizing the distribution into $2\Delta + 1$ equally spaced demand points from 0 to 2μ. Obviously, the more discrete points used, the better the approximation.

3.2. Applications

3.2.1. Two-Point Distribution. When demand $D_i = 0$ or 2μ with equal probability, then it is easy to see that

$$E[\psi] = \mu, \quad E[\tau] = 2.$$

Hence, $ACE = 0.5$. Furthermore, because $E[(\mu - D_i)^+] = \mu/2$,

$$\lim_{n \to \infty} \frac{E[MF(\mathscr{C}(n), \mathbf{D})]}{E[MF(\mathscr{F}(n), \mathbf{D})]} = 0.5 + 0.5 \times 0.5 = 0.75.$$

Thus, the chaining strategy achieves only 75% of the efficiency of the full flexibility system. This poor performance stems in part from the large variability in the demand distribution.

3.2.2. Uniform Distribution. Suppose that demand $D_i = 0, 1, \ldots, \Delta - 1, \Delta + 1, \ldots, 2\Delta - 1, 2\Delta$ with equal probability; that is,

$$P_x = \frac{1}{2\Delta} \quad \forall x = 1, \ldots, \Delta - 1, \Delta.$$

It can be shown that

$$\mathrm{E}[\tau_0] = \frac{4\Delta(2\Delta + 1)}{(\Delta + 2)(\Delta + 1)}, \quad \mathrm{E}[\psi_0] = \frac{\Delta(5\Delta + 4)}{3(\Delta + 2)},$$

$$\mathrm{E}[D_i - \Delta]^+ = \frac{\Delta + 1}{4}.$$

Hence,

$$ACE = \frac{7\Delta + 2}{12\Delta + 6}.$$

Furthermore, because $\mathrm{E}[(\Delta - D_i)^+] = (\Delta + 1)/4$,

$$\lim_{n \to \infty} \frac{\mathrm{E}[MF(\mathscr{C}(n), \mathbf{D})]}{\mathrm{E}[MF(\mathscr{F}(n), \mathbf{D})]} = \frac{7\Delta + 2}{12\Delta + 6} + \left(0.75 - \frac{1}{4\Delta}\right)$$

$$\times \frac{5\Delta + 4}{12\Delta + 6} = \frac{43\Delta^2 + 15\Delta - 4}{48\Delta^2 + 24\Delta}.$$

When demand D_i is uniformly distributed over $[0, 2\mu]$, we can obtain the *ACE* by first discretizing the interval into $2\Delta + 1$ demand points, then taking the limit as $\Delta \to \infty$. Hence,

$$ACE = \lim_{\Delta \to \infty} \frac{7\Delta + 2}{12\Delta + 6} = \frac{7}{12} \approx 58.33\%$$

and

$$\lim_{n \to \infty} \frac{\mathrm{E}[MF(\mathscr{C}(n), \mathbf{D})]}{\mathrm{E}[MF(\mathscr{F}(n), \mathbf{D})]} = \lim_{\Delta \to \infty} \frac{43\Delta^2 + 15\Delta - 4}{48\Delta^2 + 24\Delta}$$

$$= \frac{43}{48} \approx 89.58\%.$$

Note that in this case, the value of the expected max-flow in the 2-chain is already around 89.6% of the expected max-flow in the full flexibility system!

3.2.3. Normal Distribution. Suppose that demand $D_i \sim N(\mu, \sigma)$. Then, we can likewise approximate the value of *ACE* by discretization. Moreover, Corollary 1 implies that for a fixed coefficient of variation, the *ACE* is independent of the actual magnitudes of μ and σ. Table 1 summarizes how the *ACE* values change with respect to the discretization level and the coefficient of variation.

Obviously, as we increase the number of demand points, the approximation becomes finer. More importantly, the

Table 1. *ACE* for various Δ and CV values.

	Coefficient of variation (CV)						
Δ	0.33	0.31	0.29	0.27	0.25	0.23	0.21
2	0.6452	0.6509	0.6559	0.6599	0.6629	0.6649	0.6660
4	0.6895	0.7007	0.7124	0.7244	0.7365	0.7486	0.7604
6	0.6970	0.7090	0.7216	0.7348	0.7484	0.7623	0.7765
8	0.6997	0.7119	0.7248	0.7383	0.7524	0.7669	0.7819
10	0.7010	0.7133	0.7263	0.7399	0.7542	0.7690	0.7843
12	0.7017	0.7140	0.7271	0.7408	0.7552	0.7701	0.7856
14	0.7022	0.7145	0.7275	0.7413	0.7558	0.7708	0.7864

Table 2. $\mathrm{E}[MF(\mathscr{C}(n), \mathbf{D})]/\mathrm{E}[MF(\mathscr{F}(n), \mathbf{D})]$ for various CV values.

	Coefficient of variation (CV)						
	0.33	0.31	0.29	0.27	0.25	0.23	0.21
$\lim_{n \to \infty} \dfrac{\mathrm{E}[MF(\mathscr{C}(n), \mathbf{D})]}{\mathrm{E}[MF(\mathscr{F}(n), \mathbf{D})]}$	0.9614	0.9650	0.9687	0.9723	0.9758	0.9791	0.9823

value of *ACE* decreases in the coefficient of variation. This is because as relative uncertainty decreases, the need for any form of flexibility is reduced, thus improving the value of the 2-chain relative to full flexibility.

We tabulate next the ratio of the expected sales from the chaining structure and the full flexibility system in Table 2. Interestingly, even with a coefficient of variation (CV) of 0.33, the expected sales under the chaining structure are already close to 96% of the full flexibility system.

3.3. Extensions

The proposed method works as long as all products have the same demand distribution and all plants have the same capacity, even if the system is unbalanced (i.e., capacity not equal to mean demand) and the demand distribution is not symmetrical.

3.3.1. Nonsymmetrical Demand. In this case, the odd and the even cycles will have different stopping times and overshoots. To demonstrate, we consider the following example:

$$D_i = \begin{cases} 0, & \text{w.p. } 0.3 \\ 2, & \text{w.p. } 0.1 \\ 4, & \text{w.p. } 0.4 \\ 8, & \text{w.p. } 0.2 \end{cases} \quad \text{and} \quad C_i = 5.$$

Note that D_i is not symmetrical about the mean. Let τ_i and $\hat{\tau}_i$ be the stopping times for the odd and even cycles, respectively, and let ψ_i and $\widehat{\psi}_i$ denote the respective overshoots. Then,

$$\begin{bmatrix} 0.2 & 0 & 0 & -0.2 & 0 \\ -0.8 & 1 & 0 & 0 & -0.2 \\ -0.4 & -0.4 & 1 & 0 & 0 \\ -0.4 & 0 & -0.4 & 1 & 0 \\ -0.3 & -0.1 & 0 & -0.4 & 1 \end{bmatrix} \begin{bmatrix} \mathrm{E}[\tau_0] & \mathrm{E}[\psi_0] \\ \mathrm{E}[\tau_1] & \mathrm{E}[\psi_1] \\ \mathrm{E}[\tau_2] & \mathrm{E}[\psi_2] \\ \mathrm{E}[\tau_3] & \mathrm{E}[\psi_3] \\ \mathrm{E}[\tau_4] & \mathrm{E}[\psi_4] \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0.2 \\ 1 & 0.4 \end{bmatrix}$$

and

$$\begin{bmatrix} 0.8 & -0.4 & 0 & -0.1 & 0 \\ -0.2 & 1 & -0.4 & 0 & -0.1 \\ -0.2 & 0 & 1 & -0.4 & 0 \\ -0.2 & 0 & 0 & 1 & -0.4 \\ 0 & -0.2 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} E[\hat{\tau}_0] & E[\hat{\psi}_0] \\ E[\hat{\tau}_1] & E[\hat{\psi}_1] \\ E[\hat{\tau}_2] & E[\hat{\psi}_2] \\ E[\hat{\tau}_3] & E[\hat{\psi}_3] \\ E[\hat{\tau}_4] & E[\hat{\psi}_4] \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0.6 \\ 1 & 0.4 \\ 1 & 0.2 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}.$$

Hence,

$$ACE = 1 - \frac{E[\psi_0] + E[\hat{\psi}_0]}{(E[\tau_0] + E[\hat{\tau}_0])E[D_i - C_i]^+}$$

$$= 1 - \frac{0.7436 + 1.2377}{(22.7920 + 2.8798)0.6} = 0.8714.$$

In general, the above approach can be extended to obtain the asymptotic performance of the 2-chain in a balanced demand-and-supply system, as long as the demand is identical for all locations.

3.3.2. Unbalanced System. We consider next the situation when the total supply capacity might not be the same as the total demand. Consider the case when the demands are normally distributed with mean μ and standard deviation σ (with a CV of at most 0.33), but the capacity of each plant is λ.

When $\lambda = \mu$, we note that the absence of a safety capacity entails a fill rate of only $100(1 - 0.399 \times \text{CV})\% = 86.7\%$ for each product, when $\text{CV} = 0.33$. To guarantee a 97.26% fill rate for each product, a dedicated system ought to carry a safety capacity of σ units for each product, leading to a total safety capacity of $n\sigma$ units. Because full flexibility corresponds to complete capacity pooling, we can achieve a 97.26% fill rate for the entire system with only $\sigma\sqrt{n}$ units of safety capacity. This dramatic reduction in safety stock investment performance comes about with full flexibility in the production system. We investigate the corresponding performance in the case of the chaining structure. Tables 3 and 4 demonstrate that both *ACE* and the expected sales ratio of chaining to full flexibility increase in the ratio of λ/μ.

This suggests that the balanced scenario $(\lambda = \mu)$ provides a lower bound for the situation when $\lambda > \mu$ (i.e., safety capacity scenario). Therefore, with $\lambda = \mu + \sigma/\sqrt{n}$ (i.e., $\sigma\sqrt{n}$ units of total safety capacity in the system), a chaining structure can already guarantee a fill rate of at least $97.26\% \times 96.14\% = 93.5\%$. Note that the

average safety capacity per plant is decreasing in n. For the dedicated system to maintain this level of fill rate, the corresponding safety capacity investment is at least $0.5n\sigma$. This analysis suggests another advantage of flexibility in production planning: Apart from increasing the expected sales, the flexibility strategy can also help to decrease the safety capacity investment needed to maintain a required fill-rate level. In the identical demand case, we expect that the safety capacity investment that is needed should decrease roughly by a factor of $O(\sqrt{n})$.

4. General Demand and Supply

Consider a manufacturer with n plants that can be used to produce m different products. The capacities of the plants are fixed, and the demands of the products are random. In this section, we assume that demand can be arbitrarily correlated and nonidentical.

Let $\mathbf{D} = (D_1, D_2, \ldots, D_m)$ denote the demands of products and $\mathbf{C} = (C_1, C_2, \ldots, C_n)$ denote the capacities of plants. D_i is a random variable with mean μ_i. We also assume that each unit of the capacity in plant j can be used to produce one unit of product i, if $(i, j) \in \mathcal{G}$. Let $MF(\mathcal{G}, \mathbf{D})$ denote the total demand satisfied with structure $\mathcal{G}$, when demand is $\mathbf{D}$. The manufacturer wants to design a flexibility structure $\mathcal{G}$ to effectively deal with demand uncertainty. Note that a 2-chain might not perform well in this situation.

Obviously, the performance of the full flexibility structure (the complete bipartite graph $\mathcal{F}$) is the best among all flexibility structures. However, the investment and coordination cost of such a system is also the largest. We

Table 3. ACE for various λ and CV values.

λ/μ	Coefficient of variation (CV)						
	0.33	0.31	0.29	0.27	0.25	0.23	0.21
0.85	0.2986	0.2810	0.2616	0.2400	0.2162	0.1904	0.1628
0.90	0.4156	0.4035	0.3892	0.3720	0.3513	0.3265	0.2972
0.95	0.5561	0.5552	0.5531	0.5492	0.5428	0.5328	0.5180
1.00	0.7037	0.7159	0.7290	0.7428	0.7574	0.7726	0.7885
1.05	0.8314	0.8510	0.8715	0.8924	0.9136	0.9345	0.9541
1.10	0.9189	0.9365	0.9529	0.9673	0.9794	0.9886	0.9947
1.15	0.9659	0.9771	0.9859	0.9923	0.9964	0.9986	0.9996

Table 4. $E[MF(\mathcal{C}(n), \mathbf{D})]/E[MF(\mathcal{F}(n), \mathbf{D})]$ for various λ and CV values.

λ/μ	Coefficient of variation (CV)						
	0.33	0.31	0.29	0.27	0.25	0.23	0.21
0.85	0.8467	0.8475	0.8482	0.8489	0.8494	0.8497	0.8499
0.90	0.8912	0.8930	0.8948	0.8964	0.8978	0.8988	0.8994
0.95	0.9304	0.9335	0.9365	0.9394	0.9421	0.9444	0.9464
1.00	0.9614	0.9650	0.9687	0.9723	0.9758	0.9791	0.9823
1.05	0.9820	0.9852	0.9882	0.9909	0.9934	0.9955	0.9972
1.10	0.9930	0.9950	0.9966	0.9979	0.9988	0.9994	0.9998
1.15	0.9977	0.9986	0.9992	0.9996	0.9998	1.0000	1.0000

are thus interested in identifying a "sparse" structure $\mathcal{G}$, where the ability to match capacity to demand is as close to the full flexibility system as possible. Our problem boils down to finding a structure $\mathcal{G}$ with $E[MF(\mathcal{G}, \mathbf{D})]$ as close to $E[MF(\mathcal{F}, \mathbf{D})]$ as possible.

Our problem can be cast in a more general framework. Note that the maximum-flow problem $MF(\mathcal{F}, \mathbf{D})$, obtained by solving $Z_{\mathcal{F}}^*(\mathbf{D})$, is a special case of the following more general problem:

(P): $Z(b, \{1, \ldots, n\})$
$$= \max\left\{\sum_{i=1}^n c_i x_i \colon Ax \leqslant b; \ x_i \geqslant 0, \ i = 1, \ldots, n\right\},$$

where $c_i \geqslant 0$ ($\forall i = 1, \ldots, n$), A is an $m \times n$ nonnegative matrix, b is random and nonnegative, and $n \gg m$. Let

$$Z(b, S) = \max\left\{\sum_{i \in S} c_i x_i \colon A_S x_S \leqslant b; \ x_i \geqslant 0, \ i \in S\right\},$$

where A_S (respectively, x_S) denote the columns (respectively, rows) of A (respectively, x) indexed by the subset S. Our goal is to identify a subset S so that

$$E_b(Z(b, S)) \geqslant (1 - \epsilon)E_b(Z(b, \{1, \ldots, n\}))$$

and

$$|S| \approx O(m) \ll n.$$

The set S is the support for the feasible solution in (P) and corresponds to the process structure in the process flexibility problem. In this way, the design of the process structure corresponds to the selection of the variables to be retained in the support set S. We interpret this problem using the dual of (P), where the variables in (P) correspond to the constraints in the dual LP. Our problem now reduces to a constraint selection problem. We address this problem using the recent approach of constraint sampling, independently developed by Calafiore and Campi (2005) and de Farias and Van Roy (2004). We note that the approach by de Farias and Van Roy (2004), although couched in a different context, uses a sampling distribution based on an optimal policy to the original problem and is very similar to the key idea used here. However, we take a small step further and identify a condition on the optimal solution to ensure that the error (loss of optimality) introduced in the constraint sampling approach will be small.

4.1. Constraint Sampling

The problem addressed by Calafiore and Campi (2005) can be formulated as follows:

(UCP): $\big\{\min c^T x \colon f(x, \delta) \leqslant 0; \ \delta \in \mathbf{\Delta}; \ x \in \mathcal{X} \subseteq \mathbb{R}^n\big\},$

where x is the decision variable, $\mathcal{X}$ is a convex and closed region, δ is the random parameter in set $\mathbf{\Delta}$ ($\mathbf{\Delta} \subset \mathbb{R}^l$), and $f(x, \delta)$ is continuous and convex in x for all δ.

The uncertain linear program (UCP) could involve an infinite number of constraints because the set $\mathbf{\Delta}$ may already be uncountable. Instead, Calafiore and Campi (2005) studied the randomized version of UCP, where the constraint set $\mathbf{\Delta}$ is endowed with a probability distribution. They proposed a "randomized constraint sampling" approach to construct a "ϵ-robust" feasible solution (i.e., the probability that the solution obtained will violate a random constraint in $\mathbf{\Delta}$ is less than ϵ). They generated N constraints by sampling the parameter δ from $\mathbf{\Delta}$ using the endowed probability distribution, and solved the following problem:

$\text{UCP}_N \colon \big\{\min c^T x \colon f(x, \delta^k) \leqslant 0, \ k = 1, \ldots, N; \ x \in \mathcal{X} \subseteq \mathbb{R}^n\big\},$

where δ^ks are the parameters sampled.

They showed that the solution of the new problem will violate only a tiny portion of the original constraints if N is large enough. Specifically, if $N \geqslant n/(\epsilon\beta) - 1$, the probability that the optimal solution of UCP_N (say $\hat{x}_N$) is ϵ-robust feasible is more than $1 - \beta$. Here, n is the dimension of x and $\epsilon, \beta \in (0, 1]$.

It is obvious that Calafiore and Campi's (2005) result is also valid for uncertain linear programming (ULP) problems (cf. de Farias and Van Roy 2004), where the constraints $f(x, \delta) \leqslant 0$ are linear. We will briefly describe the intuition of the proof in this case, as it applies to our problem. Note that in our approach, the nonnegative constraints $x \geqslant 0$ are always included in the subproblem UCP_N, and we only sample from the other constraints in $\mathbf{\Delta}$.

Let $z^{(1)}, \ldots, z^{(N+1)}$ represent $N + 1$ parameters sampled from $\mathbf{\Delta}$ with the same endowed distribution. Construct the following problem for each $k = 1, 2, \ldots, N + 1$:

$\text{ULP}_N^k \colon \{\min c^T x \colon x \geqslant 0; \ f(x, z^{(i)}) \leqslant 0,$
$\qquad i = 1, \ldots, k-1, k+1, \ldots, N+1\}.$

Let $\hat{x}_N^k$ denote an optimal solution of ULP_N^k. In the case of multiple optimal solutions, we choose one that is lexicographically maximal. In addition, define a problem ULP_{N+1} consisting of all $N + 1$ constraints:

$\text{ULP}_{N+1} \colon \{\min c^T x \colon x \geqslant 0; \ f(x, z^{(i)}) \leqslant 0, \ i = 1, \ldots, N+1,\},$

and let $\hat{x}_{N+1}$ denote an optimal solution of ULP_{N+1}.

Note that ULP_{N+1} and ULP_N^k differ in just one constraint—$f(x, z^{(k)}) \leqslant 0$. In the event that this constraint is not tight at the optimal solution for ULP_{N+1}, its deletion from the set of constraints will not affect the optimality of $\hat{x}_{N+1}$. Hence, we must have $\hat{x}_{N+1} = \hat{x}_N^k$; that is, $f(\hat{x}_N^k, z^{(k)}) \leqslant 0$. If N is sufficiently large, and if there are relatively fewer tight constraints, the event $f(\hat{x}_N^k, z^{(k)}) \leqslant 0$ holds with very high probability.

PROPOSITION 1 (CALAFIORE AND CAMPI 2005). *The probability that $\hat{x}_N^k$ will violate the kth (sampled) constraint $f(x, z^{(k)}) \leqslant 0$ is bounded above by $n/(N+1)$; that is,*

$$P(f(\hat{x}_N^k, z^{(k)}) > 0) \leqslant \frac{n}{N+1},$$

where n is the dimension of the decision variable x.

Note that for each k, the random variable $\hat{x}_N^k$ is stochastically equivalent to $\hat{x}_N$, the solution obtained by solving ULP_N, where N constraints are independently and identically sampled from $\mathbf{\Delta}$. More precisely, if the jth constraint in $\mathbf{\Delta}$ is sampled with probability q_j, the proposition ensures that

$$P(f(\hat{x}_N^k, z^{(k)}) > 0) = \sum_{j \in \Delta} q_j P(f(\hat{x}_N^k, z^{(k)}) > 0 \mid z^{(k)} = z^j)$$

$$= \sum_{j \in \Delta} q_j P(f(\hat{x}_N, z^j) > 0) \leqslant \frac{n}{N+1}. \quad (20)$$

Note that the above results hold as long as the $z^{(k)}$s are sampled in an identical manner, using the endowed probability distribution on $\mathbf{\Delta}$.

4.2. Identifying a Sparse Support Set

Consider the problem (P):

$$Z(b, \{1, \ldots, n\}) = \max\left\{\sum_{i=1}^n c_i x_i : Ax \leqslant b; \ x_i \geqslant 0, \ i = 1, \ldots, n\right\}.$$

For ease of exposition, we use $Z(b)$ to denote $Z(b, \{1, \ldots, n\})$. The dual problem (D) is given by

$$\text{(D):} \quad Z(b) = \min\left\{\sum_{j=1}^m b_j y_j : A^T y \geqslant c; \ y_j \geqslant 0, \ j = 1, \ldots, m\right\}.$$

(D) is a linear programming problem with m variables and n constraints. If we sample N constraints from (D), and denote the constraints sampled by S, we obtain the problem

$$\text{(D(S)):} \quad Z(b, S) = \min\left\{\sum_{j=1}^m b_j y_j : A_S^T y \geqslant c_S; \right.$$

$$\left. y_j \geqslant 0, \ j = 1, \ldots, m\right\}.$$

The dual of this problem is

$$\text{(P(S)):} \quad Z(b, S) = \max\left\{\sum_{i \in S} c_i x_i : A_S x_S \leqslant b; \ x_i \geqslant 0, \ i \in S\right\}.$$

Note that unlike the uncertain convex programming problem, we do not have an endowed distribution for the set of constraints. The selection of the sampling distribution plays a key role in our analysis. We discuss next how the sampling distribution can be obtained.

Let $x^*(b)$ denote an optimal solution in $Z(b)$. Note that because b is random, $x^*(b)$ is also a random vector. We assume that problem (P) has an optimal solution $x^*(b)$ with the following property:

PROPERTY A. *$x_i^*(b) \leqslant \lambda E_b(x_i^*(b))$ almost surely for some constant $\lambda > 0$ (independent of n), and for all $i = 1, \ldots, n$.*

Property A essentially states that the optimal primal solution $x^*(b)$, for each realization of b, should not be too far above its mean value. This property holds, for example, for a truncated normal distribution and many bounded demand distributions.

THEOREM 3. *Suppose that Property A holds for (P). Then, there exists a set S with cardinality $N = O(\lambda m/\epsilon)$, such that*

$$E_b(Z(b, S)) \geqslant (1 - \epsilon) E_b(Z(b)).$$

We prove Theorem 3 using constraint sampling on the dual problem (D). The ith constraint in problem (D) is sampled with probability

$$\frac{c_i E_b(x_i^*(b))}{\sum_{j=1}^n c_j E_b(x_j^*(b))}.$$

We repeat the sampling procedure N times and denote the set of (distinct) constraints obtained by S.

PROOF. For fixed b, let $x^*(b)$ and $y^*(b)$ denote the corresponding optimal primal and dual solution in (P) and (D), respectively. Similarly, let $x^*(b, S)$ and $y^*(b, S)$ denote the corresponding optimal primal and dual solutions in (P(S)) and (D(S)), respectively:

$$Z(b, S) = \sum_{j=1}^m b_j y_j^*(b, S)$$

$$= \sum_{j=1}^m b_j y_j^*(b, S) + Z(b) - \sum_{i=1}^n c_i x_i^*(b)$$

$$\geqslant Z(b) + \sum_{j=1}^m y_j^*(b, S)\left(\sum_{i=1}^n A_{ji} x_i^*(b)\right) - \sum_{i=1}^n c_i x_i^*(b)$$

$$= Z(b) + \sum_{i=1}^n x_i^*(b)\left(\sum_{j=1}^m A_{ji} y_j^*(b, S)\right) - \sum_{i=1}^n c_i x_i^*(b)$$

$$= Z(b) + \sum_{i=1}^n \left(\sum_{j=1}^m A_{ji} y_j^*(b, S) - c_i\right) x_i^*(b).$$

Note that

$$\left(\sum_{j=1}^m A_{ji} y_j^*(b, S) - c_i\right) x_i^*(b) \geqslant -c_i x_i^*(b)$$

holds for all $y^*(b, S)$, but when the constraint $\sum_{j=1}^m A_{ji} y_j \geqslant c_i$ holds for $y^*(b, S)$, then

$$\left(\sum_{j=1}^m A_{ji} y_j^*(b, S) - c_i\right) x_i^*(b) \geqslant 0.$$

Define

$$E_S(\alpha(S)) = \int_S \alpha(S) \, dF(S),$$

where $\alpha(S)$ is a function on space S, and $F(S)$ is a cumulative distribution function on S. Hence, we can obtain

$$
E_S(Z(b, S))
$$

$$
\geqslant Z(b) + E_S\left[\sum_{i=1}^{n}\left(\sum_{j=1}^{m} A_{ji} y_j^*(b, S) - c_i\right) x_i^*(b)\right]
$$

$$
\geqslant Z(b) - \left[\sum_{i=1}^{n} c_i x_i^*(b) P\left(\sum_{j=1}^{m} A_{ji} y_j^*(b, S) < c_i\right)\right]
$$

$$
\geqslant Z(b) - \left[\sum_{i=1}^{n} c_i \lambda E_b(x_i^*(b)) P\left(\sum_{j=1}^{m} A_{ji} y_j^*(b, S) < c_i\right)\right]
$$

$$
\text{(by Property A)}
$$

$$
= Z(b) - \lambda\left(\sum_{i=1}^{n} c_i E_b(x_i^*(b))\right)
$$

$$
\cdot\left[\sum_{i=1}^{n} \frac{c_i E_b(x_i^*(b))}{\sum_{k=1}^{n} c_k E_b(x_k^*(b))} P\left(\sum_{j=1}^{m} A_{ji} y_j^*(b, S) < c_i\right)\right]
$$

$$
\geqslant Z(b) - \frac{\lambda m}{N+1}\left(\sum_{i=1}^{n} c_i E_b(x_i^*(b))\right) \quad \text{(from (20))}.
$$

Because $\sum_{i=1}^{n} c_i E_b(x_i^*(b)) = E_b(Z(b))$, by taking expectation over b,

$$
E_{b,S}(Z(b, S)) \geqslant (1 - \epsilon) E_b(Z(b)).
$$

Hence, there exists a sparse support set S with cardinality $N = O(\lambda m/\epsilon)$ (independent of n) such that $E_b(Z(b, S)) \geqslant (1 - \epsilon) E_b(Z(b))$. $\quad\square$

Note that although we did not specify the set S in the proof, a good support set can be obtained by repeating the sampling experiment. By the law of the large numbers, if we perform a sufficient number of experiments, then it is likely that the best structure obtained from the experiments will be as good as the (population) mean $E_S(Z(b, S))$.

In the rest of the section, we use the above results to identify the condition on the random parameters b, so that the existence of a near-optimal sparse support set can be guaranteed for these problems.

4.3. Sparse Process Flexibility Structure

To apply Theorem 3, we need to understand the behavior of the (random) optimal solution $x^*(b)$ as b varies. Fortunately, for the process flexibility problem, this problem is trivial under the full flexibility structure $\mathcal{F}$. Note that in this problem, with demand $\mathbf{D} = (D_1, D_2, \ldots, D_m)$ and capacity $\mathbf{C} = (C_1, \ldots, C_n)$, the max-flow problem has up to nm variables, with only $O(n + m)$ constraints.

LEMMA 2.

$$
x_{ij}^*(\mathbf{D}) = \frac{D_i C_j}{\max\{\sum_{i=1}^{m} D_i, \ \sum_{j=1}^{n} C_j\}},
$$

$$
i = 1, \ldots, m, \ \ j = 1, \ldots, n,
$$

is an optimal solution to $Z_{\mathcal{F}}^*(\mathbf{D})$ under the full flexibility structure $\mathcal{F}$. Furthermore,

$$
MF(\mathcal{F}, \mathbf{D}) = \min\left\{\sum_{i=1}^{m} D_i, \ \sum_{j=1}^{n} C_j\right\}.
$$

PROOF. For the full flexibility structure $\mathcal{F}$, it is easy to see that

$$
MF(\mathcal{F}, \mathbf{D}) = \min\left\{\sum_{i=1}^{m} D_i, \sum_{j=1}^{n} C_j\right\}.
$$

Note also that

$$
x_{ij}^*(\mathbf{D}) = \frac{D_i C_j}{\max\{\sum_{i=1}^{m} D_i, \sum_{j=1}^{n} C_j\}}
$$

$$
= \left(\frac{D_i}{\sum_{i=1}^{m} D_i}\right)\left(\frac{C_j}{\sum_{j=1}^{n} C_j}\right)\min\left\{\sum_{i=1}^{m} D_i, \sum_{j=1}^{n} C_j\right\}
$$

$$
= \left(\frac{D_i}{\sum_{i=1}^{m} D_i}\right)\left(\frac{C_j}{\sum_{j=1}^{n} C_j}\right) MF(\mathcal{F}, \mathbf{D}),
$$

so that

$$
\sum_{i=1}^{m}\sum_{j=1}^{m} x_{ij}^*(\mathbf{D}) = \sum_{i=1}^{m}\left(\frac{D_i}{\sum_{i=1}^{m} D_i}\right)\sum_{j=1}^{n}\left(\frac{C_j}{\sum_{j=1}^{n} C_j}\right) MF(\mathcal{F}, \mathbf{D})
$$

$$
= MF(\mathcal{F}, \mathbf{D}).
$$

It is also easy to see that $x^*(\mathrm{D})$ is a feasible solution to the max-flow problem. $\quad\square$

This result has numerous implications for the process flexibility problem. For example, suppose that the demand D_i is bounded between the lower bound $LE(D_i)$ (>0) and the upper bound $UE(D_i)$ (with $U \geqslant L > 0$) for all $i = 1, \ldots, n$.[5] We assume that the capacity is configured properly so that $\sum_j C_j \leqslant U(\sum_{i=1}^{n} E(D_i))$. For the max-flow problem under the full flexibility structure,

$$
x_{ij}^*(\mathbf{D}) = \frac{D_i C_j}{\max\{\sum_{i=1}^{m} D_i, \ \sum_{j=1}^{n} C_j\}} \leqslant \frac{UE(D_i)C_j}{L(\sum_{i=1}^{n} E(D_i))}
$$

and

$$
E_{\mathbf{D}}(x_{ij}^*(\mathbf{D})) = E_D\left[\frac{D_i C_j}{\max\{\sum_{i=1}^{m} D_i, \ \sum_{j=1}^{n} C_j\}}\right]
$$

$$
\geqslant \frac{LE(D_i)C_j}{U(\sum_{i=1}^{n} E(D_i))}.
$$

In this case, we have a conservative estimate that

$$
x_{ij}^*(\mathbf{D}) \leqslant \left(\frac{U}{L}\right)^2 E_{\mathbf{D}}(x_{ij}^*(\mathbf{D})).
$$

Using $\lambda = U^2/L^2$ in Theorem 3, we obtain the following corollary:

COROLLARY 2. *If demand D_i is bounded between $LE(D_i)$ and $UE(D_i)$ for all i, then there exists a process structure $\mathcal{G}$ with cardinality $|\mathcal{G}| = O((n + m)U^2/(\epsilon L^2))$, such that*

$$
\mathrm{E}[MF(\mathcal{G}, \mathbf{D})] \geqslant (1 - \epsilon)\mathrm{E}[MF(\mathcal{F}, \mathbf{D})].
$$

Note that the process structure $\mathcal{G}$ uses only $O(U^2(n+m)/L^2\epsilon)$ links, whereas the full flexibility structure $\mathcal{F}$ has nm

links. Thus, the number of links in $\mathcal{G}$ is much smaller than the full flexibility system when $U^2/(L^2\epsilon) \ll \min(m, n)$. In the event that m and n are both moderate and ϵ is small or U/L is large, then the above result is not useful. Note that the bound $U^2/(L^2\epsilon)$ is needed to facilitate the proof to our main result, and is not indicative of the actual number of links needed to attain the desired performance. Our numerical experiments (see the next section) seem to indicate that the performance of the sparse structure is not sensitive to the choice of the parameters U and L.

5. Numerical Results: Capacity Pooling Problem

In this section, we use the sampling-based approach to study a partial capacity pooling problem. Consider a manufacturer using n plants with capacity C_i ($i = 1 \ldots n$) to meet the (random) demand D_i from n different regions. Originally, each plant is supposed to serve the demand from a single (dedicated) region. To increase the manufacturer's service level, a capacity pooling method can be adopted by "pooling" the capacity of all the plants; that is, each plant can use its unused capacity to meet the demand from other regions if there is a need to. This problem can be reduced to a variant of the process flexibility problem, where there are n plants and n products. Each plant i has capacity $(C_i - D_i)^+$ (the spare capacity at plant i), which can be used to meet the demand for other products. Each product has demand $(D_i - C_i)^+$ (unfilled demand at region i). Note that in this case, both capacity and demand are random parameters in our problem, and $(C_i - D_i)^+ \times (D_i - C_i)^+ = 0$.

From Theorem 3 and the analysis in the previous sections, the existence of a sparse support structure for the capacity pooling problem is guaranteed by the following condition:

$$x_{ij}^*(\mathbf{D}) = \frac{(D_i - C_i)^+ (C_j - D_j)^+}{\max\{\sum_{i=1}^n (D_i - C_i)^+,\ \sum_{j=1}^n (C_j - D_j)^+\}}$$

$$\leqslant \lambda E_D\left[\frac{(D_i - C_i)^+ (C_j - D_j)^+}{\max\{\sum_{i=1}^n (D_i - C_i)^+,\ \sum_{j=1}^n (C_j - D_j)^+\}}\right]$$

almost surely for some $\lambda > 1$, and for all i, j.

5.1. Numerical Study

We use the data provided by Jordan and Graves (1995) to build our example of the capacity pooling problem. Consider a manufacturer with 16 plants serving 16 dedicated regions. Demand from each region is uncertain and normally distributed with standard deviation $\sigma_i = 0.4E(D_i)$. Each plant's capacity is equal to the expected demand of the region it serves, as shown in Figure 4. The regions can be divided into three subgroups: Regions 1 to 6, 7 to 13, and 14 to 16. The demands of regions in the same subgroup are pairwise correlated with a correlation coefficient of 0.3. There are no correlations between the demands of regions in different subgroups.

Figure 4. Sampled capacity pooling network with 32 arcs.
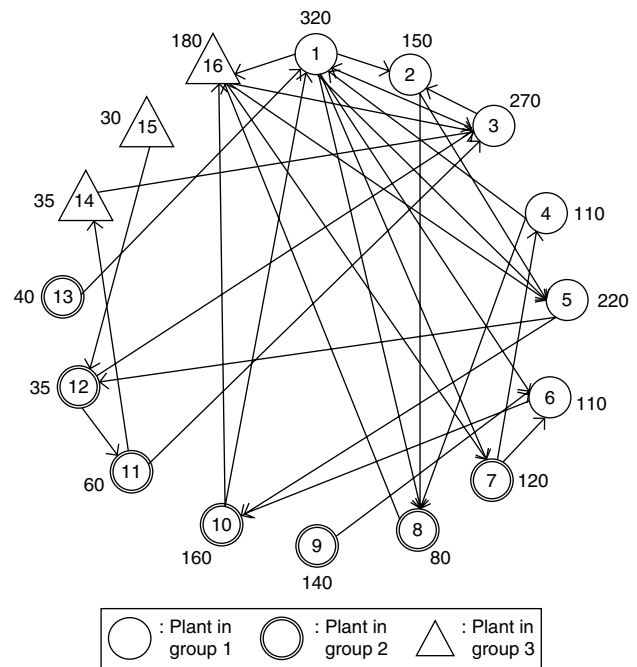
We focus on how the capacity pooling structure can be suitably designed. Here, we consider only the unidirectional pooling structure (cf. Figure 4, where plants are connected via directed arcs). An arc from plant i to j means that plant i can use its spare capacity to help plant j (serve region j). However, plant j cannot help plant i unless there is another arc (j, i) connecting them.

It is obvious that the complete pooling structure (where a plant can share its capacity with any other plant) will achieve the maximum savings. However, it would increase the complexity of the operations, as more linkages between the plants have to be pre-arranged. Therefore, a sparse partial-pooling structure is preferred.

The proof of Theorem 3 indicates that, in our sampling approach, we should set the probability of selecting arc (i, j) to be

$$p_{ij} = \frac{E_{\mathbf{D}}(x_{ij}^*(\mathbf{D}))}{\sum_{k, l} E_{\mathbf{D}}(x_{kl}^*(\mathbf{D}))}.$$

Note that

$$E_{\mathbf{D}}(x_{ij}^*(\mathbf{D})) = E_{\mathbf{D}}\left[\frac{(C_i - D_i)^+ (D_j - C_j)^+}{\max\{\sum_{i=1}^m (C_i - D_i),\ \sum_{j=1}^n (D_j - C_j)\}}\right].$$

In this example, $m = n = 16$. While these values can be computed in closed form for several demand distributions, we use instead the following simulation-based approach to approximate these values.

Step 1: Monte Carlo Simulation

1. For each region, generate 100 realizations of demand based on the specified demand distribution. Let D^k denote the demand generated in the kth instance.

2. Estimate $E_D(x_{ij}^*(\mathrm{D}))$, using

$$\tilde{E}_D(x_{ij}^*(\mathrm{D}))$$
$$= \sum_{k=1}^{100} \left[\frac{(C_i - D_i^k)^+ (D_j^k - C_j)^+}{\max\{\sum_{i=1}^{16}(C_i - D_i^k), \ \sum_{j=1}^{16}(D_j^k - C_j)\}} \right] / 100.$$

3. Estimate p_{ij}, using

$$\tilde{p}_{ij} = \frac{\tilde{E}_D(x_{ij}^*(Q))}{\sum_{k=1}^{m} \sum_{l=1}^{n} \tilde{E}_D(x_{kl}^*(\mathrm{D}))}.$$

We next sample the arcs using the estimated probabilities $\{\tilde{p}_{ij}\}$. However, to ensure that no node will be disconnected from the rest of the network, we first generate an arc into each node. Furthermore, to avoid sampling an arc twice, we remove the arc from the sampling experiment once it has been selected. We do this by normalizing the corresponding sampling probability to zero. The procedure is described next.

Step 2: Sampling Arcs

1. For $k = 1, \ldots, n$, generate a uniform random number $U_k \in (0, \sum_{j=1}^{n} \tilde{p}_{kj})$. If

$$\sum_{j=0}^{l-1} \tilde{p}_{kj} < U_k \leqslant \sum_{j=0}^{l} \tilde{p}_{kj} \quad \text{for some } l = 1, \ldots, n,$$

let $\tilde{p}_{kl} = 0$ and add arc (k, l) to the network.

2. Arrange the arcs in lexicographical order, and let $\hat{p}_{n(i-1)+j} = \tilde{p}_{ij}$.

3. Generate a random number $U_k \in (0, \sum_{i=1}^{mn} \hat{p}_i)$. If

$$\sum_{i=1}^{l-1} \hat{p}_i < U \leqslant \sum_{i=1}^{l} \hat{p}_i, \quad l = 1, 2, \ldots, mn,$$

let $\hat{p}_l = 0$ and add arc l to the network.

4. Repeat Step 3 until the number of arcs sampled reaches N.

We can generate several structures with N links by repeating Step 2. Here, 100 pooling structures are sampled.

Step 3: Structure Evaluation. The sampled structures are graded according to their ability to match supply with demand. We use a simulation here to determine the performance of each sampled structure. We generate another 100 demand scenarios, and evaluate each structure based on this set of demand. We choose the structure with the best performance.

For any given N, we can use the above sampling heuristic to design a good capacity pooling network. Figure 4, for example, is a network we obtained from this sampling-based approach, using only 32 links. The network obtained

**Figure 5.** Expected shared capacity as flexibility increases.
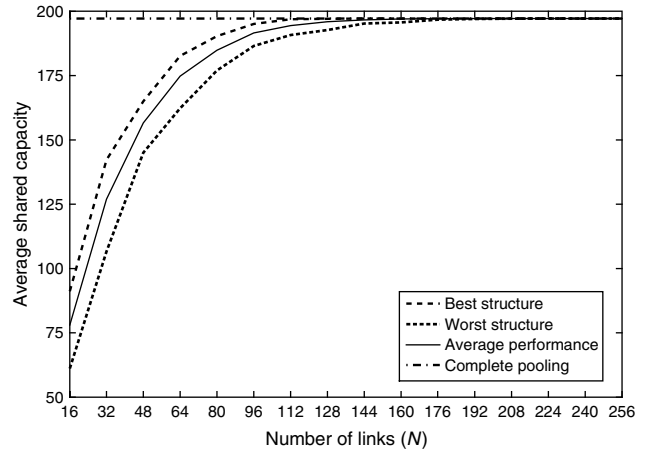
exhibits characteristics of a good capacity pooling network structure: (i) the plants with a higher average capacity should ideally be linked to more other nodes, and (ii) plants within the same group are positively correlated, and hence ideally there should only be a small number of pooling arcs within them; whereas plants in different groups tend to have more arcs linked among them because their demands are independent.

Figure 5 plots the performance of the structures obtained (in terms of average transshipped quantity) as the number of arcs increases. For each N, we plot the performance attained by the best structure, the worst structure, and the average performance of all 100 sampled structures. As shown in Figure 5, the performance gap among these three cases is very small and quickly converges to zero as the number of links N increases. This suggests that the sampling heuristic is quite stable and robust, and the performance of any sampled structure is acceptable as long as N is sufficiently large.

Another observation from Figure 5 is that, as the number of links increases, the marginal contribution of additional links diminishes for all three cases—the best, the worst, and the average case. For the best case with 16 links, for example, the expected shared capacity is only 46.7% of the complete pooling structure. After increasing the number of arcs to 96, however, the shared capacity is already close to 98.5% of the complete pooling structure. Note that the complete pooling structure has up to 240 arcs. Because the capacity pooling structure obtained using the sampling approach gives a lower bound to the performance of the optimal structure, it is expected that the optimal performance-flexibility curve should be even steeper. Nevertheless, our results further validate the fact that a partial-capacity pooling structure can achieve a performance close to that of the complete pooling system.

We also evaluate the effect of demand truncation on the performance of the sparse process structures. More specifically, instead of considering the demand in $[0, \infty)$, we

Figure 6. Expected shared capacity ratio of a sparse structure to the complete pooling for various k values.
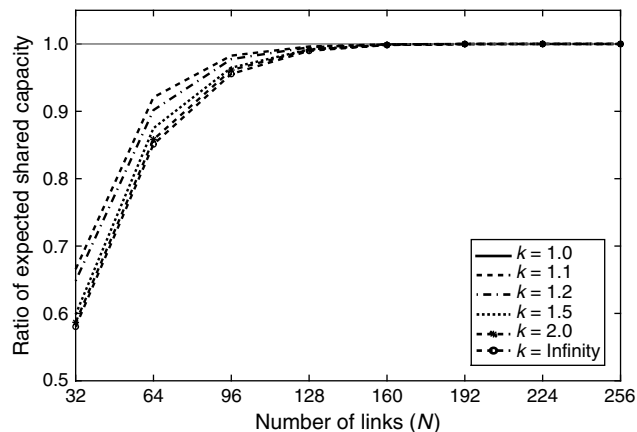
truncate the demand into the range $[E(D_i)/k, kE(D_i)]$ for product i. Figure 6 plots the expected performance of the process structures obtained by the sampling heuristic for various values of k.

When $k = 1$, the demand for each product is deterministic and exactly equals the capacity of the plant. Therefore, both the dedicated structure and the complete pooling structure have zero shared capacity. When $k = \infty$, the demand follows a truncated normal distribution in $[0, \infty)$. It appears that the choice of k has little impact on the performance of the sparse structures. For example, as shown in Figure 6, the performance-flexibility curve when $k = 2$ is already very close to that when $k = \infty$.

REMARK. The numerical study considers a balanced system, i.e., the total capacity equals the expected total demand. We have also extended our numerical study to an unbalanced system, i.e., a system with some safety capacity. Details are available upon request. From our studies, we observed that the additional safety capacity would increase the expected shared capacity for both the complete pooling structure $\mathscr{F}$ and the sampled structure $\mathscr{G}$, but would not significantly affect the relative performance of the sampled structure (i.e., the ratio of $E[MF(\mathscr{G}, \mathbf{D})]$ to $E[MF(\mathscr{F}, \mathbf{D})]$).

6. Conclusion

In this paper, we provide analytical results of the performance of the well-known chaining strategy, and we identify a class of conditions to guarantee that a sparse process structure can perform nearly as well as the dense full flexibility system. For example, when the demand distribution is uniform or normal, our method returns an efficiency measurement ACE of at least 58% and 70%, respectively. The ratio of the expected sales from chaining to the full flexibility system is even more impressive: at 89% for uniform distribution and around 96% for normal distribution, with a CV of at most 0.33. This partially confirms the popular

belief in the community that chaining already reaps a substantial portion of the benefits of full flexibility.

The proposed method works even for the unbalanced (i.e., capacity not equal to mean demand) and asymmetrical (i.e., demand not symmetrical around its mean) cases. This demonstrates the applicability of our result to an even wider range of demand distributions. We also study the process flexibility design problem when supply and demand are nonidentical. We show that partial flexibility structures, properly designed, can already accrue most of the benefits of the full flexibility system. The same insight extends to many other areas, such as capacity pooling structure design. In addition, our numerical study on the capacity pooling problem strongly supports our theoretical results. Our numerical examples, based on a simple sampling scheme, show how the sampling probabilities can be approximated and used to construct the desired network.

One of the referees pointed out that it will be useful to examine the bound $CE(n)$ as a function of n. This will allow us to scrutinize the performance of the 2-chain even for the moderate-size process flexibility problem. This problem appears difficult because we need to analyze the second-order effect of expected maximum flow in a 2-chain and a fully flexible system, respectively. The result is likely to depend on the form of the demand distribution. It is also not known whether $CE(n)$ is monotone in n for all demand distributions, although we have observed this trend in our numerical experiments.

The problem we studied in this paper is part of a broader class of the supply chain network design problem in which there are often associated concerns and new complications, such as (i) fixed cost in installing each link, (ii) different unit revenue for different product, and/or (iii) different products might consume different amounts of the plants' capacity. Such problems are often handled by numerical methods (using two-stage stochastic programming, for example), with little insight to offer on the qualitative features of a good network structure. The insights presented in this paper are in a way a small step in this direction—under appropriate assumptions, any sparse random network (sampled using appropriate distribution) can be expected to perform relatively well. Furthermore, a 2-chain should work well in the case of balanced and identical expected demand and supply environment, and we have provided rigorous bounds to assess the performance of 2-chain in such a setting. Extending these qualitative results to the more general supply chain network design problem is daunting and challenging, in part because we do not know of a good way to estimate the performance of a given process structure under random demand. In fact, to the best of our knowledge, finding a good method to estimate the expected maximum flow in an arbitrary bipartite network in random environment is already an open problem.

There are several other directions to extend the results in this paper. It will be interesting to consider price-responsive demands and formulate the manufacturer's problem as one of maximizing profits. It would also be interesting to

look at this problem in an oligopolistic framework and to examine the impact of pricing and partial flexibility on the strategic responses of the players in the market. We leave these issues for future research.

Endnotes

1. This structure is also known as the 2-chain (because each plant is connected to two products, and vice versa) or the long chain (because it is the longest possible 2-chain).

2. By sparse, we mean that the structure uses far fewer links relative to the dense process structure utilized in the full flexibility system.

3. Our technique can be modified to handle cases when the demand is not symmetrical about the mean. We focus our analysis on this case merely for ease of exposition.

4. Suppose that $P_0 > 0$. Let $P_0' = 0$, $P_x' = P_x/(1 - P_0)$ $\forall x \neq 0$. It follows that $E[\tau'] = (1 - P_0)E[\tau]$ and $E[\psi'] = (1 - P_0)E[\psi]$.

5. Consider a normal demand distribution with $\sigma = 0.2\mu$, truncated at $\mu - 3\sigma = 0.4\mu$ and $\mu + 3\sigma = 1.6\mu$. In this case, $L = 0.4$ and $U = 1.6$.

Acknowledgments

References

Akşin, O., F. Karaesmen. 2007. Characterizing the performance of process flexibility structures. *Oper. Res. Lett.* **35**(4) 477–484.

Assembly Magazine. 2005. Chrysler vice president inducted into Shingo Academy. Accessed May 2, 2005, http://assemblymag.com.

Bish, E., Q. Wang. 2004. Optimal investment strategies for flexible resources, considering pricing and correlated demand. *Oper. Res.* **52**(6) 954–964.

Bish, E., A. Muriel, S. Biller. 2005. Managing flexible capacity in a make-to-order environment. *Management Sci.* **51**(2) 167–180.

Boudette, N. 2006. Chrysler gains edge by giving new flexibility to its factories. *Wall Street Journal* (April 11).

Brusco, M., T. Johns. 1998. Staffing a multiskilled workforce with varying levels of productivity: An analysis of cross-training policies. *Decision Sci.* **29**(2) 499–515.

Calafiore, G., M. C. Campi. 2005. Uncertainty convex programs: Randomized solutions and confidence levels. *Math. Programming* **102**(1) 25–46.

Chou, M. C., C. P. Teo, H. Zheng. 2007. Process flexibility revisited: Graph expander and its applications. Working paper, National University of Singapore.

de Farias, D. P., B. Van Roy. 2004. On constraint sampling in the linear programming approach to approximate dynamic programming. *Math. Oper. Res.* **29**(3) 462–478.

DesMarteau, K. 1999. [[TC].sup.2] Leading the way in changing times. *Bobbin* **41**(2) 48–54.

Fine, C., R. Freund. 1990. Optimal investment in product-flexible manufacturing capacity. *Management Sci.* **36**(4) 449–466.

Graves, S., B. Tomlin. 2003. Process flexibility in supply chains. *Management Sci.* **49**(7) 907–919.

Gurumurthi, S., S. Benjaafar. 2004. Modeling and analysis of flexible queueing systems. *Naval Res. Logist.* **51** 755–782.

Hopp, W., E. Tekin, M. Van Oyen. 2004. Benefits of skill chaining in serial production lines with cross-trained workers. *Management Sci.* **50**(1) 83–98.

Iravani, S., M. Van Oyen, K. Sims. 2005. Structural flexibility: A new perspective on the design of manufacturing and service operations. *Management Sci.* **51**(2) 151–166.

Jordan, W., S. Graves. 1995. Principles on the benefits of manufacturing process flexibility. *Management Sci.* **41**(4) 577–594.

Karp, R. M., R. Motwani, N. Nisan. 1993. Probabilistic analysis of network flow algorithms. *Math. Oper. Res.* **18**(1) 71–97.

LaSorda, T. 2006. Chrysler Group Update. January 12. Private communication.

Lyons, R., R. Pemantle, Y. Peres. 1999. Resistance bounds for first-passage percolation and maximum flow. *J. Combin. Theory Ser. A* **86**(1) 158–168.

McCutcheon, D. 2004. Flexible manufacturing: IBM's Bromont semiconductor packaging plant. *Canadian Electronics* **19**(7) 26.

Muriel, A., A. Somasundaram, Y. Zhang. 2006. Impact of partial manufacturing flexibility on production variability. *Manufacturing Service Oper. Management* **8**(2) 192–205.

Sethi, A., S. Sethi. 1990. Flexibility in manufacturing: A survey. *Internat. J. Flexible Manufacturing Systems* **2** 289–328.

U.S. Department of Commerce. 2006. *The Road Ahead for the U.S. Auto Industry*. U.S. Department of Commerce, Washington, D.C.

Van Biesebroeck, J. 2007. Complementarities in automobile production. *J. Appl. Econometrics* **22**(7) 1315–1345.

Van Mieghem, J. A. 1998. Investment strategies for flexible resources. *Management Sci.* **44**(8) 1071–1078.

Wall, M. 2003. Manufacturing flexibility. *Automotive Industries* **183**(10) 44–45.