Production, Manufacturing and Logistics

Hub and Chain: Process Flexibility Design in Non-Identical Systems Using Variance Information

Geoffrey A. Chua*, Shaoxiang Chen, Zhiguang Han

Nanyang Business School, Nanyang Technological University, Singapore 639798, Singapore

ARTICLE INFO

ABSTRACT

In multi-product multi-plant manufacturing systems, process flexibility is the ability to produce different types of products in the same manufacturing plant or production line. While several design methods and flexibility indices have been proposed in the literature on how to design process flexibility, most of the insights generated are focused on identical production systems whereby all plants have the same capacity and all products have identically distributed demands. In this paper, we examine the process flexibility design problem for non-identical systems. We first study the effect of non-identical demand distributions on the performance of the well-known long chain design, and discover three interesting insights: (1) products with low demand mean will create a bottleneck effect, (2) products with low demand variance will result in inefficient utilization of flexibility links, and (3) long chain efficiency decreases in demand variance of any product, hence the need to provide this product with access to more capacity. Using these insights, we develop the variance-based hub-and-chain method (VHC), a simple and graphically intuitive method which decomposes the long chain into smaller chains, one of which will serve as a hub to which the other chains will be connected. Numerical tests show that VHC outperforms the long chain by 15% on average and outperforms the constraint sampling method by 38% on average. Lastly, we implement VHC on a case study in the edible oil industry in China and find substantial benefits. We then summarize with some managerial insights.

1. Introduction

"*If the last few years have taught us anything, it is that customer wants and needs can change quickly – much more quickly than we have been equipped to respond to efficiently in the past. At Michigan Assembly, we will achieve a level of flexibility we don't have in any other plant around the world, which will allow us to meet shifting consumer preferences in real time.*" - Jim Tetreault, Ford Vice President of North America Manufacturing (Ford Media, August 2010).

Ford VP Jim Tetreault declared those statements less than two years removed from the disastrous year of 2008, which saw Ford absorb a \$14.68B loss while watching Asian rivals Toyota and Honda rake in billions in profits. By 2010, Ford has started staging its much talked about turnaround on the way to accumulating over \$35B in profits in a span of four years. While pundits rave about the new car models (e.g. Fusion, Fiesta) or new car technologies (e.g. Sync) launched during that period, it is no secret that the key

reason for this turnaround is Ford becoming a more nimble and flexible automaker (LeBeau, 2010). Indeed, the company's Michigan Assembly Plant has undergone an industry-first \$550M renovation, equipping the plant not only to make a variety of models, but also different powertrains, such as conventional and electric (Woodyard, 2011). Not to mention the earlier \$200M investment in 2009 to retool Ford's Kentucky Truck Plant with a flexible body shop capable of producing a wide range of trucks and SUVs (Ford Media, April 2009). As Dave Cole, chairman emeritus for the Center for Automotive Research, puts it, "Lean isn't good enough anymore. The new reality requires being both lean and flexible." (Woodyard, 2011).

As flexibility gains importance as a strategic priority, it has also attracted strong attention from the academic community. Because flexibility comes in various forms, the earlier works (e.g. Buzacott, 1982; Gupta & Goyal, 1989; Sethi & Sethi, 1990) were mostly reviews and taxonomies, focused on understanding the nature of flexibility, developing measures and evaluation criteria, as well as categorizing the various types of flexibility. Subsequently, research focus has shifted to the strategic issue of flexibility design (i.e. what kind of flexibility to employ and how much), and has centered on *process flexibility*, defined as "the ability to produce

* Corresponding author. Tel.: +65 67906140.
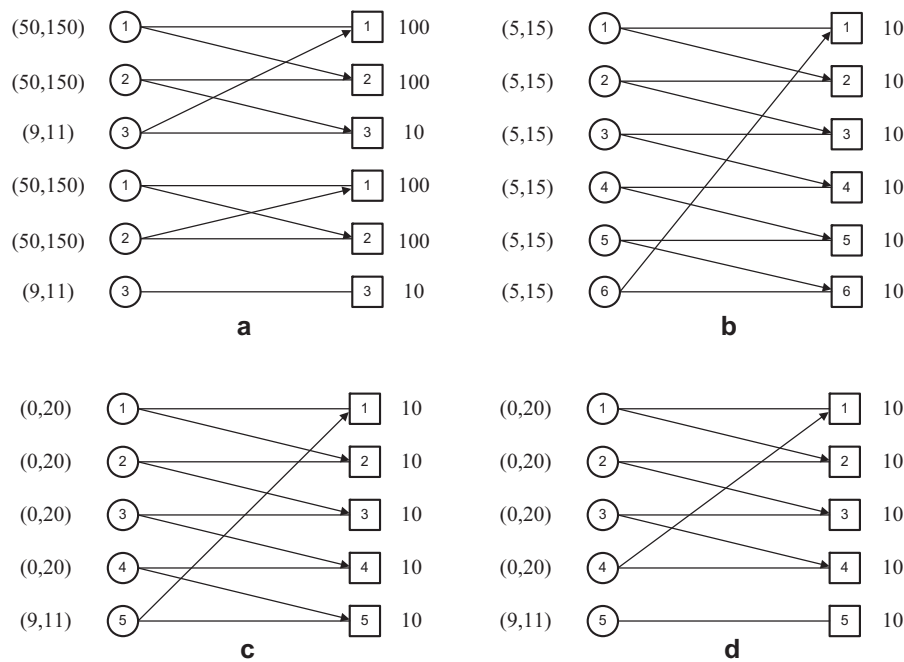 E-mail address: gbachua@ntu.edu.sg (G.A. Chua).

Fig. 1. Performance of long chains: Examples.

different types of products in the same manufacturing plant or on the same production line". Specifically, the *process flexibility design problem* is as follows: given a set of products with random demands, a set of plants with fixed capacities and a limited number of product-plant links indicating the capability of said plant to produce said product, how should one deploy these links such that system performance (i.e. expected maximum flow) is maximized? Similar to the auto-making industry, such is also the problem facing one of the largest producers of edible oil products in China whose plants have various filling and packaging lines that are used to manufacture products of various shapes and sizes. This is precisely one of the reasons that motivated us to study the process flexibility design problem. We shall report our analysis and findings for this company in Section 5.

One of the most important ideas on this topic is the concept of the *chaining strategy*. Jordan and Graves (1995) were the first to observe that a partially flexible system using only a few flexibility links, if configured the right way, can already achieve most of the benefits of full flexibility. Their proposed chaining strategy is essentially a set of guidelines on how to achieve this high-performing partially flexible configuration. Suppose we have a *balanced* production system where (1) the number of products equals the number of plants, and (2) the mean demand of each product is equal to the capacity of its primary plant.[1] Assume further that this system is *identical* such that (1) all product demands are independent and identically distributed, and (2) all plants have the same capacity. In this setting, the chaining strategy will result in a flexibility design where each product will be connected to two plants and each plant will be linked to two products such that there always exists a path connecting any pair of product and plant. This design is known as the *long chain* and Fig. 1(b) shows one example.

A number of papers in the literature examine the balanced and identical system. Aksin and Karaesmen (2007) use a network model to show that the system performance is concave in the degree of flexibility. Chou, Chua, Teo, and Zheng (2010b) develop a random walk approach to characterize asymptotic performance of the chaining strategy as system size grows very large, while Chou, Chua, Teo, and Zheng (2011) use graph expanders to examine the worst-case maximum flow. Simchi-Levi and Wei (2012) use the supermodularity property to prove the optimality of the long chain among systems that use the same amount of flexibility as well as to characterize the performance of the long chain for arbitrary system sizes. Chou, Chua, and Teo (2010a) examine the effect of response dimension of flexibility, while Chou, Chua, and Zheng (2014) study the effect of production postponement on process flexibility. Wang and Zhang (2015) provide a distribution-free analysis of the long chain, while Chen, Zhang, and Zhou (2015) extend Chou et al.'s (2011) work to probabilistic graph expanders.

For the general setting of unbalanced and non-identical systems, there are two categories of results. The first provides guidelines such as the Chaining Guidelines by Jordan and Graves (1995), the constraint sampling method by Chou et al. (2010b), and the expansion heuristic by Chou et al. (2011). The second category proposes the use of indices to rank various flexibility designs, such as the JG-Index by Jordan and Graves (1995), the Structural Flexibility (SF) indices by Iravani, Oyen, and Sims (2005), the WS-APL metric by Iravani, Kolfal, and Oyen (2007), and the Expansion Index by Chou, Teo, and Zheng (2008). However, all these methods are either sampling-based approaches (hence time-consuming and result in highly variable outputs) or heuristics that do not explicitly take into account demand variance information.

For unbalanced but identical systems, Deng and Shen (2013) develop the Enhanced Chaining Guideline but this method does not utilize the variance information nor does it work for balanced but non-identical systems. To our best knowledge, our paper is the first attempt to study balanced but non-identical systems in greater detail. This setting allows us to analytically examine the effect of non-identical demand means and variances on the performance of the long chain design. The insights we derive from the analysis will then help us develop a new flexibility design method that explicitly takes advantage of the demand variance information.

To illustrate, we consider a few examples in Fig. 1. In all these examples, each system is represented by a bipartite graph with

---

[1] Note that our definition for "balanced" system deviates from existing terminology with the addition of condition (2). We do so to better capture what is a truly balanced system and to afford us analytical tractability in examining non-identical systems.

a set of products on the left and a set of plants on the right. Each product has a two-point demand distribution with equal probability. For example, a label of (50, 150) denotes the demand has a 50–50 chance of being 50 units or 150 units. On the other hand, each plant has fixed capacity given by the label to its right. System performance is determined by solving a maximum flow problem for each demand realization and taking expectation over all possible maximum flow values. That said, we first consider the two 3×3 systems in Fig. 1(a). The long chain on top generates an expected output of 174.9 units, whereas the second design not only uses one less link but also improves the output to 184.5 units. The reason is because the third product–plant pair becomes a bottleneck when included in the long chain.

For our second example, we refer to Fig. 1(b), and define efficiency of a flexibility design as the ratio of its improvement over the dedicated design relative to full flexibility. We find that if demand distribution for any product changes from (5, 15) to (9, 11), then the efficiency of the long chain improves. This suggests that long chain efficiency decreases in demand variance of any product. Our third example compares Figs. 1(c) and (d), where two flexibility designs are employed on the same balanced but non-identical system. The outputs of the two designs are 39.75 and 39.5 units, respectively. This is only a small difference, and yet Fig. 1(d) can save one link. Meanwhile, if demand for product 5 also becomes (0, 20), the difference between the two designs increases, hence less incentive to save that one link.

In this paper, we manage to prove the three insights generated by the three above examples. Subsequently, we use these insights to devise a variance-based method for designing process flexibility in non-identical systems. The way our method works is to first isolate products with low demand variances into a group that will only employ a dedicated design. Then, we group the remaining products into subgroups according to proximity of demand means. The subgroup should also not exceed a maximum amount of group demand variability, measured by a group coefficient of variation (GCV) that we introduce in this paper. Each of these subgroups will employ the long chain design. Then, a satellite product–plant pair will be selected from each subgroup. The satellite pair of the subgroup with the largest demand means will serve as the hub to which all satellite pairs of the other subgroups will be connected. As such, we refer to our method as the variance-based hub-and-chain method or VHC. To our best knowledge, VHC is the first method that explicitly incorporates demand variance information for non-identical systems. It is also graphically intuitive and simple to explain to managers, because it is reminiscent of the hub-and-spoke network used in airlines but each node is replaced by a small chain. Moreover, our numerical tests show that it is computationally efficient and produces flexibility designs that outperform the long chain by 15% and the designs from constraint sampling method by 38%.

We make the following contributions to the literature. First, we analytically prove the effect of changes in demand means and variances on system performance and long chain efficiency. Second, we discover three interesting insights as follows: (1) products with low demand mean can create a bottleneck effect, (2) products with low demand variance will result in inefficient utilization of flexibility links, and (3) long chain efficiency decreases in demand variance of any product, hence it will be a good idea to provide this product with access to more capacity. Third, we use these insights to develop a simple and graphically intuitive variance-based VHC method for designing simple and flexible non-identical systems. Fourth, we numerically test our method for different demand distributions, and benchmark against existing methods. We do so for both simulated scenarios and a real case study from the edible oil industry.

The rest of the paper is organized as follows. In Section 2, we define the basic model and notations. Section 3 presents our main results on the effect of demand mean and variance on system performance and efficiency. We then present our hub-and-chain method in Section 4, including numerical tests to benchmark against existing methods. In Section 5, we report the results of our case study and discuss some managerial implications. Finally, Section 6 concludes.

2. Model and notation

Consider a manufacturing system with n plants, facing random demand for m products. We model this system through a bipartite graph representation $G = (A, B, \mathscr{G})$, where A denotes the set of products, B denotes the set of plants, and $\mathscr{G}$ denotes the set of edges (i, j) signifying the capability of plant j to produce product i. That is, $A = \{1, 2, \ldots, m\}$, $B = \{1, 2, \ldots, n\}$, and $\mathscr{G} \subseteq A \times B$. Next, we use the vector $\mathbf{D}$ to denote the random demand, and the vector $\mathbf{d}$ for a particular realization of $\mathbf{D}$. We further let D_i and d_i be the respective random demand and particular demand realization for product i. On the other hand, we let the capacity at plant j be fixed at c_j and let the vector $\mathbf{c} = [c_1, \ldots, c_n]^T$. Note that for a given system, every flexibility design can be uniquely represented by the edge set $\mathscr{G}$. Furthermore, given a demand realization $\mathbf{d}$, the maximum sales which can be achieved by a flexibility design $\mathscr{G}$ is defined as

$$Z(\mathbf{d}, \mathscr{G}) = \max \sum_{i \in A} \sum_{j \in B} x_{ij},$$

$$\text{s.t.} \quad \sum_{j \in B} x_{ij} \leq d_i \quad \forall i \in A$$

$$\sum_{i \in A} x_{ij} \leq c_j \quad \forall j \in B$$

$$x_{ij} \geq 0 \quad \forall (i, j) \in \mathscr{G}$$

$$x_{ij} = 0 \quad \forall (i, j) \notin \mathscr{G}$$

where x_{ij} refers to the amount of product i produced by plant j.

Under random demand $\mathbf{D}$, the maximum sales $Z(\mathbf{D}, \mathscr{G})$ is itself also a random variable. Hence, we define the *performance*, also referred to as expected maximum sales, of $\mathscr{G}$ to be $E[Z(\mathbf{D}, \mathscr{G})]$. For ease of exposition and whenever the context is clear, we also use the notation $[\mathscr{G}]$ to denote this quantity. In general, the problem is to generate a flexibility design $\mathscr{G}$ that maximizes $[\mathscr{G}]$ without exceeding a flexibility budget b. That is, we want to solve $\max_{\mathscr{G}: |\mathscr{G}| \leq b} [\mathscr{G}]$. However, this problem involves a stochastic program and is notoriously difficult and time consuming to solve. As such, most works in the literature focus on developing heuristics to solve special classes of this problem.

To classify process flexibility design problems, we introduce the following definitions. We say that the system is *balanced* if (1) $m = n$, and (2) the products and plants can be arranged such that $E[D_i] = c_i$ for all $i \in A$. Moreover, we call the system *identical* if (1) the product demands are IID, i.e. $D_i = D$ in distribution for all $i \in A$ and (2) the plant capacities are all equal, i.e. $c_j = c$ for all $j \in B$. Because of our interest in problems with heterogeneous product demand variances, we focus our study on balanced but non-identical systems.

That said, we let n denote the system size (since $m = n$), and let $\mu_i = E[D_i] = c_i$ be the mean demand and σ_i^2 be the demand variance for product i for all $i = 1, 2, \ldots, n$. We also define a few classes of flexibility designs for balanced systems of size $n \geq 2$. The dedicated design is defined as $\mathscr{D}_n = \{(i, i) | i = 1, 2, \ldots, n\}$; the open chain design is defined as $\mathscr{L}_n = \mathscr{D}_n \cup \{(i, i+1) | i = 1, 2, \ldots, n-1\}$; the long chain design is defined as $\mathscr{C}_n = \mathscr{L}_n \cup \{(n, 1)\}$; and the full flexibility design is defined as $\mathscr{F}_n = \{(i, j) | i, j = 1, 2, \ldots, n\}$. Our

goal in this paper, as in the literature, is to analyze the performance of these flexibility designs as well as to develop new methods to improve on these designs, particularly for non-identical systems.

To this end, we further define two performance measures that will be useful in developing our new flexibility design method. We define *system potential* as the difference between the performance of full flexibility and that of the dedicated design. That is, we let system potential be $p(\mathbf{D}) = [\mathscr{F}_n] - [\mathscr{D}_n]$. Next, we define *efficiency* of a flexibility design as the ratio of the difference between the performance of the given flexibility design and that of the dedicated design over system potential. That is, we let the efficiency of $\mathscr{G}$ be $f(\mathbf{D}, \mathscr{G}) = \frac{[\mathscr{G}] - [\mathscr{D}_n]}{[\mathscr{F}_n] - [\mathscr{D}_n]}$. For example, $f(\mathbf{D}, \mathscr{C}_n)$ is equivalent to the chaining efficiency used in Chou et al. (2010b).

3. Analysis and main results

In this section, we examine the effect of changes in demand means and variances on system potential and flexibility design efficiency. The results will then be used to obtain insights on how to develop methods that can generate good flexibility designs.

For analytical purposes, we consider two types of demand distributions; namely, the normal distribution and the symmetric three-point distribution. For normal distribution, we assume that for $i = 1, 2, \ldots, n$, $D_i \sim N(\mu_i, \sigma_i)$ with $3\sigma \leq \mu$ so that negative demand values are negligible. This is a reasonable assumption and is common in the literature. For symmetric three-point distribution, we assume that for $i = 1, 2, \ldots, n$,

$$D_i = \begin{cases} \mu_i - \eta_i, & \text{with probability } t_i \\ \mu_i, & \text{with probability } 1 - 2t_i \\ \mu_i + \eta_i, & \text{with probability } t_i \end{cases}$$

where $0 \leq \eta_i \leq \mu_i$ and $0 \leq t_i < \frac{1}{2}$. This is likewise a reasonable assumption because it characterizes products that have peak, shoulder, and off-peak demands. Moreover, in their study of distribution-free models, Deng and Shen (2014) show that if only the marginal mean and covariance matrix are known while the marginal distribution is unknown, the worst-case distribution when demands are independent is a three-point distribution for every product demand. That said, higher values of η_i or t_i imply higher demand variance. For both distribution types, we also assume that the product demands are independent.

With a slight abuse of notation, we let system potential $p(\mathbf{D})$ be changed to $p(\mu, \sigma)$ for normal distribution and $p(\mu, \eta, \mathbf{t})$ for symmetric three-point distribution. Similarly, we change flexibility design efficiency to $f(\mu, \sigma, \mathscr{G})$ and $f(\mu, \eta, \mathbf{t}, \mathscr{G})$ whenever appropriate. We are now ready to present our results.

Theorem 1. *Consider a balanced but non-identical system of size n such that $E[D_i] = c_i$ for all $i \in A$. For normal distribution, the system potential $p(\mu, \sigma)$ is increasing and concave in σ_k, but independent of μ_k. For symmetric three-point distribution, the system potential $p(\mu, \eta, \mathbf{t})$ is increasing and concave in η_k, increasing linearly in t_k, but independent of μ_k.*

Proof. By definition of system potential, we have

$$p(\mathbf{D}) = [\mathscr{F}_n] - [\mathscr{D}_n]$$

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

$$= \sum_{i=1}^{n} E[(D_i - \mu_i)^+] - E\left[\left(\sum_{i=1}^{n} D_i - \sum_{i=1}^{n} \mu_i \right)^+ \right]$$

For normal distribution, it is well known that if $X \sim N(0, \sigma)$, then $E[X^+] = \frac{\sigma}{\sqrt{2\pi}}$. Therefore, we can obtain $p(\mu, \sigma) = \frac{\sum_{i=1}^{n} \sigma_i}{\sqrt{2\pi}} -$

$\frac{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}{\sqrt{2\pi}}$. Hence, it follows that $\frac{\partial p(\mu, \sigma)}{\partial \sigma_k} = \frac{1}{\sqrt{2\pi}}\left(1 - \frac{\sigma_k}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}\right) \geq 0$, $\frac{\partial^2 p(\mu, \sigma)}{\partial \sigma_k^2} = -\frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}\left(1 - \frac{\sigma_k^2}{\sum_{i=1}^{n} \sigma_i^2}\right) < 0$, and $\frac{\partial p(\mu, \sigma)}{\partial \mu_k} = 0$. For three-point distribution, we take $k = n$ without loss of generality. It can be shown that

$$[\mathscr{F}_n] - [\mathscr{D}_n] = [\mathscr{F}_{n-1}] - [\mathscr{D}_{n-1}]$$

$$+ E\left[\min\left[\left(\sum_{i=1}^{n-1} D_i - \sum_{i=1}^{n-1} \mu_i \right)^+, \eta_n \right] \right] \cdot t_n$$

$$+ E\left[\min\left[\left(\sum_{i=1}^{n-1} \mu_i - \sum_{i=1}^{n-1} D_i \right)^+, \eta_n \right] \right] \cdot t_n$$

Since $[\mathscr{F}_{n-1}] - [\mathscr{D}_{n-1}]$ is independent of η_n, t_n, and μ_n, it suffices to examine the last two expectation terms. It is easy to see that these terms are increasing and concave in η_n, increasing linearly in t_n, and independent of μ_n. Hence, the results follows. □

Theorem 1 says that the potential or maximum value that can be achieved by investing in flexibility becomes larger whenever the demand variance for any product becomes larger. This makes sense because risk pooling is typically most beneficial for systems with more uncertainty. In addition, the theorem also says that potential is independent of changes in the demand mean of any product if capacity is adjusted to match changes in mean demand.

Theorem 2. *Consider a balanced but non-identical system of size n such that $E[D_i] = c_i$ for all $i \in A$. For normal distribution, the efficiency $f(\mu, \sigma, \mathscr{C}_n)$ of the long chain design is increasing in μ_k. For symmetric three-point distribution, the efficiency $f(\mu, \eta, \mathbf{t}, \mathscr{C}_n)$ of the long chain design is also increasing in μ_k.*

Proof. By Theorem 1, it suffices to show that $[\mathscr{C}_n] - [\mathscr{D}_n]$ is increasing in μ_k for both distribution types. Suppose $[\mathscr{C}_n]'$ and $[\mathscr{D}_n]'$ are the respective performances of the long chain design and the dedicated design when mean demand for product k (and capacity for plant k) increases from μ_k to $\mu_k + \delta$ where $\delta > 0$. We want to show that $[\mathscr{C}_n]' - [\mathscr{D}_n]' \geq [\mathscr{C}_n] - [\mathscr{D}_n]$, or equivalently, $[\mathscr{C}_n]' - [\mathscr{C}_n] \geq [\mathscr{D}_n]' - [\mathscr{D}_n]$.

By definition, we know that $[\mathscr{D}_n] = \sum_{i=1}^{n} E[\min(D_i, \mu_i)]$, and

$$[\mathscr{D}_n]' = \sum_{i=1, i \neq k}^{n} E[\min(D_i, \mu_i)] + E[\min(D_k + \delta, \mu_k + \delta)] = [\mathscr{D}_n] + \delta$$

Hence, it remains to be shown that $[\mathscr{C}_n]' - [\mathscr{C}_n] \geq \delta$. For every realization $\mathbf{d}$, we let $\mathbf{d}' = [d_1, \ldots, d_{k-1}, d_k + \delta, d_{k+1}, \ldots, d_n]^T$, and $x_{ij}(\mathbf{d}, \mathscr{C}_n)$ and $x_{ij}^*(\mathbf{d}, \mathscr{C}_n)$ be a feasible solution and an optimal solution, respectively, for $Z(\mathbf{d}, \mathscr{C}_n)$. Since $x_{kk}(\mathbf{d}', \mathscr{C}_n) = x_{kk}^*(\mathbf{d}, \mathscr{C}_n) + \delta$ and $x_{ij}(\mathbf{d}', \mathscr{C}_n) = x_{ij}^*(\mathbf{d}, \mathscr{C}_n)$, for all $i, j \neq k$ is a feasible solution to $Z(\mathbf{d}', \mathscr{C}_n)$, it follows that $[\mathscr{C}_n]' - [\mathscr{C}_n] \geq \delta$.

To rule out $[\mathscr{C}_n]' - [\mathscr{C}_n] = \delta$, it suffices that there exist a product–plant pair k and a demand realization $\mathbf{d}$ such that $d_{k-1} > \mu_{k-1}$, $d_k = \mu_k$ and $\mu_{k+1} > d_{k+1}$, but $\mu_k < \min(d_{k-1} - \mu_{k-1}, \mu_{k+1} - d_{k+1})$. In this case, an increase from μ_k to $\mu_k + \delta$ will result in a performance increase for the long chain design that is strictly larger than δ. □

Theorem 2 is an important result as it states that long chain efficiency (essentially how well the long chain approximates full flexibility) is increasing in the demand mean of any product. This implies that a product with low demand mean can be detrimental to the performance of the long chain design. The reason is that such product might become a bottleneck in allowing idle capacity in higher-numbered plants to serve extra demand for lower-numbered products.

The next result follows easily from Theorem 2, and provides upper and lower bounds on long chain efficiency.

Corollary 1. *Consider a balanced but non-identical system of size n such that $E[D_i] = c_i$ for all $i \in A$. For normal distribution, the efficiency of the long chain design is bounded as follows: $f(\min_i \mu_i \cdot \mathbf{1}, \boldsymbol{\sigma}, \mathscr{C}_n) \le f(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathscr{C}_n) \le f(\max_i \mu_i \cdot \mathbf{1}, \boldsymbol{\sigma}, \mathscr{C}_n)$. For symmetric three-point distribution, the efficiency of the long chain design is similarly bounded as follows: $f(\min_i \mu_i \cdot \mathbf{1}, \boldsymbol{\eta}, \mathbf{t}, \mathscr{C}_n) \le f(\boldsymbol{\mu}, \boldsymbol{\eta}, \mathbf{t}, \mathscr{C}_n) \le f(\max_i \mu_i \cdot \mathbf{1}, \boldsymbol{\eta}, \mathbf{t}, \mathscr{C}_n)$.*

In what follows, we examine the effect of changes in demand variance on long chain efficiency. For balanced but non-identical systems, this analysis proves to be unwieldy, especially for the case of normal distribution. To obtain some insights as well as to isolate the effect of demand variance, we consider the symmetric three-point distribution with homogeneous support but heterogeneous tail probabilities, i.e. $\mu_i = \mu, \eta_i = \eta = \mu$, for all $i = 1, 2, \ldots, n$. Without loss of generality, we can further assume $\mu = 1$. Hence, we have

$$D_i = \begin{cases} 0, & \text{with probability } t_i \\ 1, & \text{with probability } 1 - 2t_i \\ 2, & \text{with probability } t_i \end{cases}$$

As such, we can simplify the notation for long chain efficiency into $f(\mathbf{t}, \mathscr{C}_n)$. Our interest is to see how $f(\mathbf{t}, \mathscr{C}_n)$ behaves as t_k increases or decreases. To this end, we must first characterize the performances $[\mathscr{D}_n]$, $[\mathscr{C}_n]$, and $[\mathscr{F}_n]$ of the dedicated, the long chain, and the full flexibility designs, respectively.

From here onwards, whenever the integer k appears in a statement, we are actually referring to some $i \in \{1, 2, \ldots, n\}$ congruent to k modulo n. For instance, plant $n + 1$ refers to plant 1, while product 0 refers to product n. That said, we first present the following preliminary results that will help prove our subsequent main results.

Lemma 1. *For any balanced but non-identical system of size n such that $E[D_i] = c_i$ for all $i \in A$, we have*

$$[\mathscr{C}_n] = \sum_{i=1}^{n} ([\mathscr{C}_n \setminus \{(i, i+1)\}] - [\mathscr{C}_n \setminus \{(i-1, i), (i, i), (i, i+1)\}]).$$

Proof. This is straight-forward from Theorem 3 in Simchi-Levi and Wei (2012). □

Lemma 2. *For any balanced but non-identical system of size n such that $E[D_i] = c_i$ for all $i \in A$, and for any $i = 1, 2, \ldots, n$, we have*

$$[\mathscr{C}_n \setminus \{(i, i+1)\}] - [\mathscr{C}_n \setminus \{(i-1, i), (i, i), (i, i+1)\}]$$
$$= \mathbf{q}^0 \mathbf{A}_{i+1} \mathbf{A}_{i+2} \cdots \mathbf{A}_{i-1} \boldsymbol{\pi}^i$$

where $\mathbf{q}^0 = [1, 0]$, $\mathbf{A}_k = \begin{bmatrix} 1-t_k & t_k \\ t_k & 1-t_k \end{bmatrix}$, and $\boldsymbol{\pi}^i = [1 - t_i, 1]^T$.

Proof. The proof is similar to those of Lemma 3 and Theorem 5 in Simchi-Levi and Wei (2012). Note, however, that Simchi-Levi and Wei (2012) consider general discrete demand distributions but identical systems. □

Lemma 3. *For $j \ge i$, $\Pi_{k=i}^{j} \mathbf{A}_k = \begin{bmatrix} \frac{1 + \Pi_{k=i}^{j}(1-2t_k)}{2} & \frac{1 - \Pi_{k=i}^{j}(1-2t_k)}{2} \\ \frac{1 - \Pi_{k=i}^{j}(1-2t_k)}{2} & \frac{1 + \Pi_{k=i}^{j}(1-2t_k)}{2} \end{bmatrix}$.*

Proof. We prove this result by induction. First, observe that the lemma is true for $j = i$ by definition of $\mathbf{A}_i$. Suppose the lemma is true for $j = j_0$. We want to show that it is also true for $j = j_0 + 1$.

$$\Pi_{k=i}^{j_0+1} \mathbf{A}_k = \Pi_{k=i}^{j_0} \mathbf{A}_k \cdot \mathbf{A}_{j_0+1}$$

$$= \begin{bmatrix} \frac{1 + \Pi_{k=i}^{j_0}(1-2t_k)}{2} & \frac{1 - \Pi_{k=i}^{j_0}(1-2t_k)}{2} \\ \frac{1 - \Pi_{k=i}^{j_0}(1-2t_k)}{2} & \frac{1 + \Pi_{k=i}^{j_0}(1-2t_k)}{2} \end{bmatrix} \cdot \begin{bmatrix} 1 - t_{j_0+1} & t_{j_0+1} \\ t_{j_0+1} & 1 - t_{j_0+1} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1 + \Pi_{k=i}^{j_0+1}(1-2t_k)}{2} & \frac{1 - \Pi_{k=i}^{j_0+1}(1-2t_k)}{2} \\ \frac{1 - \Pi_{k=i}^{j_0+1}(1-2t_k)}{2} & \frac{1 + \Pi_{k=i}^{j_0+1}(1-2t_k)}{2} \end{bmatrix}$$

Lemma 4. *For any balanced but non-identical system of size n such that $E[D_i] = c_i$ for all $i \in A$, we have*

$$[\mathscr{F}_n] = [\mathscr{F}_{n-1}] + 1 - t_n \cdot P_{n-1}(n-1).$$

where $P_k(w) = Pr\{\sum_{i=1}^{k} D_i = w\}$ is the probability that the total demand for the first k products is w.

Proof. By conditioning on D_n, we get

$$[\mathscr{F}_n] = [\mathscr{F}_{n-1}] + Pr\{D_n = 0\} \cdot Pr\left\{\sum_{i=1}^{n-1} D_i > n-1\right\} \cdot 1$$
$$+ Pr\{D_n = 1\} \cdot 1 + Pr\{D_n = 2\}$$
$$\cdot \left[Pr\left\{\sum_{i=1}^{n-1} D_i \ge n-1\right\} \cdot 1 + Pr\left\{\sum_{i=1}^{n-1} D_i < n-1\right\} \cdot 2\right]$$
$$= [\mathscr{F}_{n-1}] + t_n \cdot \frac{1 - P_{n-1}(n-1)}{2} + (1 - 2t_n)$$
$$+ t_n \cdot \left[\frac{1 + P_{n-1}(n-1)}{2} + \frac{1 - P_{n-1}(n-1)}{2} \cdot 2\right]$$
$$= [\mathscr{F}_{n-1}] + 1 - t_n \cdot P_{n-1}(n-1)$$

The second equation is due to the symmetry of the distribution of $\sum_{i=1}^{n-1} D_i$. □

It is now time to present our next main result, which characterizes the performances of the dedicated, the long chain, and the full flexibility designs.

Theorem 3. *Consider a balanced but non-identical system of size n such that $E[D_i] = c_i$ for all $i \in A$. For the symmetric three-point distribution with $\mu_i = 1, \eta_i = 1$, for all $i = 1, 2, \ldots, n$, we have*

(a) $[\mathscr{D}_n] = n - \sum_{i=1}^{n} t_i$
(b) $[\mathscr{C}_n] = n - \frac{1}{2} \sum_{i=1}^{n} t_i - \frac{1}{2} K(n) \sum_{i=1}^{n} \frac{t_i}{1 - 2t_i}$, where $K(n) = \Pi_{i=1}^{n}(1 - 2t_i)$ *is the probability that all products have demand equal to the mean.*
(c) $[\mathscr{F}_n] = n - \sum_{i=1}^{n} t_i \cdot P_{i-1}(i-1)$.

Proof. For (a), the result follows easily from the fact that $[\mathscr{D}_n] = \sum_{i=1}^{n} E[\min(D_i, 1)]$.

For (b), we combine Lemmas 1 and 2 to obtain $[\mathscr{C}_n] = \sum_{i=1}^{n} \mathbf{q}^0 \Pi_{k=1, k \ne i}^{n} \mathbf{A}_k \boldsymbol{\pi}^k$. By applying Lemma 3 and performing standard matrix multiplication, we get

$$[\mathscr{C}_n] = n - \frac{1}{2} \sum_{i=1}^{n} t_i - \frac{1}{2} \sum_{i=1}^{n} t_i \cdot \Pi_{k=1, k \ne i}^{n}(1 - 2t_k).$$

By definition of $K(n)$, the result follows.

For (c), the result follows by applying Lemma 4 recursively and because $[\mathscr{F}_1] = 1$. □

Corollary 2. *We have the following partial derivatives and difference equations.*

(a) $\frac{\partial}{\partial t_n}[\mathscr{D}_n] = -1$
(b) $\frac{\partial}{\partial t_n}[\mathscr{C}_n] = -\frac{1}{2} - \frac{1}{2} K(n-1) + K(n-1) \sum_{i=1}^{n-1} \frac{t_i}{1 - 2t_i}$
(c) $\frac{\partial}{\partial t_n}[\mathscr{F}_n] = -P_{n-1}(n-1)$

(d) $[\mathscr{D}_n] - [\mathscr{D}_{n-1}] = 1 + t_n \cdot \frac{\partial}{\partial t_n}[\mathscr{D}_n] = 1 - t_n$

(e) $[\mathscr{C}_n] - [\mathscr{C}_{n-1}] = 1 + t_n \cdot \frac{\partial}{\partial t_n}[\mathscr{C}_n] = 1 + t_n(-\frac{1}{2} - \frac{1}{2}K(n-1) + K(n-1)\sum_{i=1}^{n-1}\frac{t_i}{1-2t_i})$

(f) $[\mathscr{F}_n] - [\mathscr{F}_{n-1}] = 1 + t_n \cdot \frac{\partial}{\partial t_n}[\mathscr{F}_n] = 1 - t_n \cdot P_{n-1}(n-1)$

Next, we present an interesting connection between the sensitivity of long chain efficiency to demand variance (represented by tail probability) and the sensitivity of long chain efficiency to system size.

Theorem 4. $f(\mathbf{t}, \mathscr{C}_n)$ *is decreasing in* t_n *if and only if* $f(\mathbf{t}, \mathscr{C}_n)$ *is decreasing in n.*

Proof. By definition of $f(\mathbf{t}, \mathscr{C}_n)$ and using Corollary 2 (d)–(f), we have

$$\frac{\partial}{\partial t_n} f(\mathbf{t}, \mathscr{C}_n) \le 0 \Leftrightarrow \frac{\partial}{\partial t_n}\frac{[\mathscr{C}_n] - [\mathscr{D}_n]}{[\mathscr{F}_n] - [\mathscr{D}_n]} \le 0$$

$$\Leftrightarrow \frac{([\mathscr{F}_n] - [\mathscr{D}_n]) \cdot \frac{\partial}{\partial t_n}([\mathscr{C}_n] - [\mathscr{D}_n]) - ([\mathscr{C}_n] - [\mathscr{D}_n]) \cdot \frac{\partial}{\partial t_n}([\mathscr{F}_n] - [\mathscr{D}_n])}{([\mathscr{F}_n] - [\mathscr{D}_n])^2}$$

$$\le 0$$

$$\Leftrightarrow ([\mathscr{F}_n] - [\mathscr{D}_n]) \cdot \frac{([\mathscr{C}_n] - [\mathscr{D}_n]) - ([\mathscr{C}_{n-1}] - [\mathscr{D}_{n-1}])}{t_n}$$

$$- ([\mathscr{C}_n] - [\mathscr{D}_n]) \cdot \frac{([\mathscr{F}_n] - [\mathscr{D}_n]) - ([\mathscr{F}_{n-1}] - [\mathscr{D}_{n-1}])}{t_n} \le 0$$

$$\Leftrightarrow -([\mathscr{F}_n] - [\mathscr{D}_n])([\mathscr{C}_{n-1}] - [\mathscr{D}_{n-1}])$$

$$+ ([\mathscr{C}_n] - [\mathscr{D}_n])([\mathscr{F}_{n-1}] - [\mathscr{D}_{n-1}]) \le 0$$

$$\Leftrightarrow \frac{[\mathscr{C}_n] - [\mathscr{D}_n]}{[\mathscr{F}_n] - [\mathscr{D}_n]} \le \frac{[\mathscr{C}_{n-1}] - [\mathscr{D}_{n-1}]}{[\mathscr{F}_{n-1}] - [\mathscr{D}_{n-1}]}$$

The analysis involved to show that long chain efficiency decreases in system size has proven unwieldy. In fact, previous works (Chou et al., 2010b; Simchi-Levi & Wei, 2012) are unable to prove this relationship even for the simplest case of balanced and identical systems. Nonetheless, several numerical studies have supported the belief that this relationship holds. To conduct similar numerical studies for our balanced but non-identical case, we first note the following recursive equations for $P_k(w)$. For $k = 1, \ldots, n$,

$$P_k(1) = (1 - 2t_k)P_{k-1}(0) + t_k P_{k-1}(1)$$

$$P_k(w) = (1 - 2t_k)P_{k-1}(w - 1) + t_k P_{k-1}(w)$$

$$\qquad + t_k P_{k-1}(w - 2), \quad \forall w = 2, \ldots, k-1$$

$$P_k(k) = (1 - 2t_k)P_{k-1}(k - 1) + 2t_k P_{k-1}(k - 2)$$

with boundary conditions $P_0(0) = 1$, and $P_k(0) = \Pi_{i=1}^{k} t_i$. For completeness, we can obtain the rest of the probabilities with $P_k(w) = P_k(2k - w), \forall w = k+1, k+2, \ldots, 2k$.

The above recursive equations allow us to calculate $P_k(k)$ for all $k = 1, 2, \ldots, n$, which in turn allows us to calculate $[\mathscr{F}_n]$ according to Theorem 3. From the same theorem, the values of $[\mathscr{D}_n]$ and $[\mathscr{C}_n]$ are more straight-forward to calculate. For different system sizes $n = 2, 3, \ldots, 20$, we randomly generated 10,000 sets of tail probabilities $t_i \in (0, 0.5), \forall i = 1, 2, \ldots, n$ and calculated the long chain efficiency $f(\mathbf{t}, \mathscr{C}_n)$. For all instances, we find that $f(\mathbf{t}, \mathscr{C}_n)$ is, indeed, decreasing in n. Hence, we also have good reason to believe that $f(\mathbf{t}, \mathscr{C}_n)$ is decreasing in t_n (equivalently, demand variance).

The next result follows easily from Theorem 4, and also provides upper and lower bounds on long chain efficiency.

Corollary 3. *For symmetric three-point distribution with* $\mu_i = 1, \eta_i = 1$ *for all* $i = 1, 2, \ldots, n$, *suppose* $f(\mathbf{t}, \mathscr{C}_n)$ *is decreasing in n. Then,* $f(\max_i t_i \cdot \mathbf{1}, \mathscr{C}_n) \le f(\mathbf{t}, \mathscr{C}_n) \le f(\min_i t_i \cdot \mathbf{1}, \mathscr{C}_n)$.

For our next result, we refer to the edges in $\mathscr{G}$ as flexibility links. We then examine the efficiency of link usage for products

with low demand variance. Interestingly, we find that link efficiency decreases and goes to zero as variance decreases and goes to zero. This implies that when a product has very low demand variance, including it in the long chain does not cause any harm to system performance, but the benefit is minimal, resulting in a poorly utilized flexibility link.

Theorem 5. *Consider a balanced system of size n such that* $E[D_i] = c_i$ *for all* $i \in A$ *and two flexibility designs; namely,* $\mathscr{C}_n$ *and* $\mathscr{C}_{n-1} \cup \{(n, n)\}$. *The difference in performance* $[\mathscr{C}_n] - [\mathscr{C}_{n-1} \cup \{(n, n)\}]$ *is increasing in* t_n *and equals 0 when* $t_n = 0$.

Proof. It is easy to see that $[\mathscr{C}_{n-1} \cup \{(n, n)\}] = [\mathscr{C}_{n-1}] + (1 - t_n)$. From Corollary 2(e), we have

$$[\mathscr{C}_n] - [\mathscr{C}_{n-1}] - (1 - t_n)$$

$$= t_n\left(\frac{1}{2} - \frac{1}{2}K(n-1) + K(n-1)\sum_{i=1}^{n-1}\frac{t_i}{1-2t_i}\right).$$

Except in the trivial case when $t_i = 0$ for all $i = 1, 2, \ldots, n-1$, the coefficient of t_n above is strictly positive and independent of t_n. Hence, the result follows. $\square$

3.1. Discussion of results

While the above results are interesting and useful in themselves, we further discuss how they can be used to develop a method to create flexibility designs that can outperform the long chain. First, Theorem 2 tells us that a product with low demand mean can create a bottleneck if included in a long chain design. Therefore, we will likely be better off with having a few smaller chains where each chain will have relatively similar demand means. However, what we gain from bottleneck removal, we might lose due to a reduction in risk pooling. In fact, there exists no risk pooling at all between products from different small chains.

To address this issue, we turn to Theorems 1 and 5. Theorem 1 says that there is little value in installing flexibility for products with very low demand variance. Similarly, Theorem 5 tells us that the link used to include a product with low demand variance in a chain (either long or short) brings minimal value. This finding suggests us to isolate products with low demand variance in order to free up previously poorly utilized links which can then be used to connect the small chains created by the aforementioned bottleneck removal. By doing so, we can address the problem of reduced risk pooling.

That said, the next question becomes how to form the smaller chains. That is, how many small chains should there be and which products should belong to which chain? To answer this, we attempt to combine Corollaries 1 and 3. We do so numerically and find that for both normal distribution and symmetric three-point distribution, our numerical examples consistently show that

$$f(\min_i \mu_i \cdot \mathbf{1}, \max_i \sigma_i \cdot \mathbf{1}, \mathscr{C}_n) \le f(\min_i \mu_i \cdot \mathbf{1}, \boldsymbol{\sigma}, \mathscr{C}_n) \le f(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathscr{C}_n)$$

$$f(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathscr{C}_n) \le f(\max_i \mu_i \cdot \mathbf{1}, \boldsymbol{\sigma}, \mathscr{C}_n) \le f(\max_i \mu_i \cdot \mathbf{1}, \min_i \sigma_i \cdot \mathbf{1}, \mathscr{C}_n).$$

For normal distribution, we generate 100 random scenarios. For each scenario, we randomly sample $\mu_i \in [50, 100]$, and $\sigma_i \in [0, 25], \forall i = 1, \ldots, n$. Fig. 2 shows that the above bounds persist in this set of numerical tests. For symmetric three-point distribution, we likewise generate 100 random scenarios. For each scenario, we randomly sample $\mu_i \in [50, 100]$, $\eta_i \in [0, 50]$, and $t_i \in (0, 0.5)$, $\forall i = 1, \ldots, n$. Fig. 3 also confirms the above bounds on long chain efficiency, in this case for the three-point distribution.

The first inequality above says that the long chain efficiency of any balanced but non-identical system is bounded below by the long chain efficiency of an associated balanced and identical system. In this system, the products all have demand mean equal to
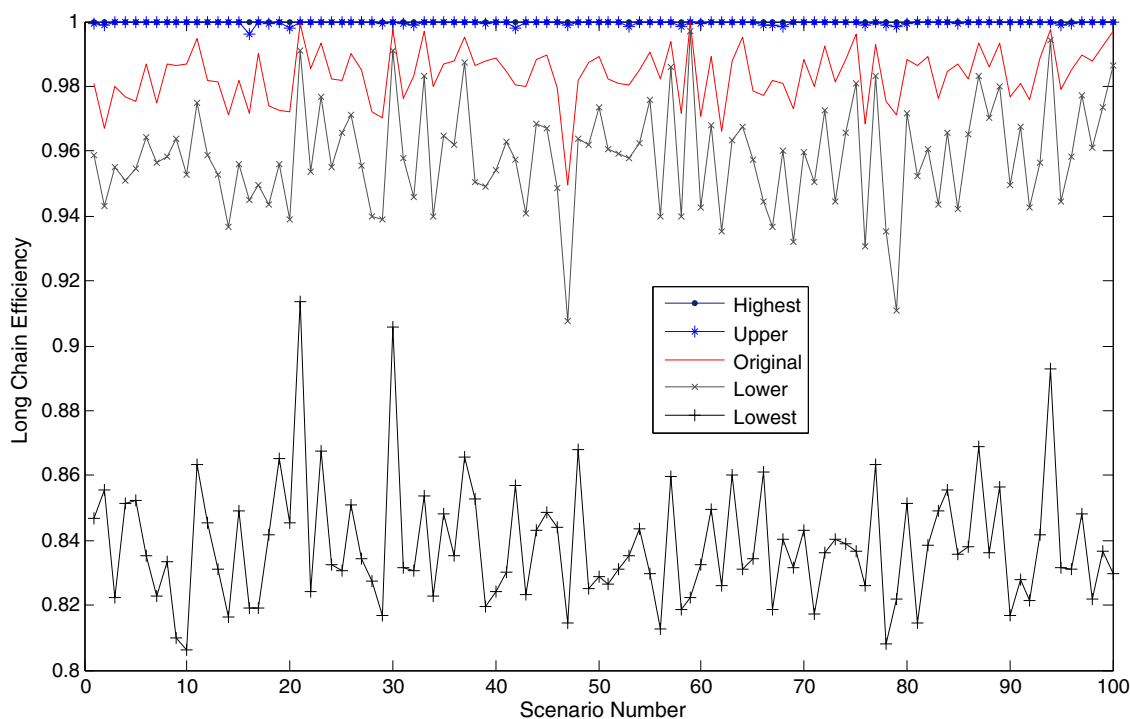
Fig. 2. Long chain efficiency bounds for normal distribution.
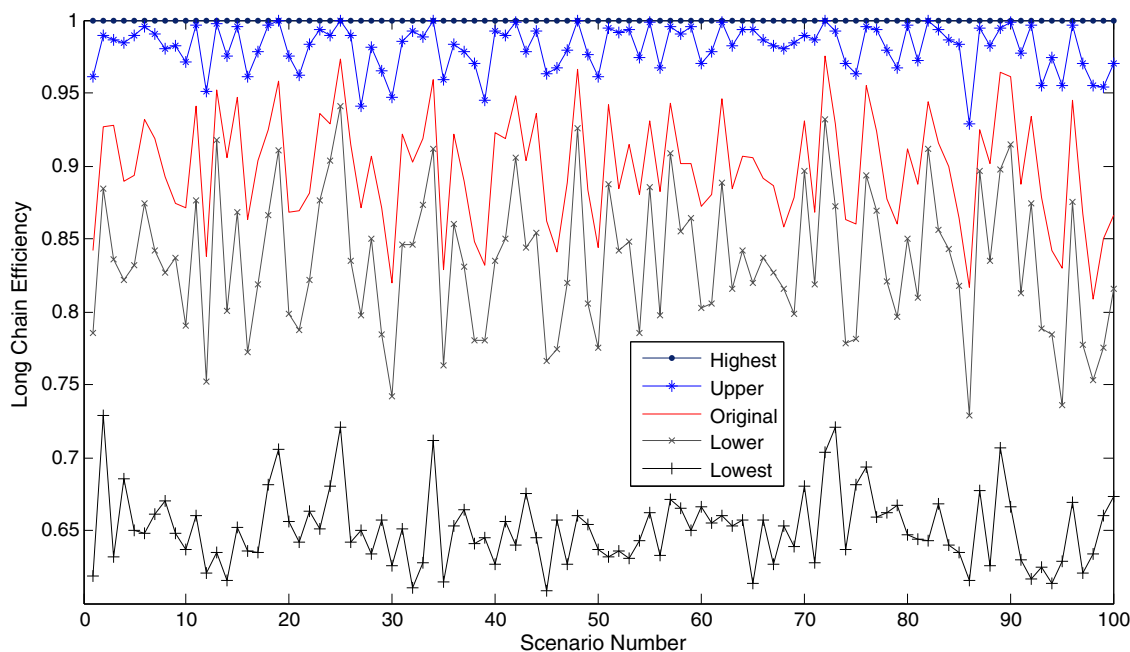
Fig. 3. Long chain efficiency bounds for symmetric three-point distribution.

the smallest demand mean $\min_i \mu_i$ in the non-identical system, and demand variance equal to the largest demand variance $\max_i \sigma_i$ in the non-identical system. This brings us back to the question of how to form smaller chains. The above seems to suggest a group index for each small chain to indicate the amount of variability within the chain, and hence whether products in that chain form a good grouping. We define this group index as follows.

Definition 1. For a balanced but non-identical system or subsystem G, we define its **group coefficient of variation** as follows.

$$GCV(G) = \frac{\max_{i \in G} \sigma_i}{\min_{i \in G} \mu_i}$$

There are two reasons to use $GCV(G) = \frac{\max_{i \in G} \sigma_i}{\min_{i \in G} \mu_i}$ rather than $\frac{\sum_{i \in G} \sigma_i}{\sum_{i \in G} \mu_i}$ (or some other index). First, when GCV is small, the balanced and identical system associated with the lowest bound $f(\min_i \mu_i \cdot \mathbf{1}, \max_i \sigma_i \cdot \mathbf{1}, \mathscr{C}_n)$ would already have high long chain efficiency. Hence, it would also be good to use the long chain design for the non-identical system G. Secondly, when GCV is high, the index clearly guides us in identifying which product to exclude

from the system G. Specifically, the denominator suggests to remove the product with the lowest demand mean in order to reduce the GCV, which is consistent with our intention to remove the bottleneck. On the other hand, the numerator implies concern about the product with the largest demand variance. We shall discuss later how our proposed method will provide access to more capacity for these high-variance products.

In summary, the GCV index will become a key component of our proposed method to generate good flexibility designs for balanced but non-identical systems, particularly those with heterogeneous demand variances. In the next section, we provide more details on how this method works.

4. The hub-and-chain method

For a balanced but non-identical system, we seek a flexibility design $\mathscr{G} \subseteq A \times B$ that maximizes $[\mathscr{G}]$ such that $|\mathscr{G}| \leq b = 2n$. As mentioned earlier, this is a difficult stochastic mixed integer linear program, which leads us to focus on heuristic methods to find good enough flexibility designs within reasonable computational time. Clearly, choosing $\mathscr{G} = \mathscr{C}_n$ involves negligible computational time and is a feasible design. However, as explained in the previous section, a product with low demand mean might cause a bottleneck and substantially reduce the effectiveness of this long chain design. As such, we might be better served using smaller chains with a provision to link together these small chains. And therein lies the inspiration for our proposed method.

In what follows, we present our variance-based hub-and-chain method, or VHC for short. The main idea is to partition the product set A into $s + 1$ subsets $G_0, G_1, \ldots, G_s$ such that the subsystem corresponding to G_0 will employ a dedicated design, while the subsystems corresponding to $G_1, \ldots, G_s$ will each employ a long chain design. Then, we choose a product–plant pair in each subsystem $G_1, \ldots, G_s$ and call it the *satellite pair* for that subsystem. Finally, we use the links freed up by G_0 to connect the satellite pair of G_1 to those of $G_2, \ldots, G_s$. We denote the resulting flexibility design as $\mathscr{H}(G_0, G_1, \ldots, G_s)$.

Our problem becomes to maximize $[\mathscr{H}(G_0, G_1, \ldots, G_s)]$ such that $\bigcup_{i=0}^{s} G_i = A$ and $G_i \bigcap G_j = \emptyset, \forall i, j = 0, 1, \ldots, s, i \neq j$. This is still a difficult problem to solve for at least two reasons. First, the objective function is the expected optimal value of another optimization problem. As a remedy, we can consider a substitute objective function which is to minimize the maximum of the GCVs of the subsets $G_1, \ldots, G_s$. That is, we want to partition such that the worst group variability of all the subsystems is minimized.

$$\min \quad z$$

$$\text{s.t.} \quad \sum_{j=0}^{s} y_{ij} = 1, \quad \forall i = 1, \ldots, n$$

$$z \geq \text{GCV}(G_j) = \frac{\max_{i=1}^{n} \sigma_i y_{ij}}{\min_{i=1}^{n} \mu_i y_{ij}}, \quad \forall j = 1, \ldots, s$$

$$z \geq \frac{\sigma_i}{\mu_i} \cdot y_{i0} \quad \forall i = 1, \ldots, n$$

$$s = 1 + \frac{1}{2} \sum_{i=1}^{n} y_{i0}$$

$$y_{ij} \in \{0, 1\}, \quad \forall i = 1, \ldots, n, \forall j = 0, 1, \ldots, s$$

$$z \geq 0; s \geq 1 \text{ and is an integer}$$

where y_{ij} decides whether product i is included in subset G_j or not. In addition, the third and fourth constraints ensure that we do not include too many products into G_0. The third constraint prefers that products with low demand variability be isolated, while the fourth constraint ensures that the total number of flexibility links used stays at $2n$.

The second reason why our problem is difficult is that the integer s is also a decision variable, which depends on the number of products selected into the set G_0. In fact, even if we fix s, we are still left with a difficult binary nonlinear program. To avoid the long computational time associated with the above optimization problem, we propose the VHC method detailed in Algorithm 1. Re-

---

Algorithm 1 VHC method for balanced case.

Step 1

1: $G_0 \leftarrow \emptyset; G_1 \leftarrow A; n_1 \leftarrow n; x \leftarrow \frac{\min_{i \in G_1} \sigma_i}{\sum_{i=1}^{n} \sigma_i}$;

2: **while** $\frac{\min_{i \in G_1} \sigma_i}{\sum_{i=1}^{n} \sigma_i} < \theta_1$ and $x < \theta_2$ **do**

3: $\quad i^* \leftarrow \text{argmin}_{i \in G_1} \sigma_i; G_0 \leftarrow G_0 \cup \{i^*\}; G_1 \leftarrow G_1 \setminus \{i^*\}$;

4: $\quad x \leftarrow x + \frac{\min_{i \in G_1} \sigma_i}{\sum_{i=1}^{n} \sigma_i}; n_1 \leftarrow n_1 - 1$;

5: **end while**

Step 2

1: $j \leftarrow 1$;

2: **while** $G_j \neq \emptyset$ **do**

3: $\quad G_{j+1} \leftarrow \emptyset; n_{j+1} \leftarrow 0$;

4: $\quad$ **while** $GCV(G_j) > \theta_3$ **do**

5: $\quad\quad i^* \leftarrow \text{argmin}_{i \in G_j} \mu_i; \quad G_{j+1} \leftarrow G_{j+1} \cup \{i^*\}; \quad G_j \leftarrow G_j \setminus \{i^*\};$
$\quad\quad n_{j+1} \leftarrow n_{j+1} + 1$;

6: $\quad$ **end while**

7: $\quad j \leftarrow j + 1$;

8: **end while**

9: $s \leftarrow j - 1$.

Step 3

1: $l_1^* \leftarrow \text{argmax}_{i \in G_1} \sigma_i; \mathscr{G} \leftarrow \mathscr{D}_n$;

2: **for** $j = 1$ to s **do**

3: $\quad$ **for** $i = 1$ to $n_j - 1$ **do**

4: $\quad\quad \mathscr{G} \leftarrow \mathscr{G} \cup \{(G_j(i), G_j(i+1))\}$;

5: $\quad$ **end for**

6: $\quad \mathscr{G} \leftarrow \mathscr{G} \cup \{(G_j(n_j), G_j(1))\}$;

7: $\quad$ **if** $j \neq 1$ **then**

8: $\quad\quad l^* \leftarrow \text{argmax}_{i \in G_j} \sigma_i; \mathscr{G} \leftarrow \mathscr{G} \cup \{(l_1^*, l^*), (l^*, l_1^*)\}$;

9: $\quad$ **end if**

10: **end for**

---

call that A is the product set, $GCV(G_j)$ is according to Definition 1, and let $G(i)$ be the ith smallest element in set G.

The VHC method involves three main steps. In Step 1, products with very low demand variance will be isolated into the subset G_0, in which only a dedicated flexibility design will be employed. The decision on how many such products will be chosen depends on two pre-specified parameters $\theta_1, \theta_2 \in (0, 1)$. θ_1 refers to the maximum standard deviation for a single product relative to total standard deviation for all products in order for that single product to be relegated to G_0. On the other hand, θ_2 refers to the maximum proportion of total standard deviation to be allowed for a group of products to be relegated to G_0. Clearly, higher values of θ_1 and θ_2 will result in more products being included in subset G_0. All the remaining products will be placed in subset G_1. In Step 2, the subset G_1 will be further divided into more subsets $G_1, G_2, \ldots, G_s$ such that the GCV of each subset does not exceed a pre-specified threshold value $\theta_3 \in (0, 1)$. Our choice of θ_3 will obviously influence the number s of subsets as well as the number of elements in each subset. Finally, in Step 3, we form the hub-and-chain design. We first let the subsystem G_0 employ a dedicated design and each of the subsystems $G_1, \ldots, G_s$ employ a long chain design. For each non-dedicated subsystem, we let the product–plant pair with the largest demand variance be the satellite pair for that subsystem. Then, we label the subsystem G_1 as the hub whose satellite pair will be connected (via two links) to the satellite pair of

Table 1
An example of a system scenario.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
μ	489	231	201	148	408	121	251	116	439	419	336	105	246	434	279	407	173	106	252	405
σ	59	105	55	13	147	59	45	20	133	137	114	51	120	200	136	163	40	19	37	97

every other subsystem $G_2, \ldots, G_s$. Note that the reason for choosing pairs with maximum demand variance as satellite pairs is because they are the most likely sources of excess demand and idle capacity. Furthermore, this is in line with the discussion at the end of Section 3.1 that high-variance products should be provided more access to additional capacity.

Having said that, we now demonstrate how to implement the VHC method and benchmark against two existing methods; namely, the long chain design and the constraint sampling method developed in Chou et al. (2010b). Before we proceed, we observe that the number of flexibility links utilized by the hub-and-chain design is not always equal to $2n$, but largely depends on our choice of values for the parameters θ_1, θ_2, and θ_3. However, in most instances, the number of links deviates from $2n$ by only a small number. On one hand, a firm with no hard budget on flexibility links may want to just run the VHC method using some pre-tested parameters values. On the other hand, a firm with a strict budget of $2n$ links may still use the VHC method as a subroutine in a broader iterative method. We discuss both cases in the next subsections.

4.1. Implementation: No flexibility budget

In this section, we consider a 20×20 balanced but non-identical system. We build a simulator that randomly generates a system scenario as follows. A scenario consists of 20 normally distributed product demands with $\mu_i \in \{100, 101, \ldots, 500\}$ and $\sigma_i \in \{0, 1, 2, \ldots, \lfloor \frac{\mu_i}{2} \rfloor \}$ for $i = 1, 2, \ldots, 20$. Because the system is balanced, the capacity for plant i is $c_i = \mu_i$ for all $i = 1, 2, \ldots, 20$. Table 1 shows an example of a system scenario randomly generated by our simulator. Observe that product 4 and product 18 have low demand standard deviations of 13 and 19, respectively.

We now apply the VHC method to the example in Table 1 by setting the parameter values at $\theta_1 = 0.01, \theta_2 = 0.1$, and $\theta_3 = 0.6$. These values were selected after a few rounds of calibration. That said, the output of the VHC method is illustrated in Figs. 4 and 5. Fig. 4(a) presents the pre-established dedicated design, while Fig. 4(b) shows how the product–plant pairs are partitioned into the dedicated group G_0, and the non-dedicated groups G_1, G_2, and G_3. Not surprisingly, the low-variance products, including product 4 and product 18, are isolated into G_0, while G_1 contains the products with the largest demand means. Finally, Fig. 5 shows how the hub-and-chain design is formed. The satellite pairs for G_1, G_2, and G_3 are product-plant pairs 14, 15, and 6, respectively, which precisely correspond to the products with the largest demand variance in each group. G_1 then serves as the hub, to whose satellite pair the satellite pairs of all other small chains G_2 and G_3 are connected.

We use our simulator to randomly generate 30 system scenarios like the example above. For each scenario, we implement the VHC method (again using pre-calibrated parameter values $\theta_1 = 0.01, \theta_2 = 0.1$, and $\theta_3 = 0.6$) and obtain a hub-and-chain design. For each of the 30 scenarios, we benchmark the hub-and-chain design against two existing methods; namely, the long chain design and the constraint sampling method developed in Chou et al. (2010b). For the long chain, we follow the natural sequence of the products, i.e. product 1 is connected to plant 2, product 2 is connected to plant 3, and so on until product n is connected to plant

1. While the sequence of the long chain certainly impacts its performance, numerical studies have shown the performance does not vary substantially. For the constraint sampling method, Chou et al. (2010b) applied the method to a transshipment problem. To accommodate our problem, we redefine the following components accordingly.

$$\tilde{E}(x_{ij}^*) = \sum_{k=1}^{10000} \left[\frac{d_i^k c_j}{\max\{\sum_i d_i^k, \sum_j c_j\}} \right] / 10000,$$

$$\tilde{p}_{ij} = \frac{\tilde{E}(x_{ij}^*)}{\sum_{i=1}^{20} \sum_{j=1}^{20} \tilde{E}(x_{ij}^*)}$$

where the sampling probabilities $\tilde{p}_{ij}$ are used to generate 100 flexibility designs from which the best design is selected.

For benchmarking purpose, we randomly generate 10,000 demand realizations and compute the performance (i.e. expected sales) for each of these three methods. In addition, we also compute the performance of the dedicated as well as the full flexibility designs, using the same 10,000 demand realizations.

The results are summarized in Table 2, where columns 2–6 contain the respective performances of the dedicated, the long chain, the full flexibility, the constraint sampling, and the hub-and-chain designs. Columns 7–9 present the number of non-dedicated groups, the number of products in the dedicated group, and the number of flexibility links used in the hub-and-chain design, respectively. Finally, the last three columns show the efficiency of the hub-and-chain design, as well as the improvement of this design over both the long chain and the best design from constraint sampling. Here, we define improvement of the hub-and-chain design over a flexibility design $\mathscr{G}$ as

$$\frac{[\text{VHC}] - [\mathscr{D}_n]}{[\mathscr{G}] - [\mathscr{D}_n]} - 1 = \frac{[\text{VHC}] - [\mathscr{G}]}{[\mathscr{G}] - [\mathscr{D}_n]}.$$

In all scenarios, VHC achieves at least 91% of the benefits of full flexibility. Moreover, in all but one scenario, VHC obtains at least 94% of the benefits of full flexibility, and in more than 70% of the scenarios, at least 96%.

Furthermore, we see from the last two columns that on average, the hub-and-chain design improves on the long chain by 16%, and improves on the best design from constraint sampling by 38%. These are substantial improvements, which corroborate our earlier result that product demand variance plays a significant role in the performance of flexibility designs. Hence, it is not surprising that a method like the VHC, which takes into account demand variances, outperforms existing methods. In terms of computational time, the VHC method requires only at most 1% of the time required to run and compare 100 designs using constraint sampling.

It is important to note the following caveats. First, the long chain requires negligible amount of time to design and was also designed for a balanced and identical system. Hence, it is expected to perform worse than a variance-based method. Second, the constraint sampling was proposed based on some restriction on demand bounds. Therefore, heterogeneous demand variances with large values may adversely affect the performance of this method. However, this method works even for unbalanced and non-identical systems.

Another observation is that since we have no flexibility budget, one concern is that the VHC method might use too many flexibility
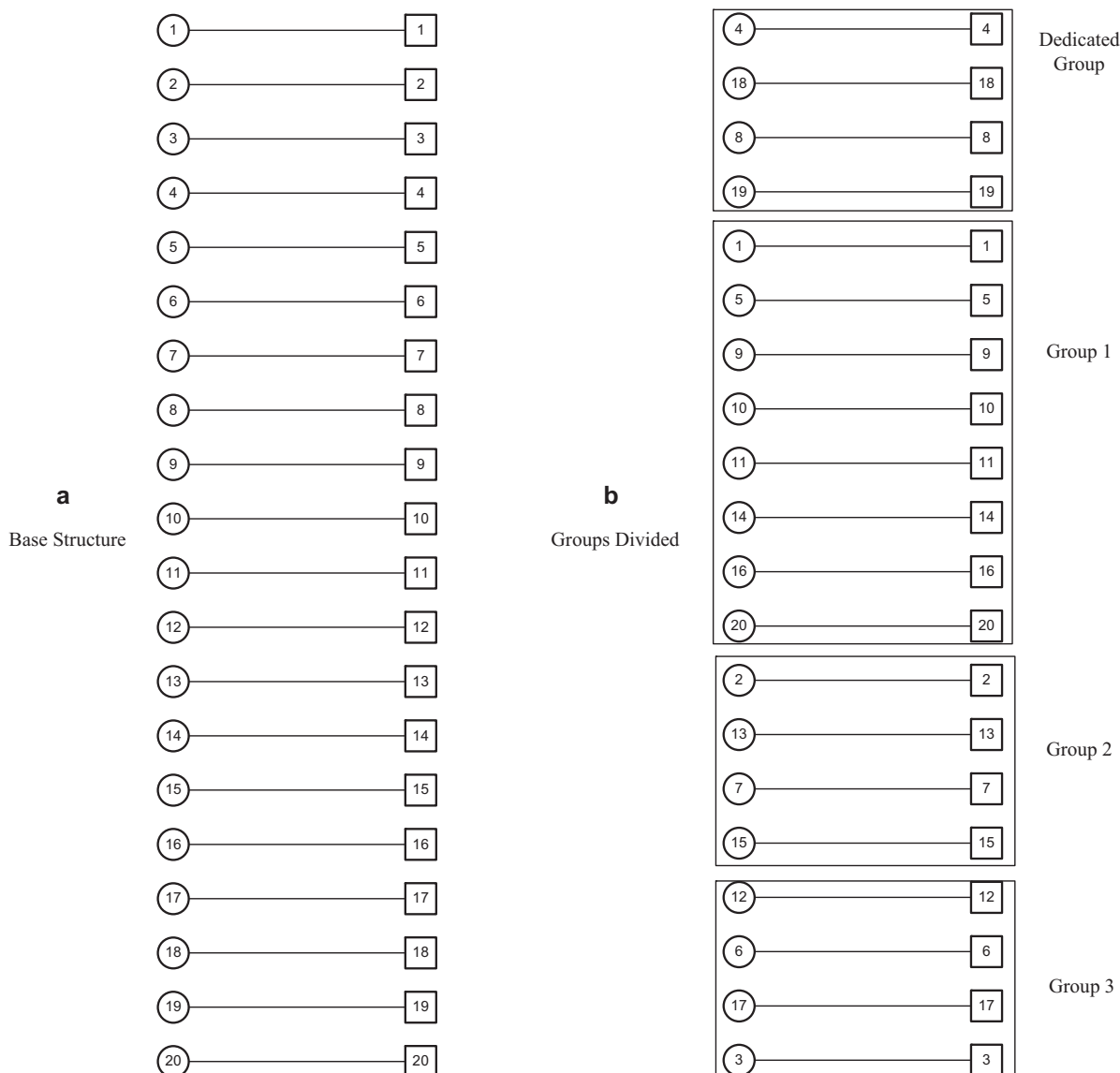
Fig. 4. How the product–plant pairs are grouped.

links compared to the 40 links used in the long chain. Fortunately, a quick look at Table 2 reveals that the VHC method only use, on average, about 1.6 more links. In fact, in some instances, the VHC even utilizes less links than the long chain, and still obtains substantial gains.

4.2. Implementation: With flexibility budget

In this section, we consider a hard budget b on the number of flexibility links that can be used. To obtain the optimal hub-and-chain design, we implement the following iterative method which uses Steps 2 and 3 of Algorithm 1 as a subroutine. That is, we no longer need the parameters θ_1 and θ_2 to determine the number $|G_0|$ of dedicated product–plant pairs. Instead, we initialize $|G_0| = 2$, and later incrementally add 2 more products to G_0, i.e. $|G_0| \leftarrow |G_0| + 2$. For each value of $|G_0|$, we iteratively run Steps 2 and 3 of Algorithm 1, starting with $\theta_3 = \max_{i=1}^{n} \frac{\sigma_i}{\mu_i}$. If the resulting number of non-dedicated subsets is $s > \frac{b+|G_0|}{2} - n + 1$, then we increment θ_3 by a step size $\Delta\theta_3$ until we satisfy $s \leq \frac{b+|G_0|}{2} - n + 1$. While a binary search might be faster, this line search for θ_3 does not depend on system size n, hence is already computationally

very efficient. Finally, we choose the best among hub-and-chain designs generated by different values of $|G_0|$ based on their performance according to 10,000 randomly selected demand realizations.

For our numerical implementation, we choose $n = 20, b = 40, \Delta\theta_3 = 0.01$ and let $|G_0| = 2, 4, \ldots, 12$. As in the no flexibility budget case, we use our simulator to randomly generate 30 system scenarios. For each scenario, we implement the above iterative method. The results are summarized in Table 3, where columns 2–5 contain the performances of the dedicated, the long chain, the full flexibility, and the constraint sampling designs. In columns 6–8, we report the performance of the hub-and-chain designs for $|G_0| = 2, 4,$ and 6, respectively. For each scenario, we use an underline to indicate the performance of the best hub-and-chain design. Finally, the last three columns show the efficiency of the hub-and-chain design, as well as the improvement of this design over both the long chain and the best design from constraint sampling.

It is interesting to note that the performance of the hub-and-chain design is unimodal in $|G_0|$. This implies that as more products are isolated into the dedicated group, the loss of flexibility in this group increases to an extent no longer matched by the benefit from linking together a larger number of small chains. Computationally, this means that once [VHC] decreases for some $|G_0|$, we

Table 2
Performance comparison: No flexibility budget.

Scenario	$[\mathscr{D}_n]$	$[\mathscr{C}_n]$	$[\mathscr{F}_n]$	[CS]	[VHC]	s	$\lvert G_0 \rvert$	$\lvert$VHC$\rvert$	$f(\boldsymbol{\mu}, \boldsymbol{\sigma}, \text{VHC})(\%)$	Improvement over	
										$\mathscr{C}_n(\%)$	CS (%)
1	5328	5713	5848	5731	5831	4	2	44	96.8	30.9	24.8
2	6227	6585	6643	6500	6623	3	3	39	95.3	10.8	45.4
3	4981	5273	5330	5248	5321	3	1	43	97.4	16.3	27.3
4	5750	6081	6127	5990	6111	3	2	41	95.6	9.2	50.1
5	5809	6153	6229	6151	6212	3	3	41	96.0	17.2	17.7
6	4884	5089	5112	5045	5105	3	3	41	96.6	7.6	37.1
7	4979	5295	5352	5251	5347	4	1	44	98.6	16.6	35.4
8	5498	5863	5943	5823	5928	3	2	42	96.7	17.8	32.3
9	6133	6546	6644	6511	6638	4	0	46	98.9	22.3	33.4
10	6422	6825	6878	6739	6870	4	2	42	98.2	11.1	41.4
11	6105	6470	6539	6403	6524	2	2	40	96.4	14.8	40.7
12	5882	6251	6303	6158	6280	3	4	40	94.7	8.0	44.5
13	5826	6207	6291	6150	6275	3	3	41	96.4	17.7	38.5
14	5551	5880	5992	5811	5967	3	6	37	94.3	26.4	59.8
15	5514	5886	5975	5868	5960	4	2	44	96.8	20.0	25.9
16	4596	4882	4959	4836	4941	3	4	40	95.1	20.8	43.7
17	6149	6490	6554	6426	6538	3	4	39	96.0	14.1	40.4
18	5495	5763	5796	5681	5785	3	3	40	96.6	8.3	56.2
19	5295	5598	5656	5538	5646	3	4	40	97.2	15.8	44.4
20	6004	6410	6471	6344	6457	3	2	42	97.1	11.8	33.2
21	5803	6218	6320	6191	6294	3	2	42	95.0	18.2	26.6
22	5101	5449	5524	5393	5505	4	2	43	95.5	16.1	38.4
23	5185	5502	5568	5464	5560	4	2	43	97.8	18.2	34.2
24	5136	5487	5530	5397	5516	3	1	43	96.5	8.3	45.8
25	6231	6580	6637	6522	6632	4	1	44	98.9	15.1	37.7
26	5866	6215	6279	6158	6276	3	0	44	99.3	17.6	40.3
27	4929	5289	5404	5306	5389	3	1	43	97.0	27.7	22.1
28	5138	5504	5596	5474	5582	3	2	42	96.9	21.3	32.0
29	4791	5017	5072	4976	5047	4	7	39	91.1	13.4	38.4
30	5481	5777	5831	5713	5820	3	4	39	96.9	14.8	46.6

Table 3
Performance comparison: With flexibility budget.

| Scenario	$[\mathscr{D}_n]$	$[\mathscr{C}_n]$	$[\mathscr{F}_n]$	[CS]	[VHC] when $\lvert G_0 \rvert =$			$f(\boldsymbol{\mu}, \boldsymbol{\sigma}, \text{VHC})(\%)$	Improvement over	
					2	4	6		$\mathscr{C}_n(\%)$	CS (%)
1	5342	5727	5862	5700	5828	5814	5790	93.6	26.4	35.9
2	6232	6596	6656	6504	6636	6627	6604	95.4	11.1	48.7
3	4987	5279	5336	5203	5321	5311	5291	95.7	14.2	54.1
4	5750	6071	6118	5985	6102	6058	6068	95.8	9.6	49.9
5	5803	6150	6233	6090	6211	6208	6187	94.8	17.5	42.2
6	4884	5088	5113	5030	5106	5101	5073	96.8	8.7	51.6
7	4986	5300	5354	5232	5337	5321	5296	95.1	11.4	42.2
8	5496	5866	5948	5805	5924	5911	5866	94.6	15.7	38.2
9	6123	6536	6622	6439	6585	6565	6537	92.7	12.0	46.4
10	6422	6818	6876	6709	6861	6846	6818	96.7	10.9	52.7
11	6110	6470	6538	6420	6522	6512	6493	96.3	14.6	32.9
12	5885	6253	6299	6164	6280	6282	6271	95.8	8.0	42.3
13	5839	6222	6303	6171	6284	6279	6263	96.0	16.3	34.2
14	5558	5897	6005	5877	5983	5993	5985	97.5	28.4	36.4
15	5516	5886	5979	5827	5948	5948	5927	93.3	16.9	39.1
16	4598	4886	4965	4845	4937	4949	4929	95.4	21.5	41.8
17	6140	6481	6546	6426	6537	6531	6513	97.9	16.5	39.1
18	5492	5762	5798	5692	5789	5776	5755	97.1	10.1	48.6
19	5301	5597	5651	5544	5641	5641	5609	97.2	14.8	39.9
20	6011	6428	6493	6347	6473	6457	6432	95.9	11.0	37.6
21	5787	6199	6298	6162	6259	6259	6238	92.4	14.4	25.7
22	5089	5443	5520	5368	5495	5485	5461	94.1	14.7	45.7
23	5204	5518	5578	5442	5566	5558	5534	96.6	15.3	51.6
24	5123	5477	5523	5361	5501	5484	5464	94.6	6.9	58.8
25	6224	6571	6625	6488	6607	6588	6562	95.4	10.4	45.1
26	5861	6209	6278	6116	6245	6230	6205	92.0	10.3	50.4
27	4938	5291	5403	5252	5355	5363	5339	91.3	20.4	35.3
28	5127	5490	5583	5440	5566	5550	5518	96.3	21.0	40.3
29	4799	5027	5084	4987	5073	5070	5061	96.1	20.4	45.5
30	5488	5776	5822	5691	5811	5812	5789	96.9	12.4	59.3
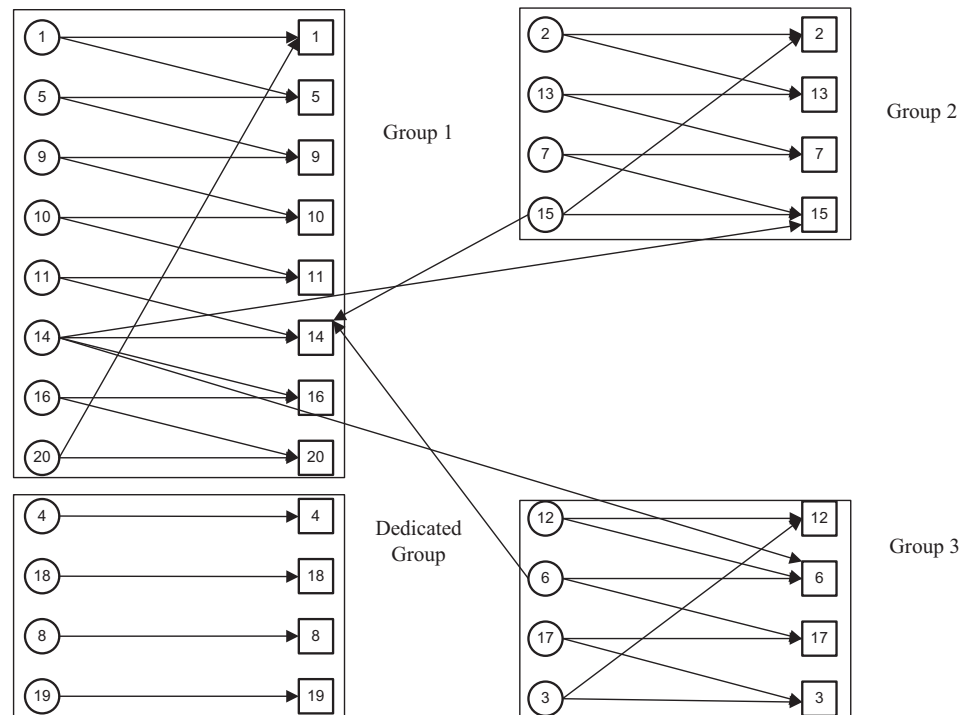
Fig. 5. The hub-and-chain design.

can terminate our iterative method. This is the reason why we only report the performances for $|G_0| = 2, 4, 6$. Moreover, we observe that the best hub-and-chain designs mostly coincide with $|G_0| = 2$ or $|G_0| = 4$. Hence, the computational requirements for our VHC method is reasonably low.

In terms of solution quality, we can see from Table 3 that VHC achieves at least 91% of the benefits of full flexibility in all scenarios sampled. VHC also outperforms the long chain by 15% and outperforms constraint sampling by 44%. In fact, Fig. 6 clearly shows that the efficiency of the hub-and-chain design is better than those of both constraint sampling and the long chain design. Furthermore, we see that the hub-and-chain design is already almost as good as full flexibility, and leaves little room for further improvement. That is, any additional flexibility link will be of little value and hence, inefficiently utilized. Finally, we also observe that the efficiency of the hub-and-chain design is less variable (more stable) than those of the other two methods.

5. Case study and managerial implications

In this section, we discuss the practical relevance of our paper by demonstrating how our VHC method can be used in a case study. Afterwards, we also discuss and summarize some managerial insights that can be gleaned from our work.

5.1. Edible oil case study

Company Y is one of the largest producers of consumer-pack edible oil products (e.g. soybean oil, sunflower oil) in China. These bottled oil products are consumed in Chinese households on a daily basis, primarily for cooking. The company's plants located in Guangzhou and Shenzhen have 16 filling-and-packaging lines which can manufacture various size formats and bottle shapes. These production lines can cope with 1.6-, 1.8-, 2-, 2.5-, and 3-L formats for the round bottle shape, 4-, 4.5-, 5-, and 6-L formats for both round and square bottle shapes, and 5.1999-, 5.258-, 5.435-L promotion formats for the round bottle shape. While each line

can technically be re-engineered to produce all formats and bottle shapes, doing so will require the plants to keep many types of machine parts to switch between lines. This is costly and will greatly complicate the scheduling problem. Currently, each line is built and designed for one format and bottle shape. The capacity of the lines (in terms of number of bottles) will not vary significantly even if switched to fill another size format due to the nature of this company's specific filling and packaging process.

Recently, the company's management started an initiative to integrate the operations of the two plants and to restructure the flexibility of the production lines without overinvesting in capacity flexibility. The main decision lies in determining which secondary product formats and bottle sizes the production lines must be equipped to produce. We obtained historical data on various information about the company's problem from their Enterprise Resource Planning system (SAP) as summarized in Table 4. This data set includes descriptions of product formats and bottle shapes, average demands for each format and shape as well as their standard errors, and the production line capacities. We observed that as the production lines are partially automated, their capacities are always adjusted to approximately match the average demands of the dedicated products.

To help solve Company Y's problem, we employ our VHC method as follows. We execute our algorithm with a budget of 32 links, which is double the number of production lines as the company wanted. We find that the best flexibility design is to isolate lines 6, 13, 15 and 16, and partition the remaining lines into three groups, as in Fig. 7. The output of this hub-and-chain design is 647,850 bottles per day compared to 587,349 bottles for the dedicated design, 627,620 bottles for the long chain design, 618,192 bottles for constraint sampling, and 652,565 bottles for the fully flexible design. We can see that such production line re-engineering can bring about capacity improvement of about 60,000 bottles per day, equivalent to the capacity of production line 11 and greater than the average capacity of 42,525 bottles per line per day. Deploying a fully flexible design can only add at most another 5000 bottles a day at enormous cost, whereas using the long
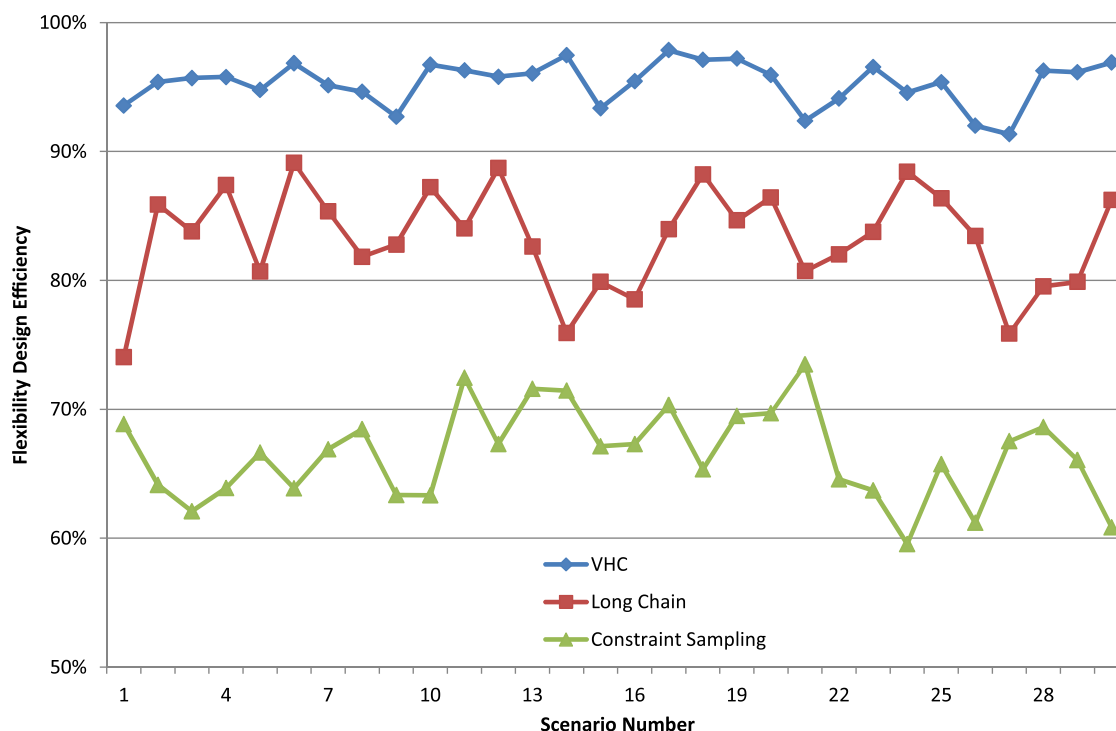
Fig. 6. Efficiency of three flexibility designs.

Table 4
Data for edible oil case study.

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16
Format	4L	5.435L	2L	5.1999L	2.5L	5.258L	4.5L	5L	5L	3L	1.8L	1.6L	6L	4.5L	6L	4L
Shape	S	R	R	R	R	R	R	S	R	R	R	R	S	S	R	R
Mean demand	48.4	70.8	65.5	48.8	21.6	4.7	27	96	96	30	60	67.5	8.1	18	6	12
Standard error	20.1	25.1	18.4	22	5.8	2.2	13	38.4	34.7	14	15.7	33	4	7	3	1.8
Capacity	48.4	70.8	65.5	48.8	21.6	4.7	27	96	96	30	60	67.5	8.1	18	6	12

chain or constraint sampling adds less than the average capacity per line.

5.2. Managerial insights

We are now ready to summarize the managerial implications that one can take away from this paper. First, we confirm that similar to identical systems, a partially flexible design can still perform very well in non-identical systems. This implies that indeed a little flexibility can go a long way in designing flexible manufacturing capacity. Second, demand variances of the various products, together with mean demand values, play a significant role in the performance of these partially flexible designs, particularly the long chain. For the long chain, we find that non-identical systems create the following asymmetries: (1) low demand mean results in a bottleneck effect, (2) low demand variance results in link inefficiency, and (3) high demand variance worsens long chain performance. These findings have important managerial implications such as (1) grouping products and lines into partially flexible sub-chains according to mean demand, (2) freeing up links from low variance products, and (3) using freed up links for high variance products. Third and last, the VHC method is computationally easy and graphically intuitive that managers are likely to deploy such approach and its resulting flexible designs. This implies that companies do not need to resort to sophisticated software to implement stochastic optimization or constraint sampling. With the hub-and-chain

design, they can also more easily explain and communicate the rationale behind the new flexible design.

6. Conclusions

In this paper, we study the process flexibility design problem for non-identical systems. While the long chain design can be employed in this setting, we find that in many cases, it results in poor performance. To devise a method that can improve on the long chain design, we first analytically examine the effect of non-identical demand distributions on system performance as well as long chain efficiency. We find a number of interesting results. First, long chain efficiency increases as the demand mean of any product increases, suggesting a bottleneck effect caused by low demand means. Second, long chain efficiency decreases with the demand variance of any product, implying the benefit of providing high-variance products with access to more capacity. Third, the efficiency of a flexibility link decreases as demand variance of the product for which the link is used decreases. This suggests that the link can be utilized more profitably elsewhere, such as for products with high demand variance.

With these insights, we develop a simple and graphically intuitive VHC method to generate flexible designs for non-identical systems. We consider two cases, with and without a fixed flexibility budget. In both cases, we show numerically that VHC already achieves at least 91% of the benefits of full flexibility. Moreover, VHC is, on average, 15% better than the long chain, and 38%
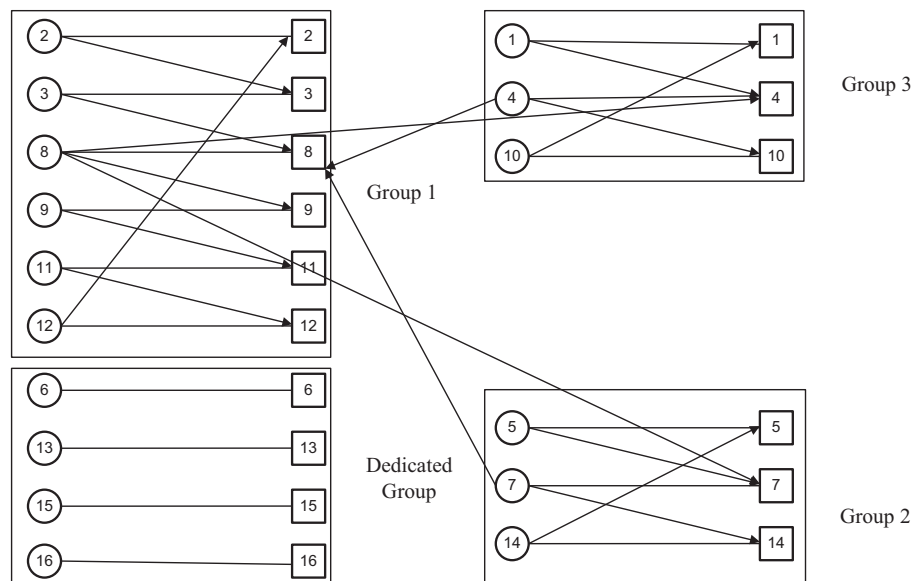
Fig. 7. The hub-and-chain design for edible oil case study.

better than constraint sampling. Furthermore, VHC uses only 1% of the computational requirements of the constraint sampling method. Finally, we implement the VHC method on a case study in the edible oil industry in China and find substantial benefits brought about by the hub-and-chain design. We also summarize some managerial insights that can be gleaned from this work.

References

Aksin, O. Z., & Karaesmen, F. (2007). Characterizing the performance of process flexibility structures. *Operations Research Letters, 35*(4), 477–484.

Buzacott, J. (1982). The fundamental principles of flexibility in manufacturing systems. In *Proceedings of the first international conference on flexible manufacturing sytems* (pp. 13–22). Brighton, UK: Elsevier.

Chen, X., Zhang, J., & Zhou, Y. (2015). Optimal sparse designs for process flexibility via probabilistic expanders. *Operations Research, 63*(5), 1159–1176.

Chou, M. C., Chua, G. A., & Teo, C. P. (2010a). On range and response: Dimensions of process flexibility. *European Journal of Operational Research, 207*(2), 711–724.

Chou, M. C., Chua, G. A., Teo, C. P., & Zheng, H. (2010b). Design for process flexibility: Efficiency of the long chain and sparse structure. *Operations Research, 58*(1), 43–58.

Chou, M. C., Chua, G. A., Teo, C. P., & Zheng, H. (2011). Process flexibility revisited: The graph expander and its applications. *Operations Research, 59*(5), 1090–1105.

Chou, M. C., Chua, G. A., & Zheng, H. (2014). On the performance of sparse process structures in partial postponement production systems. *Operations Research, 62*(2), 348–365.

Chou, M. C., Teo, C. P., & Zheng, H. (2008). Process flexibility: Design, evaluation and applications. *Flexible Services and Manufacturing Journal, 20*(1-2), 59–94.

Deng, T., & Shen, Z. J. (2013). Process flexibility design in unbalanced networks. *Manufacturing Service Operations Management, 15*(1), 24–32.

Deng, T., & Shen, Z. J. (2014). A demand-distribution-free analysis of process flexibility design. Working Paper. University of California at Berkeley.

Ford, Media. April (2009). New era of flexible manufacturing begins at ford's kentucky truck plant in louisville. http://www.media.ford.com/, Accessed 15.09.14.

Ford, Media. August (2010). Transformed michigan assembly plant sets new standard for flexible manufacturing at ford. http://www.media.ford.com/, Accessed 15.09.14.

Gupta, Y. P., & Goyal, S. (1989). Flexibility of manufacturing systems: concepts and measurements. *European Journal of Operational Research, 43*(2), 119–135.

Iravani, S. M. R., Kolfal, B., & Oyen, M. P. V. (2007). Call-center labor cross-training: It's a small world after all. *Management Science, 53*(7), 1102–1112.

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

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

LeBeau, P. (2010). Ford's flexibility is the story to watch. CNBC news. http://www.cnbc.com, Accessed 15.09.14.

Sethi, A., & Sethi, S. (1990). Flexibility in manufacturing: A survey. *The International Journal of Flexible Manufacturing Systems, 2*(4), 289–328.

Simchi-Levi, D., & Wei, Y. (2012). Understanding the performance of the long chain and sparse designs in process flexibility. *Operations Research, 60*(5), 1125–1141.

Wang, X., & Zhang, J. (2015). Process flexibility: A distribution-free bound on the performance of k-chain. *Operations Research, 63*(3), 555–571.

Woodyard, C. (2011). Ford focuses on flexibility at its factories. http://www.usatoday.com, Accessed 15.09.14.