Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Conveying Demand Information in Serial Supply Chains with Capacity Limits

Roman Kapuściński, Rodney P. Parker

informs
http://pubsonline.informs.org/journal/opre

OPERATIONS RESEARCH
Vol. 70, No. 3, May–June 2022, pp. 1485–1505
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Contextual Areas

Conveying Demand Information in Serial Supply Chains with Capacity Limits

Roman Kapuściński,[a] Rodney P. Parker[b,*]

[a] The University of Michigan, Ross School of Business, Ann Arbor, Michigan 48109; [b] Indiana University, Kelley School of Business, Bloomington, Indiana 47405
*Corresponding author

Contact: roman.kapuscinski@umich.edu, https://orcid.org/0000-0003-2972-3168 (RK); rodp@indiana.edu, https://orcid.org/0000-0001-7688-5122 (RPP)

Received: August 14, 2017
Revised: August 12, 2019; April 2, 2021; August 27, 2021
Accepted: November 25, 2021
Published Online in Articles in Advance: February 23, 2022

Area of Review: Operations and Supply Chains

https://doi.org/10.1287/opre.2021.2251

Abstract. For serial multiechelon systems subject to production capacity limits at every stage, we consider a class of modified echelon base stock (*MEBS*) policies. To evaluate information requirements of such systems, we consider two separate inventory management mechanisms operated in a decentralized manner. For ordering decisions, these mechanisms utilize local knowledge only and are distinguished by the timing of the orders being conveyed upstream from installation to installation. We demonstrate that these mechanisms can duplicate the shipment quantities in the modified echelon base-stock policy that uses full information. Thus, although full demand information will not be conveyed up the channel due to the demand censoring effects of capacity, we demonstrate that *sufficient* information about the market demand is conveyed via the orders. This suggests that local information is sufficient to make ordering decisions that replicate the policy's orders, a significant finding for implementing supply chain inventory policies in practice, where dynamic state information may not be readily accessible. We extend this local information result to serial channels with completely general capacity configurations acting under the corresponding echelon policies. We demonstrate the strong relationship between these two mechanisms that relate to serial capacitated channels of differing lengths. We augment our main results with two important extensions. (1) Because our focus is on *MEBS* policies, which are not necessarily optimal for longer supply chains, we evaluate their performance. We numerically show that they perform very well in general. We also provide upper and lower bounds, which further justify their strong performance. Given that these policies are close to optimal, that they are easy to interpret, and that they can operate with only local information, they are appealing in practical applications. (2) We compare the usage of local information to operate a capacitated system versus incentives to sustain such a system. We show that, similar to the noncapacitated case, it is possible to design an alternative incentive-compatible performance mechanism such that local managers will follow the centralized solution, albeit with more demanding information requirements.

Supplemental Material: The e-companion is available at https://doi.org/10.1287/opre.2021.2251.

Keywords: capacity limits • local information • inventory • multi-echelon supply chain • incentive compatible mechanism

1. Introduction and Literature

1.1. Problem Description and Preview of Results

In models with a central decision maker, the logical interpretation is that all the available information about the state of the system is used. On the other hand, in decentralized settings, the information available to one party may not be available to the others. In practice, this phenomenon of local information can also exist in centrally owned channels where local managers may covet and hoard their own information. We consider a retailer who sees demand information, but short of special arrangements that provide visibility of the retailer's inventory, further upstream tiers do not have access to that information. It has been shown (Axsäter and Rosling 1993) that it is possible to optimally operate a system with no capacity constraints when only local inventory is visible. Such a result depends on the policy parameters being given to local installations, possibly by a centralized decision maker. We evaluate whether such an assertion is true when capacity constraints are present. Although the literature on systems with capacity constraints is described below, a preview of the results is that while indeed the capacity limits do *censor* the demand information as it is passed up the supply chain, *sufficient* information (for modified echelon base-stock policies) is conveyed up the channel. We illustrate this through employing an alternative production mechanism that results in the shipment decisions of a desirable full-information inventory policy being mimicked, implying that the operators of individual installations can make decisions using *local information alone*. This implies that either a centrally- or decentrally operated channel with capacity limits may be operated in an informationally decentralized manner. In practice, imperfections in IT systems may prevent real-time distribution of accurate information, and our results suggest these challenges may be overcome by considering local information alone.

Timing of the ordering decisions is important for the policies we consider. In the literature, papers analyzing decentralized inventory systems tend to assume a simultaneous game. However, in the decentralized mechanisms we employ, we relax this assumption. In a simultaneous game, the retailer's decision is based on the current inventory. Because the supplier (and his supplier) have not yet received the order from the retailer, their decisions are based on what happened in the previous period. One can imagine two extreme highly coordinated systems where decisions are taken sequentially, rather than simultaneously. Imagine that installation 1, the retailer, makes ordering decisions at 8:01 a.m., installation 2, the supplier, makes ordering decisions at 8:02 a.m., etc. Clearly, the information about customer orders within minutes is transferred to the highest echelon. We will call such a system *Sequential Fast* (*SF*) because the orders are passed up the chain rapidly, within the same time period. One can imagine, however, another system where the higher echelon makes their decision before the lower echelon. For example, the retailer orders at 8:10 a.m., while her supplier orders at 8:09 a.m., and echelon N orders N minutes before 8:11 a.m. We will label such a system *Sequential Slow* (*SS*) because there is obviously a delay before an order is received by an upstream installation. We will analyze these two systems when capacity constraints are present and relate them to policies for centralized and decentralized systems. A summary of our results appears in Table 1. We show that the modified echelon base-stock policy (*MEBS*), which is a centrally operated inventory policy with full information is equivalent to Sequential Fast (*SF*), a decentrally operated policy with only local information. In addition, we show the Sequential Slow (*SS*) policy, which is also a decentrally operated policy with local information is equivalent to a longer *SF* system. Finally, we show that the primary result, that a centralized inventory policy may be operated with local information, is preserved under a variety of capacity configurations.

Table 1. Summary of Results

Operating mode	Centralized		Decentralized		
Policy	MEBS		SF		SS
Number of stages	N	≡	N	≡	M^*
Local information			✓		✓
Full information	✓				

* Where $2M - 1 = N$

Because our paper shows sufficiency of local information to mimic operations of a centralized system using *MEBS* policy, in an extension we justify the use of *MEBS* policies. The optimal policy for capacitated multistage systems remains intractable, but several papers (Parker and Kapuściński 2004, 2011; Janakiraman and Muckstadt 2009; Huh et al. 2010) have provided some structure, particularly describing *MEBS*. We demonstrate that *MEBS* policies, despite suboptimality, perform very well, and we also provide lower bounds on the optimal policy and upper bounds on the *MEBS* policy, which indicates that in the most demanding cases, *MEBS* and the optimal policy are very close to each other. Thus, from a practical point of view, *MEBS* policies seem to be the only realistic path to control capacitated systems. We show that *MEBS* policies are not only very efficient, but also can be operated using local information alone.

Also, as an extension, we discuss the incentives of local managers to maintain target inventory levels. This was the focus of Lee and Whang (1999) for uncapacitated systems, where they use the Veinott (1966) induced penalty cost functions to design an incentive compatibility mechanism. We show that the results for capacitated systems mirror the ones for uncapacitated systems and that an incentive compatible payment system may be designed. We describe the contingent information requirements of such a mechanism.

1.2. Literature Review

Because our paper shows sufficiency of local information to replicate operations of centralized multiechelon capacitated systems, it is important to put it in the broader context of relevant literature. We can naturally divide the literature into that involving inventory (with or without capacity limits), decentralized operation, information decentralization, and information sharing.

The literature dealing with inventory optimization in serial channels originated with Clark and Scarf's (1960) seminal work showing the optimality of an *echelon base-stock* policy, a result extended to the infinite-time horizon by Federgruen and Zipkin (1984). The optimality results pertaining to single-echelon inventory systems with capacity limits in the infinite horizon are due to Federgruen and Zipkin (1986a, b), where a base-stock level is sought when possible, limited by the capacity constraint, a policy later referred to as *modified base-stock*. The policy for multistage capacitated system was not characterized for many years. In highly visible papers Glasserman and Tayur (1994, 1995) assume suboptimal echelon-based-stock policies and provide a method to calculate parameters of such policies. They also illustrate that the echelon-based policy perform very well in capacitated systems. Several other visible papers, for example, Huh et al. (2016) use (noncapacitated) echelon-base-stock policies to shed light on capacitated systems. Parker and Kapuściński (2004) characterized the optimal inventory policy for a two-echelon system when the more limiting capacity is at the retailer and there is a one-period leadtime upstream of installation 2. The optimal policy, known as *modified echelon base-stock* (*MEBS*), has each echelon seeking a desired target but with the inventory at the higher installation limited to no more than what the retailer can process in one period. Although Speck and van der Wal (1991) provided convincing logic for the difficulty in characterizing policies with more stages, Janakiraman and Muckstadt (2009) characterized a number of properties of the optimal policy in the *N*-stage system, bounding the number of target inventory levels, and deriving a "two-tier base-stock" policy when there is a two-period leadtime upstream of installation 2 in a two-stage system. Huh et al. (2010) illustrate a sample-path correspondence between the echelon shortfalls in a serial system with capacity constraints and those of a single-stage system. Huh and Janakiraman (2010) characterize convexity properties of capacitated assembly systems when operating under base-stock policies while Angelus and Zhu (2013) describe the collapse of the same assembly system to a capacitated serial system, *à la* Rosling (1989). In this paper, similarly to the above-cited papers, we assume *MEBS* policies to operate a system with local information as well as full information.

Decentralized systems can use either "full" or "local" information. Many decentralized systems make the assumption that all information (inventory levels at all installations, demand realizations) is available to each firm, the firms just differ in their objective functions, including Parlar (1988), Lippman and McCardle (1997), Netessine and Rudi (2003), and Nagarajan and Rajagopalan (2009). For decentralized serial channels, Cachon and Zipkin (1999) considered echelon and local information policies in a decentralized setting of Federgruen and Zipkin (1984), finding and comparing the equilibria, for cases that correspond to different information about the current inventory. Parker and Kapuściński (2011) subsequently establish the decentralized policy for the same two-echelon system with capacity constraints, finding it is also *MEBS*. As opposed to papers analyzing decentralized systems, we do not focus on different objective functions, but instead we look at sufficiency of local information to execute *MEBS* policies.

Interestingly, although many papers consider multiple decision makers with differing objectives, they commonly assume access to all information. Lee and Whang (1999, p. 634) recognize that *information decentralizability* is a property of a viable measurement scheme and observe that the profit of an inventory policy utilizing richer echelon information cannot be worse than one using only local information, "the incremental benefit should be traded off with the cost of implementing information sharing." The informational requirements of Clark and Scarf's (1960) echelon policies and their descendants by definition require any installation to be aware of all inventories downstream or at least their sum. Axsäter and Rosling (1993) demonstrate that, in general, echelon inventory policies are superior to installation inventory policies when considering (Q, r) rules. However, more pertinent to our analysis is their result which shows, in uncapacitated systems, an equivalence between echelon and installation order-up-to policies, where the latter does not require visibility of all inventory levels through the channel. This result implies, as Lee and Whang (1999) highlight, that an optimal echelon policy may be replicated with an installation policy using local information, for serial systems under an assumption of stationarity and an initial inventory condition. In this paper, we demonstrate a similar result for serial systems subject to capacity constraints. A related paper, Shang et al. (2009), compares coordination mechanisms for serial systems operating with batch ordering with echelon, local, and quasi-local information scenarios, without any capacity constraints. Their result suggests quasi-local information (local information plus consumer demand information) can restore optimality while local information alone cannot. We do not include batch ordering but do include capacity constraints and illustrate that local operation can replicate full-information policies. One more relevant paper is Hariharan and Zipkin (1995) who show that an inventory replenishment leadtime is the complete opposite to a customer informational leadtime. This is relevant to our scenario because our *SS* scenario embraces a delay in the orders being transmitted upstream and we subsequently show an equivalence to a longer *SF* channel.

Lastly, the topic of information decentralization overlaps with the topic of *information sharing*, for which there is a substantial literature. Comprehensive reviews on information sharing appear in Chen (2003), Lau (2007), and Choi (2010). The general theme of the information sharing literature is to ascertain the value of information for varying degrees of sharing or incentives for sharing information in competitive scenarios with asymmetric information. This dynamic information includes the inventory levels at each installation and realized consumer demand. Gavirneni et al. (1999) is an example of early work in this area, with Li (2002), Li and Zhang (2008), and Ha et al. (2011) examining richer supply chain structures. Our perspective is not to examine the efficiency gain of sharing information but to establish the sufficiency of known *local* information in executing the optimal or equilibrium decisions. Specifically we investigate whether the actions derived from a *given* policy characterized with full information may be replicated by a policy with limited (local) information.

1.3. Structure of the Paper

The remainder of the paper is structured as follows. In Section 2, we introduce the model and associated notation. In Section 2.1, the Sequential Fast model is introduced and analyzed and the Sequential Slow model is similarly considered in Section 2.2. Section 2.3 establishes the strong relationship between the *SF* and *SS* systems. Section 3.1 demonstrates that the Modified Echelon Base-Stock policy performs extremely well numerically in three- and four-stage systems, justifying its earlier usage, whereas in Section 3.2 we use bounds to illustrate conditions when *MEBS* performs well. In Section 4, we describe how the results extend to some other supply chain configurations. Section 4.1 considers a system where the most constraining capacity is found internal to the serial channel, dividing it into two bands and shows the sufficiency of the information conveyed in the censored orders. This discussion is continued for serial systems with any capacity configuration in Section 4.2. In Section 5, we discuss incentives for local managers to adhere to the system target levels in capacitated systems. We conclude the paper with a discussion in Section 6.

2. Models and Analysis

In this section, we describe the settings of our model, which is serial channel with capacity limits at each installation, with two types of ordering systems (slow and fast). For each of these, we show that it can operate solely based on local information.

Consider a serial channel with multiple installations $i = 1, \dots, N$ where installation 1 delivers to the final customers and N is the uppermost installation. For the purposes of elucidation, we refer to the lowest installation as a "retailer." Each installation i represents a facility conducting a transformative production process, limited by capacity constraint, K_i. The installation closest to the customer (retailer) has the smallest capacity, $K_1 \le K_i$. The delivery leadtime between installations i and $i+1$ is one period for $i = 2, \dots, N$.[1] The leadtime between installations 1 and 2 can be a general integer number of periods, although for the purposes of exposition we will present it as one period. We assume stationarity in all economic and model parameters. Our global objective is to examine whether *MEBS* policies[2] that use full information can be replicated by local policies, that is, policies that use purely local information, in serial systems with capacity constraints.

Centralized systems naturally use all *full* information, where each installation knows all the economic parameters and state variables at the beginning of each period as well as demand realizations. The seminal paper of Clark and Scarf (1960) established that the optimal policy for a multistage system is an echelon policy. It defines *echelon* inventory ($X^j \doteq \sum_{i=1}^{j} x^i$ where x^i is the *installation i*'s local inventory) and shows that for the centralized system without capacity limits it is optimal to raise each of the echelon levels X^j to a desired target level Z^j. The echelon policy clearly requires information about inventory levels at other stages.

In a system with local information each installation cannot directly observe the current inventory level of their upstream supplier and, thus, cannot limit their order by the inventory available at that immediate upstream supplier.[3] Local information is where each installation knows its local inventory, the backlog owed to its immediate downstream customer installation, and the backlog owed to it by its immediate upstream supplier installation. The local inventory can simply be counted, the local backlog is simply what the installation owes the immediate downstream customer due to a current or past delivery shortfall, and the upstream backlog is simply the cumulative difference between what the installation has ordered and what has been delivered. These are quantities that the installation will naturally know merely through following the local base-stock policy. Local information applies to the *SF* and *SS* systems we analyze. Because some of the order amounts may exceed the available inventory, we assume that an order is an *authorization* to deliver immediately or whenever sufficient material becomes available. Therefore, any installation will track a cumulative backlog of unmet orders from their immediate customer and to their immediate supplier.

We will first analyze the full information model and then attempt to show a lack of dependency of decisions on any other firm's inventory level, under the auspices of stationary target installation up-to levels.

For systems operating with full or local information, the following assumption holds.

Assumption 1. *Each installation $i \in \{1, \dots, N\}$ operates under capacity limit K_i. $K_1 \le K_i$ for all i. We will label $K := K_1$ for convenience.*

Local information systems are the focus of the paper. The following assumptions hold only for systems operating with local information.

Assumption 2. *The firms operate under given stationary, local base-stock levels.*

Assumption 3. *No installation other than the retailer observes demands.*

Assumption 4. *Each firm ships whatever is in stock at the beginning of the period in order to satisfy the current order and any existing backlog of unmet orders from the immediate downstream customer.*

Assumption 1 reflects a common system configuration seen in the literature (Parker and Kapuściński 2004, 2011; Janakiraman and Muckstadt 2009). This assumption will be relaxed in Section 4. Assumption 2 consists of three elements. Firstly, base-stock levels are the focus of all theoretical work that builds on Clark and Scarf (1960), in numerical evaluations (e.g., Glasserman and Tayur 1995, 1996), and dominant in practice (for example, Hall and Rust 2000). Axsäter and Rosling (1993) assumed the parameters for the operating policy that can result in local information operation are given, possibly by a centralized decision maker. We likewise assume the policy parameters are given to the local installations. The stationarity policy parameters (base-stock levels) described in Assumption 2 can arise due to stationary parameters, stationary demand, and a long time horizon. These are utilized by Rosling (1989), Axsäter and Rosling (1993), and Chen (1998) amongst others. Section 3.1 is devoted to further evaluation and justification of this setting. Assumption 3 is the key element of our objective of examining whether sufficient demand information is conveyed up the supply chain. Assumption 4 describes the mechanics of operation where a firm will deliver goods demanded (either through the current order or the local backlog) from the physical inventory available at the beginning of the period.

The notation we use is as follows. At the beginning of period t, installation i has local inventory of x_t^i, has a local up-to target of z^i, and places an order with its upstream supplier of a_t^i. Unsatisfied market demand is expressed as negative x_t^1. For installations $i>1$, installation i has a local backlog of $B_t^{i-1,i}$ that represents orders from installation $i-1$ that were not met in periods prior to t, and $q_t^{i-1,i}$ that represents the shipment from installation i to installation $i-1$ in period t. Let Z^i be the centralized echelon i base-stock levels and $z^i \geq 0$ be the corresponding installation i base-stock levels. Z^i and z^i are stationary (Assumption 2). Set $Z^1 = z^1$ and $Z^i = Z^{i-1} + z^i$ for $i>1$. Equivalently, $Z^i = \sum_{j=1}^{i} z^j$ for $i>1$. We denote $x \wedge y = \min(x,y)$ and $x^+ = \max(0,x)$.

The state fully describing the system at the beginning of period t is:

$$(\mathbf{x}_t, \mathbf{B}_t) = (x_t^1, \ldots, x_t^i, \ldots, x_t^N, B_t^{12}, \ldots, B_t^{i-1,i}, \ldots, B_t^{N-1,N}).$$

For the purposes of clarity, we now formally define full and local information. Full information is consistent with the centralized operation of a system, where the operator knows all the inventory levels at the beginning of each period, the target levels, the realized demands (in past periods), and all economic costs. This is particularly pertinent in echelon inventory policies (echelon base-stock or *MEBS*, say) where an installation's echelon inventory consists of the sum of local and all downstream inventory. In reality, widespread knowledge of such dynamic quantities is unrealistic, because inventory levels will vary from time period to period. Local information is where an installation i (say, for $i>1$) is aware of quantities, which are only naturally known through operation of the installation: the local installation inventory (x_t^i), the local backlog the installation owes the immediate downstream installation ($B_t^{i-1,i}$), the local backlog the immediate supplier owes the installation ($B_t^{i,i+1}$), the local target level (z^i), and the immediate downstream order (a_t^{i-1}) (in the case of Sequential Fast but not in the case of Sequential Slow, as described below).

In any period, first all orders are placed and received. (The detailed timing is what differentiates *SS* and *SF* and is described in Sections 2.1 and 2.2.) Then demand arrives, and finally, at the end of the period, costs are evaluated. The inventory transition functions are as follows:

$$x_{t+1}^1 = x_t^1 + q_t^{12} - d_t, \tag{1}$$

$$x_{t+1}^i = x_t^i + q_t^{i,i+1} - q_t^{i-1,i}, i = 2, \ldots, N. \tag{2}$$

The quantity shipped between installations i and $i-1$ ($i = 2, \ldots, N$) is as follows:

$$q_t^{i-1,i} = (a_t^{i-1} + B_t^{i-1,i}) \wedge x_t^i, \tag{3}$$

and $q_t^{N,N+1} = a_t^N$. The new increment to the backlog between installations $i-1$ and i ($i = 2, \ldots, N$) is as follows:

$$C_{t+1}^{i-1,i} = a_t^{i-1} - (q_t^{i-1,i} - B_t^{i-1,i})^+. \tag{4}$$

The backlog transition function between installations $i-1$ and i ($i = 2, \ldots, N$) is as follows:

$$\begin{aligned} B_{t+1}^{i-1,i} &= (B_t^{i-1,i} - q_t^{i-1,i})^+ + C_{t+1}^{i-1,i} = B_t^{i-1,i} + a_t^{i-1} - q_t^{i-1,i} \\ &= (B_t^{i-1,i} + a_t^{i-1} - x_t^i)^+. \end{aligned} \tag{5}$$

Equations (1–5) describe the evolution of the system from period t to period $t+1$, whether it be under the operation of *SF* or *SS*, reflecting the single-period delivery leadtimes. Equation (3) describes the shipment quantity given the order amount, which is the lesser of the on-hand inventory and what is owed (current order plus backorder). Hereafter, the specific systems will be denoted with visual mnemonics: system *SF* will be denoted with an arrow indicating speed (e.g., $\vec{a}_t^i$) and system *SS* will be denoted with a tilde implying the slowness of an undulating wave (e.g., $\tilde{a}_t^i$).

We will now consider the two models *SF* and *SS* in detail.

2.1. Sequential Fast Model

The Sequential Fast system is a serial decentralized multistage system, where each stage orders up to a stationary target *after* receiving the order from the immediate downstream stage. An exception to this is the retailer who chooses her order quantity *prior* to the demand realization.[4]

The timing of the N-stage system is as follows:

1. For $i = 1$ to N, installation i places an order;
2. For $i = 1$ to N, the order is shipped (from installation $i+1$) and arrives (to installation i), that is, as much of installation i's current order and its backlog is satisfied as possible from installation $i+1$'s inventory at the beginning of the period, $\vec{x}_t^{i+1}$;
3. demand is realized and is only seen by the retailer; and
4. costs are assessed.

Specifically, the order quantities are ($i = 2, \ldots, N-1$):

$$\vec{a}_t^1 = \left[\vec{z}^1 - \left(\vec{x}_t^1 + \vec{B}_t^{12}\right)\right] \wedge K \tag{6}$$

$$\vec{a}_t^i = \vec{z}^i - \left[\left(\vec{x}_t^i + \vec{B}_t^{i,i+1}\right) - \left(\vec{q}_t^{i-1,i} + \vec{B}_{t+1}^{i-1,i}\right)\right] \tag{7}$$

$$= \vec{z}^i - \left[\vec{x}_t^i - \left(\vec{a}_t^{i-1} + \vec{B}_t^{i-1,i}\right) + \vec{B}_t^{i,i+1}\right] \tag{8}$$

$$\vec{a}_t^N = \vec{z}^N - \vec{x}_t^N + \vec{B}_t^{N-1,N} + \vec{a}_t^{N-1} \tag{9}$$

Equations (6–9) in conjunction with the transition functions, Equations (1–5), fully characterize the operation of the *SF* operating regime. Note that we do not explicitly impose the capacity constraints on the order quantities or shipment quantities in installations above the retailer. Also, at the retailer we only impose that the order quantity is limited to K and we do not constrain her incoming shipment quantity. Although these capacity constraints are not imposed here, the resulting dynamics are unaffected by their omission (see Proposition 1).

The following definition provides equivalences when inventory policies are mated with demands. This facilitates comparisons of policy-demand combinations in our subsequent results.

Definition 1.

a. *We consider pairs (Policy1, demand1) and (Policy2, demand2) as equivalent (≡), if all shipments between installations and all installation inventories are identical.*

$$\textit{Policy1} \oplus \textit{demand1} \equiv \textit{Policy2} \oplus \textit{demand2}$$

b. *The two pairs will be considered x-Retailer equivalent ($\overset{xR}{\equiv}$), if all shipments between installations and all installation inventories except (excluding) those of the retailer are identical. We denote the equivalence as*

$$\textit{Policy1} \oplus \textit{demand1} \overset{xR}{\equiv} \textit{Policy2} \oplus \textit{demand2}$$

c. *If, for any demand, we have Policy1 ⊕ demand ≡ Policy2 ⊕ demand, we will state that Policy1 and Policy2 are equivalent.*

We notice that if $Policy1 \oplus d \overset{xR}{\equiv} Policy2 \oplus d$, then $Policy1 \oplus d \equiv Policy2 \oplus d$.

For the purposes of clarity, let us define the following inventory policy.

Definition 2. *The modified echelon base-stock policy, or MEBS (Parker and Kapuściński 2004) can be written as the mapping from initial echelon inventories (X^j) to echelon ordering levels (Y^j) where $Y^j \geq X^j$ for all j, attempting to reach the target echelon levels (Z^j):*

$$Y^1 = \min(Z^1, X^1 + K, X^2)$$
$$Y^j = \min(Z^j, Y^{j-1} + K, X^{j+1}) \text{ for } j = 2, \ldots, N-1$$
$$Y^N = \min(Z^N, Y^{N-1} + K).$$

We call this policy MEBS(K). When we want to explicitly show the dependence on targets Z, we will refer to the policy as MEBS(Z,K).

We observe that $MEBS(\infty)$ is equivalent to the echelon base-stock policy of Clark and Scarf (1960) (i.e., no capacity constraints). The critical factors that differentiate $MEBS(K)$ from *SS* and *SF* are as follows:

a. *MEBS* is a full information policy and the *MEBS* policy does not impose a capacity constraint at any installation but instead imposes that each installation will never *stock* more than K, $Y^j \leq Y^{j-1} + K$ for $j > 1$, thus introducing a "band" limitation.

b. *SS* and *SF* are local information policies and they impose capacity constraints upon the lowest installation (only) $y^1 - x^1 \leq K$.

Under the operation of an *SF* system, we offer a number of propositions. In order to clearly describe the dependence of policies on capacity level K, we will use $SF(K)$ and $SS(K)$. We first describe an existing result for uncapacitated systems.

Consider first a system without capacity constraints. Note that Axsäter and Rosling (1993) effectively assumes *SF*. Their proposition 1, for batch sizes of one unit, states that any echelon base-stock policy can be replicated by a Sequential Fast System. Using our notation, this means for any demand d, $SF(\infty) \oplus d \equiv MEBS(\infty) \oplus d$. This implies that when the up-to levels are identical, the quantities delivered in both systems will be identical: $\vec{q}_t^{i,i+1}(SF(\infty)) = a_t^i(MEBS(\infty))$ for $i \leq N$. The nontechnical explanation for this is that while $MEBS(\infty)$ requires knowing the whole vector of inventories across all installations, $SF(\infty)$ achieves the same result by placing unconstrained orders up to ideal target levels and remembering backlogs of locally unsatisfied orders. Thus, as expected for a system with no capacity constraints, the sufficient information about demand is perfectly passed to higher echelons.

The following definitions will prove useful. It is convenient to modify the periodic demand d_t to $d_t(K)$, where the latter has any raw demand in excess of K carried into future periods (the time subscript will be omitted when the demand process is being applied in conjunction with an inventory policy, as defined in Definition 1). We use the same logic to define $z^i(K)$ and $Z^i(K)$:

Definition 3.

a. *Let d_t be the periodic demand, with $D_t = \sum_{i=1}^{t} d_i$. We define $d_t(K)$ inductionally, as $d_t(K) = (D_t - D_{t-1}(K)) \wedge K$, where $D_t(K) = \sum_{i=1}^{t} d_i(K)$.*

b. *Let $z^i(K) = (Z^i - Z^{i-1}(K)) \wedge K$ for $i > 1$ where $Z^i(K) = \sum_{j=1}^{i} z^j(K)$.*

We now consider capacity constraints in a serial system for an N-stage system operating under the *MEBS* policy. The proof will follow from more general results that bridge several models: Sequential Fast, the centralized echelon base-stock policy $MEBS(\infty)$ (derived in Clark and Scarf 1960), and the Modified Echelon Base-Stock policy $MEBS(K)$.

Theorem 1. $SF(K) \oplus d \equiv MEBS(K) \oplus d$.

The equivalency $SF(K) \oplus d \equiv MEBS(K) \oplus d$ will follow from the following relationships: $SF(K) \oplus d \overset{xR}{\equiv} SF(\infty) \oplus d(K) \equiv MEBS(\infty) \oplus d(K) \equiv MEBS(K) \oplus d(K) \overset{xR}{\equiv} MEBS(K) \oplus d$. The second equivalence, for infinite capacities, $SF(\infty) \oplus d(K) \equiv MEBS(\infty) \oplus d(K)$, is shown in Axsäter and Rosling (1993) and, therefore, we can directly leverage it.

The full information optimal policy is known for a two-stage system.

Corollary 1. *The optimal centralized full information policy for a stationary capacitated two-stage supply chain, modified echelon base-stock policy, or MEBS(K) (Parker and Kapuściński 2004) is equivalent to (can be replicated by) Sequential Fast, SF(K).*

We will use the following condition in some of the following results. It affects the inventory levels and local backlogs in the first period of the horizon only. Ordering up to stationary local levels will result in local inventory below the targets.[5] Note that stationary policies with initial inventory conditions have been used before (see Rosling 1989 and Axsäter and Rosling 1993, for example).

Condition 1. *Let $x_1^i = z^i$ and $B_1^{i,i+1} = 0$ for $i = 1, \ldots, N$.*

Lemmas 1–4 establish the bridging relationships needed for Theorem 1. Lemma 1 states installation 1 orders the cumulative demands in the previous periods in excess of K, and that an uncapacitated system with the censored part of the cumulative demands is equivalent to a capacitated system with the raw demand for all installations above the retailer. Lemma 2 states that the *MEBS* policy with the censored target levels is equivalent to the *MEBS* policy with the regular target levels so long as the initial inventory levels are low enough. Lemma 3 states the uncapacitated system policy with the censored demand process is equivalent to *MEBS*, the capacitated system policy. Lemma 4 states the *MEBS* policy applied to the capacitated system with the censored demand process is equivalent to the same with the uncensored demand process.

Lemma 1. *Under* Condition 1, *(a) $\vec{a}_t^1 = d_{t-1}(K)$, and (b) $SF(\infty) \oplus d(K) \overset{xR}{\equiv} SF(K) \oplus d$.*

Proof. We first show (a) by induction that

$$d_{t-1}(K) = \vec{a}_t^1. \tag{10}$$

Assume this holds for t. Clearly,

$$\vec{x}_{t+1}^1 = \vec{x}_1^1 + \sum_{i=1}^{t} \vec{q}_i^{12} - \sum_{i=1}^{t} d_i = \vec{z}^1 + \sum_{i=1}^{t} \vec{q}_i^{12} - \sum_{i=1}^{t} d_i \tag{11}$$

and

$$\vec{B}_{t+1}^{12} = \vec{B}_t^{12} + \vec{a}_t^1 - \vec{q}_t^{12} = \sum_{i=1}^{t} \vec{a}_i^1 - \sum_{i=1}^{t} \vec{q}_i^{12}. \tag{12}$$

Thus,

$$\begin{aligned}
\vec{a}_{t+1}^1 &= \left[\vec{z}^1 - (\vec{x}_{t+1}^1 + \vec{B}_{t+1}^{12})\right] \wedge K \\
&= \left[\vec{z}^1 - \vec{z}^1 - \sum_{i=1}^{t} \vec{q}_i^{12} + \sum_{i=1}^{t} d_i - \vec{B}_{t+1}^{12}\right] \wedge K \\
&= \left[-\sum_{i=1}^{t} \vec{a}_i^1 + \sum_{i=1}^{t} d_i\right] \wedge K \\
&= \left[D_t - \sum_{i=1}^{t} \vec{a}_i^1\right] \wedge K = d_t(K)
\end{aligned}$$

where the first equality arises from Equation (11), the second arises from Equation (12), and the fourth from Equation (10). (b) The retailer's orders in $SF(K) \oplus d$ are the same as in $SF(\infty) \oplus d(K)$. Because all higher stages (installations $i \geq 2$) are identical in $SF(\infty)$ and $SF(K)$ systems, we have $SF(K) \oplus d \overset{xR}{\equiv} SF(K) \oplus d(K)$. □

Lemma 2. *For any demand d and $X_1^i \leq Z^i(K)$ for $i = 1, \ldots N$, $MEBS(Z, K) \oplus d \equiv MEBS(Z(K), K) \oplus d$.*

Proof. Consider two systems, A and B. System A operates under $MEBS(Z,\ K)$ and system B operates under $MEBS(Z(K), K)$. Note that $Z^i(K) = Z^{i-1}(K) + (Z^i - Z^{i-1}(K)) \wedge K$. Consider a specific period (index omitted) and then installations 1 to N. For installation 1, $Z^1 = Z^1(K)$ and so $Y^{1A} = Y^{1B}$ because both systems start from the same inventory vector. Whenever the inventory availability limit binds, the ordering decision will be common between the systems so we will omit for conciseness. Assume $Y^{i-1,A} = Y^{i-1,B}$.

For installation i, if $X^i < Z^{i-1}(K)$ then $Y^{iA} = Y^{iB} = Y^{i-1} + K$. If $X^i \geq Z^{i-1}(K)$, $Y^{iA} = Y^{iB} = Z^i(K)$ and we have two cases (i) $Z^i - Z^{i-1}(K) < K$ resulting in $Y^{iA} = \min(Z^i, X^{i+1}) = \min(Z^i(K), X^{i+1}) = Y^{iB}$, and (ii) $Z^i - Z^{i-1}(K) \geq K$ resulting in $Y^{iA} = \min(Z^{i-1}(K) + K, X^{i+1})$. Thus, $Y^{iA} = \min(\min(Z^i, Z^{i-1}(K) + K), X^{i+1}) = \min(Z^{i-1}(K) + \min(Z^i - Z^{i-1}(K), K), X^{i+1}) = \min(Z^i(K), X^{i+1}) = Y^{iB}$. The ordering decisions will be the same in systems A and B at all echelons, so each system will begin with identical inventory vectors in the following period. □

From now on, without loss of generality we replace Z with $Z(K)$.

Lemma 3. *MEBS(∞)* $\oplus\, d(K) \equiv$ *MEBS(K)*$\oplus\, d(K)$.

Proof. While *MEBS*(∞) defines $Y_t^i = X_t^{i+1} \wedge Z^i$, *MEBS*($K$) defines $Y_t^i = (Y_t^{i-1} + K) \wedge X_t^{i+1} \wedge Z^i$. Thus, it is sufficient to show that in *MEBS*(∞), $x_t^i \leq K$ for all $i > 1$, which immediately implies that the additional restriction does not change any dynamics. To justify this result, we will consider a sequence of policies π^i. Policy π^i is defined as follows: for all installations $k \leq i-1$, it operates under *MEBS*(∞), and for all installations $k \geq i$ it operates under *MEBS*(K). The proof is by demonstrating that for all i, $\pi^i = \pi^{i+1}$.

Consider policy π^{i+1}. For installation i, recall that $x_{t+1}^i = x_t^i + a_t^i - a_t^{i-1}$. Consider two cases:

a. $a_t^{i-1} < x_t^i$ (not all inventory is shipped to installation $i-1$), which implies $Y_t^{i-1} = Z^{i-1}$ (the target inventory is reached). Because $Y_t^i \leq Z^i$, we have $x_{t+1}^i = Y_t^i - Y_t^{i-1} \leq Z^i - Z^{i-1} \leq K$, by construction of levels Z^i and Z^{i-1}.

b. $a_t^{i-1} = x_t^i$ (all inventory is shipped to installation $i-1$), which implies $x_{t+1}^i = a_t^i$. Clearly $a_t^i \leq x_t^{i+1}$. However, in policy π^{i+1}, we have $x_t^{i+1} \leq K$.

Thus, in both cases $x_t^i \leq K$ and $\pi^{i+1} = \pi^i$. □

Lemma 4. *MEBS*(K)$\oplus\, d(K) \overset{xR}{\equiv}$ *MEBS*(K)$\oplus\, d$.

Proof Each system operates *MEBS(K)* indicating that the installation inventories for $i > 1$ are $x_t^i \leq K$. At the start of period t, both systems have identical inventories other than at the retailer. Let us adopt the nomenclature of A for the system operating under demand stream $d(K)$ and B for d.

We show that in all periods $a_\tau^{1A} = a_\tau^{1B}$ and $X_{\tau+1}^{1A} - X_{\tau+1}^{1B} = D_\tau - D_\tau(K)$. Clearly it holds for $\tau = 0$. Let $\tau \geq 1$ and assume $a_{\tau-1}^{1A} = a_{\tau-1}^{1B}$ and $x_\tau^{1A} - x_\tau^{1B} = D_{\tau-1} - D_{\tau-1}(K)$. Consider two cases: (a) $D_{\tau-1} = D_{\tau-1}(K)$, and (b) $D_{\tau-1} > D_{\tau-1}(K)$. In (a), the induction assumption gives $x_\tau^{1A} = x_\tau^{1B}$ and so $a_\tau^{1A} = x_\tau^2 \wedge (z^1 - x_\tau^{1A}) = x_\tau^2 \wedge (z^1 - x_\tau^{1B}) = a_\tau^{1B}$. For (b), we show that $d_{\tau-1}(K) = K$. Since $D_{\tau-1} > D_{\tau-1}(K)$, $\sum_{i=1}^{\tau-1} d_i > \sum_{i=1}^{\tau-1} d_i(K)$ implying $d_{\tau-1} + \sum_{i=1}^{\tau-2} d_i - \sum_{i=1}^{\tau-2} d_i(K) > d_{\tau-1}(K) = (D_{\tau-1} - D_{\tau-2}(K)) \wedge K = (d_{\tau-1} + D_{\tau-2} - D_{\tau-2}(K)) \wedge K = K$. Now, since $D_{\tau-1} > D_{\tau-1}(K)$, $x_\tau^{1A} > x_\tau^{1B}$ and so $z^1 - x_\tau^{1B} > z^1 - x_\tau^{1A} = z^1 - x_{\tau-1}^{1A} - a_{\tau-1}^{1A} + d_{\tau-1}(K) \geq d_{\tau-1}(K) = K \geq x_\tau^2$ so $a_\tau^{1A} = x_\tau^2 \wedge (z^1 - x_\tau^{1A}) = x_\tau^2 = x_\tau^2 \wedge (z^1 - x_\tau^{1B}) = a_\tau^{1B}$. Now $x_{\tau+1}^{1A} - x_{\tau+1}^{1B} = x_\tau^{1A} + a_\tau^{1A} - d_\tau(K) - (x_\tau^{1B} + a_\tau^{1B} - d_\tau) = x_\tau^{1A} - x_\tau^{1B} + d_\tau - d_\tau(K) = D_{\tau-1} - D_{\tau-1}(K) + d_\tau - d_\tau(K) = D_\tau - D_\tau(K)$. Because the retailer orders are identical for systems A and B, all upstream inventories will remain identical between both systems. □

Proposition 1. *For SF(K) with installation target levels z^i, when Condition 1 holds:*

a. *the orders from each installation are identical, $\vec{a}_t^i = \vec{a}_t^{i+1}$; and*

b. *the amount shipped from installation $i+1$ to i does not exceed K. That is, $\vec{q}_t^{i,i+1} \leq K$.*

Proof. For (a), assume $\vec{x}_t^i = \vec{z}^i - \vec{B}_t^{i,i+1} + \vec{B}_t^{i-1,i}$ for $i = 2, \ldots, N$. Installation 1 places an order $\vec{a}_t^1 = [\vec{z}^1 - (\vec{x}_t^1 + \vec{B}_t^{12})] \wedge K \leq K$. Installation i orders $\vec{a}_t^i = \vec{z}^i - (\vec{x}_t^i + \vec{B}_t^{i,i+1}) + (\vec{a}_t^{i-1} + \vec{B}_t^{i-1,i}) = \vec{z}^i - (\vec{z}^i - \vec{B}_t^{i,i+1} + \vec{B}_t^{i-1,i} + \vec{B}_t^{i,i+1}) + (\vec{a}_t^{i-1} + \vec{B}_t^{i-1,i}) = \vec{a}_t^{i-1}$. Now consider the inventory transition function for installation $i > 1$, $\vec{x}_{t+1}^i = \vec{x}_t^i + \vec{q}_t^{i,i+1} - \vec{q}_t^{i-1,i}$. Add $\vec{B}_{t+1}^{i,i+1} - \vec{B}_{t+1}^{i-1,i}$ to both sides: $\vec{x}_{t+1}^i + \vec{B}_{t+1}^{i,i+1} - \vec{B}_{t+1}^{i-1,i} = \vec{x}_t^i + \vec{q}_t^{i,i+1} - \vec{q}_t^{i-1,i} + \vec{B}_{t+1}^{i,i+1} - \vec{B}_{t+1}^{i-1,i}$. Because, $\vec{B}_{t+1}^{i,i+1} + \vec{q}_t^{i,i+1} = \vec{a}_t^i + \vec{B}_t^{i,i+1}$, using the induction assumption, $\vec{x}_t^i = \vec{z}^i - \vec{B}_t^{i,i+1} + \vec{B}_t^{i-1,i}$, we have $\vec{x}_{t+1}^i + \vec{B}_{t+1}^{i,i+1} - \vec{B}_{t+1}^{i-1,i} = \vec{x}_t^i + \vec{a}_t^i + \vec{B}_t^{i,i+1} - \vec{a}_t^{i-1} - \vec{B}_t^{i-1,i} = \vec{z}^i + \vec{a}_t^i - \vec{a}_t^{i-1} = \vec{z}^i$. The induction assumption holds trivially for $t = 1$.

Because the retailer limits her order quantity to be no greater than K, this property is inherited by each higher installation: $\vec{a}_t^i = \vec{a}_t^1 \leq K$ for all $i > 1$.

The proof of (b) is by induction on the echelon, starting from the uppermost installation. For installation N, $\vec{B}_t^{N,N+1} = 0$ (because the outside supplier has an unlimited supply, $\vec{x}_t^{N+1} = \infty$) and $\vec{q}_t^{N,N+1} = (\vec{a}_t^N + \vec{B}_t^{N,N+1}) \wedge \vec{x}_t^{N+1} = \vec{a}_t^N \leq K$. Assume $\vec{q}_t^{i+1,i+2} \leq K$. Then we have two cases: (b1) $\vec{B}_t^{i,i+1} = 0$, and (b2) $\vec{B}_t^{i,i+1} > 0$. Under case (b1), $\vec{q}_t^{i,i+1} = (\vec{a}_t^i + \vec{B}_t^{i,i+1}) \wedge \vec{x}_t^{i+1} \leq \vec{a}_t^i \leq K$. Under case (b2), we wish to show that if there is a backorder at installation $i+1$ (owed to installation i), installation $i+1$ exhausted its supply, prior to receiving its own shipment, in the previous period. Consider $\vec{B}_t^{i,i+1} = \vec{B}_{t-1}^{i,i+1} + \vec{a}_{t-1}^i - \vec{q}_{t-1}^{i,i+1}$. $\vec{B}_t^{i,i+1} > 0$ if $\vec{q}_{t-1}^{i,i+1} < \vec{B}_{t-1}^{i,i+1} + \vec{a}_{t-1}^i$ and therefore $\vec{q}_{t-1}^{i,i+1} = \vec{q}_t^{i,i+1} = (\vec{a}_t^i + \vec{B}_t^{i,i+1}) \wedge \vec{x}_t^{i+1} = \vec{x}_{t-1}^{i+1}$. Now, $\vec{x}_t^{i+1} = \vec{x}_{t-1}^{i+1} + \vec{q}_{t-1}^{i+1,i+2} - \vec{q}_{t-1}^{i,i+1} = \vec{q}_{t-1}^{i+1,i+2} \leq K$. Because installation $i+1$ has no more than K units in period t, he cannot ship more than K units. □

The orders will be identical for each installation when starting off at the target level. It should be noted that when those initial inventory levels are not at the targets, it takes a finite number of periods to reach a situation where the target levels are achieved, when expected demand is less than K. The proof of the central result of the section now follows.

Proof of Theorem 1. Using Lemma 1, Lemma 3, Lemma 4, and Axsäter and Rosling's (1993) proposition 1, we have the following relationships: $SF(K)\oplus d \overset{xR}{\equiv} SF(\infty)\oplus d(K) \equiv MEBS(\infty)\oplus d(K) \equiv MEBS(K)\oplus d(K) \overset{xR}{\equiv} MEBS(K)\oplus d$. Although the two policies are equivalent, we highlight that *MEBS* uses the concept of the inventory band (range of installation inventory levels), whereas *SF* uses the concept of capacity (how much may be processed in a period). □

The above theorem is a key result: local information in $SF(K)$ is sufficient to replicate the $MEBS(K)$ policy that uses full information. This result is interesting because the upstream installations never become aware of the realized demand observed by the retailer. The "censored" demand that the retailer passes upward in her order is *sufficient* for all other installations to make decisions that result in shipments corresponding to $MEBS(K)$. The suggested mechanism is quite natural (tracking any unmet orders as a local backlog), with a recognition that, at all installations above the retailer, the backlog may not be combined in the same state variable as local inventory. Although the retailer's orders do not exceed K, not-satisfied orders become an authorization to deliver goods in later periods. The theorem shows that all the upper installations *naturally* limit their shipments through the following means: Even though an installation may be authorized to send more goods than K (through the incoming order and current backlog), he does not have more than K inventory to send, and when he has more than K inventory to ship, his customer will never order more than K.

2.2. Sequential Slow Model

In the Sequential Slow system, each firm chooses their order quantity prior to receiving their customer's order. The timing of the N-stage system is identical as for the Sequential Fast system with the exception that the sequence of installations in step 1 is reversed, as follows:

1. For $i = N$ to 1, installation i places order;
2. For $i = 1$ to N, the order arrives, that is, as much of installation i's current order and its backlog is satisfied as possible from installation $i+1$'s inventory at the beginning of the period, $\tilde{x}_t^{i+1}$;
3. demand is realized and is only seen by the retailer; and
4. costs are assessed.

The equations that govern the operation of the system follow. The order quantities are ($i = 2, \ldots, N-1$):

$$\tilde{a}_t^1 = [\tilde{z}^1 - (\tilde{x}_t^1 + \tilde{B}_t^{12})] \wedge K \tag{13}$$

$$\tilde{a}_t^i = \tilde{z}^i - \tilde{x}_t^i + \tilde{B}_t^{i-1,i} - \tilde{B}_t^{i,i+1} \tag{14}$$

$$\tilde{a}_t^N = \tilde{z}^N - \tilde{x}_t^N + \tilde{B}_t^{N-1,N} \tag{15}$$

Again, notice that Equations (13–15) with the transition functions (Equations (1–5)) fully characterize the *SS* operating regime. The derivation of $\tilde{a}_t^i$ is similar to that for the Sequential Fast order quantity (Equations (6–9)) other than this quantity is chosen *before* receiving the order from installation $i-1$. The following result involves both *SF* and *SS* systems, whereas the remaining results characterize various properties of the *SS* system. (The proofs of the following results may be found in Online Appendix A.)

Lemma 5 (SF and SS). *Under Condition 1,* $\vec{a}_t^1 = \tilde{a}_t^1 = d_{t-1}(K)$.

Lemma 6. *Under Condition 1,* $SS(K)\oplus d \overset{xR}{\equiv} SS(K)\oplus d(K) \equiv SS(\infty)\oplus d(K)$.

Lemma 7 (for *d(K)* and under Condition 1). $\sum_{k\le i}\tilde{a}_t^k = Z^i - X_t^i - \tilde{B}_t^{i,i+1}$.

Lemma 8. *Under Condition 1,* $\tilde{a}_{t+1}^{i+1} = \tilde{a}_t^i$ *for* $i = 1, \ldots, N-1$.

2.3. Comparison of *SF* and *SS* Models

It is unsurprising that the *SF* system could be more effective than an equivalently sized *SS* system. In this section, we refine this comparison and distill some specific properties of two analogous systems. Intuitively, the *SS* system passes information more slowly than in an *SF* system because each installation orders before receiving their immediate customer's order. Thus, in an *SF* system where a customer's order is incorporated into the firm's order, two *SF* installations (nonretailer) would approximate a single *SS* installation.

For the following results, consider an *SS* system with N installations and an *SF* system with $2N-1$ installations: (i) $\vec{z}^{2i-2} + \vec{z}^{2i-1} = \tilde{z}^i$, and (ii) $\vec{z}^{2i-2} = \tilde{z}^i$ if $\tilde{z}^i < K$ and $\vec{z}^{2i-2} = K$ if $\tilde{z}^i \ge K$, for all $i > 1$.

Proposition 2. *For an installation $i = 2, \ldots, N$ in SS,* $\tilde{x}_t^i + d_{t-i} - \tilde{B}_t^{i-1,i} + \tilde{B}_t^{i,i+1} = \tilde{z}^i$ *where* $\tilde{B}_t^{N,N+1} = 0$.

Proof. The result follows from Equation (14) and Lemmas 5 and 8. □

Proposition 3. *For two paired installations $i = 2, \ldots, N$ in SF,* $\vec{x}_t^{2i-2} + \vec{x}_t^{2i-1} + \vec{B}_t^{2i-1,2i} - \vec{B}_t^{2i-3,2i-2} = \vec{z}^{2i-2} + \vec{z}^{2i-1}$ *where* $\vec{B}_t^{2N-1,2N} = 0$.

Proof. The result follows by adding Equation (8) for two consecutive stages and Proposition 1. □

Theorem 2. *For* $i = 2, \ldots, N$,

$$a(t,i) \qquad \tilde{x}_t^i = \vec{x}_{t-i+1}^{2i-2} + \vec{x}_{t-i+1}^{2i-1} - \vec{q}_{t-i+1}^{2i-3,2i-2}$$

$$b(t,i) \quad \tilde{B}_t^{i-1,i} = \vec{B}_{t-i+2}^{2i-3,2i-2}$$

$$c(t,i) \quad \tilde{q}_t^{i-1,i} = \vec{q}_{t-i+2}^{2i-3,2i-2}$$

Proposition 4 (Relation between SF and SS). *Consider an SS system with N installations and an SF system with* $2N-1$ *installations,*

i.
$$\tilde{x}_t^N + d_{t-N}(K) = \vec{x}_{t-N+2}^{2N-2} + \vec{x}_{t-N+2}^{2N-1}$$

ii.
$$\sum_{j=2}^{N} \tilde{x}_t^j + \sum_{j=2}^{N} d_{t-j}(K) = \sum_{j=2}^{2N-1} \vec{x}_t^j$$

and

iii.
$$\tilde{x}_t^1 = \vec{x}_t^1.$$

These propositions draw comparisons between the *SF* and *SS* systems. In short, Proposition 3 shows that an N-stage *SS* system can be mimicked by a longer $(2N-1)$-stage *SF* system, where system *SS*'s installation $i>1$ corresponds to system *SF*'s installations $2i-2$ and $2i-1$ together. That is, the corresponding *SF* system will have double the number of nonretailer installations. Arguably, the most important elements of the theorem are that the shipments, backlogs, and the inventory presented at the retailer will be identical while the upstream inventories in the *SF* system will have a clear relationship to those in the *SS* system, as described in $a(t, i)$ of Theorem 2.

It is not surprising that a longer supply chain (the *SF* system) possesses more inventory than the shorter one (the *SS* system). Proposition 4 shows the difference between the sum of inventory in the respective systems is simply the sum of demands over the previous $N-1$ periods. This intuitively arises due to the additional number of periods of demand experienced as goods traverse the longer *SF* channel. Notably, there is now a relationship between the mechanics of the *SS* system and system operated under the *MEBS* policy. This relationship does not necessarily extend to having identical costs because the holding costs at a single (nonretailer) *SS* installation would be distributed over two corresponding *SF* installations. Thus, the equivalence between *SF* and *MEBS* (and the local information results) extends to *SS*, via its longer equivalent *SF* channel.

3. Performance of MEBS

Because we limited our attention to *MEBS* policies, even though they are not optimal for more than two stages, in this section we investigate how well *MEBS* performs. In Section 3.1 we numerically compare *MEBS* with the optimal policy for three- and four-stage systems. Following this, in Section 3.2 we establish lower bounds on the optimal policy and upper bounds on *MEBS* and discuss the gap between them. Both of these sections illustrate that *MEBS* performs very well, despite not being optimal.

3.1. Numerical Performance of MEBS

We numerically evaluate systems consisting of three and four installations with capacities $K=10$, across a wide set of parameters. These parameter sets were chosen to consider both typical and extreme values. While keeping the most constraining capacity constant, the demand distribution is varied from low to high utilization and from low to high coefficient of variation. The demand distributions have the following characteristics: $\mu := \mathrm{E}[D] \in \{6, 9, 9.8\}$ and c.v.$\in \{0.2, 0.5, 1.0, 1.4\}$. To preclude nontrivial cases in this capacity-limited context, all demand distributions have support over nonnegative integers with $\Pr(D > K) > 0$. The unit shortage cost, $p \in \{1, 5, 20, 50\}$ that were chosen relative to a constant $H_1 = 1$, in keeping with practical contexts where stockout costs usually exceed holding costs. The sets of holding costs consider various cases: constant increments, both smaller and larger; and uneven increments. However, these were necessarily different across the three and four installation channels. For $N=3$, $(H_1, H_2, H_3) \in \{(1, 0.99, 0.98), (1, 0.95, 0.90), (1, 0.75, 0.50), (1, 0.99, 0.80), (1, 0.81, 0.80)\}$. For $N=4$, $(H_1, H_2, H_3, H_4) \in \{(1, 0.99, 0.98, 0.97), (1, 0.95, 0.90, 0.85), (1, 0.75, 0.50, 0.25), (1, 0.99, 0.98, 0.80), (1, 0.82, 0.81, 0.80)\}$. Finally, the discount factor is $\alpha = 0.9$. These parameter combinations result in 240 separate cases for each channel length.

For the numerical exercise in this section, we consider the optimality actions for any initial inventory position within a band $\mathcal{B} := \{Y \mid Y^i \le Y^{i+1} \le Y^i + K, i = 1, \ldots, N-1\}$ and below the target levels. This is partly to factor in that the problem is exercised using backward induction over a finite horizon (described). Toward the end of the horizon there may be nonstationary behavior but as the horizon lengthens, the behavior at the start of the horizon (which has converged in value) will be stationary, consistent with our analytical models.

The solution procedure is the following. Beginning at the final period with a zero terminal value, for each initial state inventory position in the band $\mathcal{B}$ (justified by lemma 1 in Parker and Kapuściński 2004) the optimal actions, constrained by the action set, are identified along with the corresponding optimal value. The value function is iterated until the difference between the value function in successive periods for every state value is less than 0.005 (an arbitrary value).

Separately, in every period we implement a *MEBS* policy as defined in the following algorithm:

1. For $t = 0$, set $J_t^{MEBS}(\boldsymbol{Y}) = L(\boldsymbol{Y})$
2. Set $\boldsymbol{Z} := \arg\min_{\boldsymbol{Y}\in\mathcal{B}} J_t^{MEBS}(\boldsymbol{Y})$.
3. Let $V_t^{MEBS}(\boldsymbol{X}) = T(\boldsymbol{Z})J_t^{MEBS}(\boldsymbol{Y})$.
4. Let $J_{t+1}^{MEBS}(\boldsymbol{Y}) = L(\boldsymbol{Y}) + \alpha\mathrm{E}[V_t^{MEBS}(\boldsymbol{Y} - \boldsymbol{D})]$.
5. Go to Step 2.

where the $T(\boldsymbol{Z})$ operator maps $J_t^{MEBS}(\boldsymbol{Y})$ to $V_t^{MEBS}(\boldsymbol{X})$ using the *MEBS* policy parameterized by $\boldsymbol{Z}$.

For each state value (the vector of starting inventories), the error of the *MEBS* policy relative to the optimal value function is calculated. Summary error statistics are the average error across stages, up to K below Z (bigger differences lead to smaller errors), the maximum error found across the entire band (across different initial inventory states), the error at Z, and the error at the optimal target level. Summary statistics can be found in Table 2 for $N = 3$ and $N = 4$. The first observation we see is the large number of cases (out of 240) below 0.1%, 0.5%, and 1%, across the various parameters described, suggesting that while *MEBS* is not optimal, its errors tend to be very low whether considering average errors around $\boldsymbol{Z}$ or the maximum error in the band. Secondly, we observe that errors are larger for the four-stage system than the three-stage system, although the errors are still very small. Across the 240 cases for $N = 3$, the average of the average errors is 0.10% and the average of the maximum errors is 0.23%. The corresponding numbers for $N = 4$ are 0.22% and 0.44%. In contrast, within their testbed of 72 problems Glasserman and Tayur (1996) find an average error of 1.9% using modified base-stock policies (with inventory visibility) at each stage, over those found using IPA in Glasserman and Tayur (1995). Speck and van der Wal (1991) are credited with constructing a counter example to base-stock policies, under somewhat contrived circumstances, but also illustrate that neighboring base-stock policies are within 1.7% of optimality.

To further explore the efficacy of *MEBS* we also considered another metric where the "pipeline" cost $(H_2 + H_3 + H_4)\mathrm{E}[D]/(1 - \alpha)$ (which should be common to the optimal and *MEBS* policies) is removed and the error recalculated.[6] The "absolute" and "pipeline-adjusted" (PA) errors are summarized in the histograms in Figure 1. Clearly the PA errors are larger (as expected) but still perform well. Also, the larger PA errors are for the same cases, which yield larger absolute errors, as we discuss.

Figure 2 illustrates how the *MEBS* errors behave with respect to demand characteristics across the three- and four-stage systems. These data represent an average across the 20 cases representing a mean-c.v. combination. We first note (as already observed), the average and maximum errors for $N = 4$ are worse than for $N = 3$. Second, we see that as the c.v. increases, the errors increase. Third, we note that as the mean increases, the errors decrease. Last, we see that these errors are very small everywhere but they are worst at the low demand with high variance combination. All these observations remain true for the PA errors, too. The insight from this is that *MEBS* performs well precisely where we would want it to, where the demands and capacities will be colliding. In contrast, the larger errors (absolute or PA, average or maximum) occur for the cases where the mean demands are lower ($\mathrm{E}[D] = 6$ c.f. $K = 10$), where a conventional echelon base-stock policy may be more apt. Larger errors were also observed for cases when $p = 1$ (c.f. $H_1 = 1$), which is the least realistic/practical value of the unit stockout cost.

Although these errors are quite small, there are several reasons why these errors are conservative (that is, biased against *MEBS*):

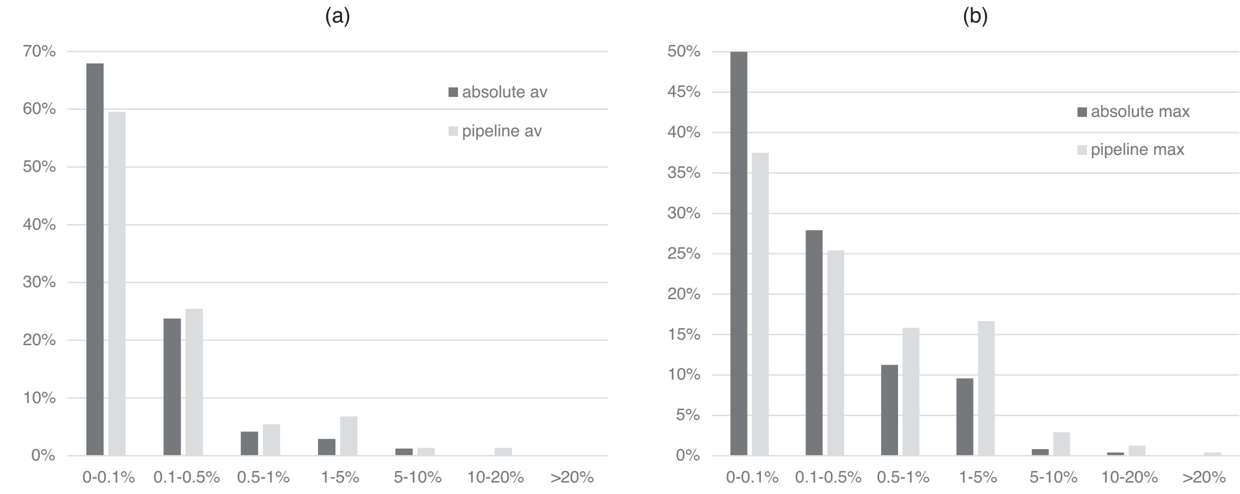
- The error metrics are conservative. The maximum error is found for a state value within the band but may not be in any typical sample path. The errors should be measured across the more typical transient states under each policy to better reflect the typical errors experienced.
- We used a discount factor of $\alpha = 0.9$. Some numerical experiments with higher discount factors have shown even smaller errors (average and maximum) arise.
- The holding cost parameters exercised are those that would disadvantage against *MEBS*-like behavior.
- The parameter sets chosen include extreme cases where *MEBS* policies would not be favored.

We attempted to consider longer chains ($N > 4$). However, even $N = 5$ appeared to be prohibitively time consuming while the $N = 4$ case was possible.[7]

Table 2. Count of Cases from 240 for $N = 3$ and $N = 4$

	$N = 3$		$N = 4$	
	Average error	Maximum error	Average error	Maximum error
Below 0.1%	204	150	163	120
Below 0.5%	229	178	220	187
Below 1%	233	230	230	214

Figure 1. Average and Maximum Error Histograms for Absolute and Pipeline-Adjusted Errors



Notes. (a) Average errors histogram. (b) Maximum errors histogram.

3.2. Bounds

In this section we define several bounds on the optimal value and the value when using *MEBS*. Ideally, bounds are used to establish an analytical performance gap, by establishing an upper bound on a proposed suboptimal policy (*MEBS*) and a lower bound on the optimal policy. Our proposed bounds rely on the steady state shortfall distribution for a single-stage capacity-limited system (see Tayur 1993, Glasserman and Tayur 1994, Glasserman 1997, Huh et al. 2010, Huh et al. 2016 for various treatments of this topic). We denote this shortfall as S with distribution G. G may be found through the following Lindley equation recursion: $S_{n+1} = \max(0, S_n + D_n - K)$, $S_0 = 0$, and $S =_d \lim_{n\to\infty} S_n$. The most important aspect to note is that there is no closed-form expression for the cost of the single-stage capacity-limited system. As noted, $S \sim G, D \sim F$ and $\mu := \mathrm{E}[D]$.

We define one lower bound on the optimal cost:

LB1 This bound includes the optimal cost for a single installation with capacity limit with a single-period leadtime with access to an unlimited supplier. The higher installations are assessed the holding costs for the average demand. As such, this encapsulates the costs of the original model but with fewer constraints, because it ignores any inventory availability limits.[8] The bound has the cost:

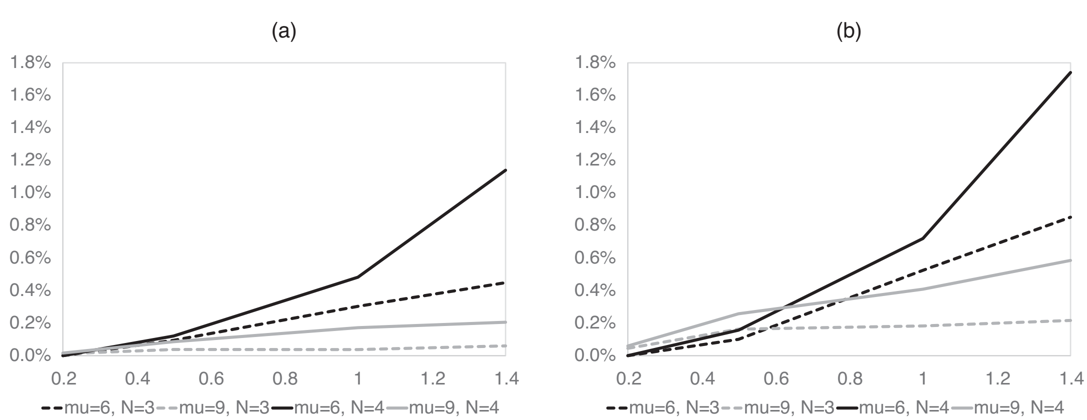
$$\frac{1}{1-\alpha}\left[H_1 \mathrm{E}[(z-D-S)^+] + p\mathrm{E}[(D+S-z)^+] + \sum_{j=2}^{N} H_j \mu\right]$$

where the expectation is taken over $S + D$ and z solves $G * F(z) = p/(p + H_1)$.

We define three upper bounds on *MEBS* cost:

UB1 This bound is for installation 1 to be modeled as a single-installation capacity-limited system with a

Figure 2. Average of Average Errors and Maximum Errors as a Function of Coefficient of Variation



Notes. (a) Average error as a function of coefficient of variation. (b) Maximum error as a function of coefficient of variation.

one-period leadtime. Assume that installations $2, \ldots, N$ stock K units, which is the maximum needed under any policy with capacity of K at the lowest stage. In this way, installation 1 will never be limited by installation 2's inventory availability and will behave as ordering from unconstrained inventory, so when optimizing the inventory there, it will follow Federgruen and Zipkin's (1986b) modified base-stock policy for a single-stage system. Thus, the cost consists of holding K units in all installations other than installation 1 and holding and backlogging cost at installation 1:

$$\frac{1}{1-\alpha}\left[H_1 \mathrm{E}[(z-D-S)^+] + p\mathrm{E}[(D+S-z)^+] + K\sum_{j=2}^{N}(j-1)h_j\right]$$

where the expectation is taken over $S+D$ and z solves $G * F(z) = p/(p+H_1)$.

UB2 This bound is established by considering a single-stage capacity-limited system with a leadtime equal to $N-1$, the number of stages (other than installation 1). The holding cost applied will be a period-specific unit holding cost for each unit of inventory as it traverses toward installation 1, corresponding to the holding costs of the system described, appropriately discounted and summed and applied to the single installation. The rationale for why this is an upper bound to *MEBS* is that the economics are identical but there are fewer decisions to be made to control the system but with a common installation 1 decision (modified base-stock). The cost is:

$$\frac{1}{1-\alpha}\left[H_1 \mathrm{E}[(z-S-D)^+] + p\mathrm{E}[(S+D-z)^+] + \mu\sum_{j=2}^{N}H_j\right].$$

where the expectation is taken over $S+D$ and z is the target level determined by the critical fractile and the convolution of the shortfall and the $(N-1)$-fold convolution of demand: $G * F^{N-1}(z) = p/(p+H_1)$.

UB3 For this bound each installation orders μ and installations $2, \ldots, N$ stock μ. Unlike the other bounds, this bound does not follow a base-stock type policy and thus does not have renewal properties. Notice that because $a_t^1 = \mu < K \le \min_j K_j$, the effect of the most limiting capacity does not have any effect. The costs of this bound is as follows:

$$\sum_{t=1}^{\infty}\alpha^{t-1}\left\{H_1 \mathrm{E}[(X_t^1 + \mu - D_t)^+] + p\mathrm{E}[(D_t - X_t^1 - \mu)^+] + \sum_{j=2}^{N}H_j\mu\right\}$$

where $X_{t+1}^1 = X_t^1 + \mu - D_t$. The initial inventories at all higher installations will be μ but the retailer's initial inventory will need to be determined.

We observe the following based on numerical experiments:

- LB1 becomes tighter for lower demand coefficients of variation.
- Both UB1 and UB2 are tighter than UB3.
- UB1 tends to be tighter than UB2 for higher mean demand distributions but this diminishes as the demand coefficient of variation increases.
- The gap between UB1 and LB1 narrows as the demand mean increases.

The last bullet is of particular interest: LB1 and UB1 have similar forms, which allow us to evaluate the difference between them. The difference between bounds UB1 and LB1 yields $\sum_{j=2}^{N} H_j(K-\mu)/(1-\alpha)$ and allows us to bound the difference between *MEBS* and the optimal policy but also to provide additional explanation for why *MEBS* perform so well. Evidently, as the mean demand increases relative to the capacity, we should expect the cost of operating the system to likewise increase, so this narrowing difference between UB1 and LB1 has real meaning. If taken as a relative measure of LB1, the measure becomes

$$\Delta := \frac{\text{UB1-LB1}}{\text{LB1}} = \frac{\sum_{j=2}^{N} H_j(K-\mu)}{H_1 \mathrm{E}[(z-D-S)^+] + p\mathrm{E}[(D+S-z)^+] + \sum_{j=2}^{N} H_j\mu} \tag{16}$$

and while it is not as compact, it is more meaningful. Specifically, as the mean demand increases, the numerator decreases and the denominator increases, so Δ decreases. Now consider increasing the coefficient of variation of demand: Keeping the mean constant but increasing the standard deviation will lessen Δ. Also, Δ decreases in p. These statements are formalized in the following result.

Proposition 5. *Δ is nonincreasing in the demand mean, the demand standard deviation, the unit shortage cost, installation 1's holding cost, and nondecreasing in H_j for $j > 1$ and N.*

Table 3 (Panel A) provides numerical illustrations. It shows that Δ is decreasing in the demand mean and coefficient of variation. These numbers show Δ is less than 1% for $\mu = 9.8$, which is important because this is a situation where the capacity limit has real effect. For $\mu = 6$ the values of Δ are far larger, greater than 18%. However, Table 3 (Panel B) shows the average errors of *MEBS* compared with the optimal values for $N = 3$. Each cell represents the average of the average errors of the 20 cases for that mean-c.v. combination. As these numbers for $\mu = 6$ show, they are far smaller in scale than Δ indicating that the real errors are far smaller. The averages for the maximum errors are similarly small. Interestingly, the values of the bound (Δ) are lowest at high coefficient of variability (actually very low). But the actual differences, when we travel from a high coefficient of variability to a low one, are becoming even lower. As with the results for

Table 3. Comparison for Demand Means and Coefficients of Variation

		Panel A: Value of Δ			
		Coefficient of variation			
		0.2	0.5	1.0	1.4
μ	9.8	0.95%	0.3%	0.12%	0.12%
	9	8.72%	4.78%	1.96%	1.12%
	6	55.4%	46%	29.7%	18.5%
		Panel B: Average of average errors			
		Coefficient of variation			
		0.2	0.5	1.0	1.4
μ	9.8	0.06%	0.11%	0.08%	0.20%
	9	0.02%	0.09%	0.17%	0.21%
	6	0.00%	0.12%	0.48%	1.14%

Section 3.1, the table also indicates that for higher demand means, *MEBS* performs superbly, but it may perform very well across all demand characteristics.

When the whole system is not heavily utilized, the echelon-based policy is naturally optimal or very close to optimal (because an echelon-based policy is optimal for uncapacitated systems). When the system reaches very high utilizations, we observe that there is very little room for cost differences for holding inventory at the higher installations (as μ increases, Δ decreases). This is driven by the band structure of *MEBS* policies, which limits inventory to capacity K per installation and the natural need to maintain flow at all installations, which requires average inventory to be above average demand. The strong performance of *MEBS* combined with its sufficient simplicity and interpretability make *MEBS* effectively an obvious policy from a practical point of view. In this context, the question of running *MEBS* with local information is an important one.

4. Extensions to Other Capacity Configurations

In this section, we discuss how the properties we show above extend to other capacity configurations. In Section 4.1, we examine a system comprised of two distinct capacities with the smaller capacity internal to the channel. We call this a dual-band system (or 2-band system) and demonstrate that the local information is sufficient for shipments to represent the orders of a *MEBS*-like policy, just as found for the single-band system in Section 2. In Section 4.2, we describe how this principle can be extended to serial systems with any capacity configurations.

4.1. Serial Systems with Dual-Bands

In this section, we consider a system with the tightest capacity internal to the serial channel. We continue with Assumptions 2–4. For systems operating with full or local information, the following assumption describes the capacity configuration, replacing Assumption 1.

Assumption 5. *Each installation $i \in \{1,\dots,N\}$ operates under capacity limit K_i. Installation n has the smallest capacity, $K_n < K_1$ and $1 < n < N$. In addition, $K_n \le K_i$ for $i = n+1,\dots,N$ and $K_1 \le K_i$ for $i = 2,\dots,n-1$.*

This assumption allows the system to be effectively divided into two segments, each operating as *MEBS*-type systems. With such a structure of capacities a "dual-band" property would hold: it cannot be optimal to store more inventory than K_1 at any installation $j = 2,\dots,n-1$ or store more than K_n at any installation $j = n+1,\dots,N$. This is formalized in Lemma 9. It is, therefore, needed to use a modification of *MEBS* to "dual-band *MEBS*" (2-*MEBS*, defined), that limits inventory to the dual-band. Let $B1 = \{2,\dots,n-1\}$ and $B2 = \{n+1,\dots,N\}$.

Lemma 9 (Generalization of lemma 1 in Parker and Kapuściński (2004)). *Assume $K_1 \le K_j$ for $j \in B1$, $K_n < K_1$, and $K_n \le K_j$ for $j \in B2$. For any X_t all optimal Y_t satisfy $y_t^j \le \max(K_1, x_t^j - a_t^{j-1})$ for $j \in B1$, and $y_t^j \le \max(K_n, x_t^j - a_t^{j-1})$ for $j \in B2$.*

Proof See Online Appendix A.

Corollary 2 (Generalization of corollary 2 in Parker and Kapuściński (2004)). *Assume that $K_1 \le K_j$ and $X_t^j - X_t^{j-1} \le K_1$ for all $j \in B1$ and $X_t^j - X_t^{j-1} \le K_n$ for all $j \in B2$. Then:*

a. *The optimal Y_n^j satisfy $Y_n^j - Y_n^{j-1} \le K_1$ for all $j \in B1$ and $Y_n^j - Y_n^{j-1} \le K_n$ for all $j \in B2$.*

b. *If the optimal policy is followed, then the inventory positions satisfy $X_{t'}^j - X_{t'}^{j-1} \le K_1$ for all $j \in B1$ and $X_{t'}^j - X_{t'}^{j-1} \le K_n$ for all $j \in B2$, for $t' < t$.*

c. *All capacities in $B1$ may be replaced with capacities equal to K_1 and capacities in $B2$ by K_n without affecting costs.*

Note that the results guarantee that the optimal policies are in dual-band. Therefore, for dual-band, we replicate single-band results. Recall that for limiting capacity at retailer, a policy in single-band is optimal. *MEBS* is intuitive and well-performing policy in single band, but not necessarily optimal. We showed that *MEBS* is equivalent to *SF*. For two limiting capacities operating in dual-band is optimal, but dual-*MEBS* policy itself is not necessarily optimal. We argue that *SF* in such a case is equivalent to 2-*MEBS* (defined below).

We continue the earlier notation and extend our notation to dual-band. The transition functions (Equations (1–5)) from Section 2 continue to hold. We will presume the *SF* timing where each installation (other than the retailer) orders after receiving their immediate customer's order. The order quantities in this dual-band system under *SF* are as follows:

$$\vec{a}_t^1 = \left[\vec{z}^1 - \left(\vec{x}_t^1 + \vec{B}_t^{12}\right)\right] \wedge K_1 \tag{17}$$

$$\vec{a}_t^i = \vec{z}^i - \left[\left(\vec{x}_t^i + \vec{B}_t^{i,i+1}\right) - \left(\vec{q}_t^{i-1,i} + \vec{B}_{t+1}^{i-1,i}\right)\right] \tag{18}$$

$$= \vec{z}^i - \left[\vec{x}_t^i - \left(\vec{a}_t^{i-1} + \vec{B}_t^{i-1,i}\right) + \vec{B}_t^{i,i+1}\right] \tag{19}$$

$$\text{for } i = 2, \ldots, n-1, n+1, \ldots, N-1$$

$$\vec{a}_t^n = \left[\vec{z}^n - \left[\vec{x}_t^n - \left(\vec{a}_t^{n-1} + \vec{B}_t^{n-1,n}\right) + \vec{B}_t^{n,n+1}\right]\right] \wedge K_n \tag{20}$$

$$\vec{a}_t^N = \vec{z}^N - \vec{x}_t^N + \vec{B}_t^{N-1,N} + \vec{a}_t^{N-1} \tag{21}$$

As before we do not explicitly impose the capacity constraints on the order quantities at any installations other than the retailer and installation n. Also, we do not constrain the incoming shipment quantities at installations 1 and n, they will naturally never exceed the capacity levels. The order quantity equations may be interpreted as before.

For the purposes of clarity, let us define the following full-information inventory policy.

Definition 4. *The dual-band modified echelon base-stock policy, or 2-MEBS, can be written as the mapping from initial echelon inventories (X^j) to echelon ordering levels (Y^j) where $Y^j \geq X^j$ for all j, attempting to reach the target echelon levels (Z^j):*

$$Y^1 = \min(Z^1, X^1 + K_1, X^2)$$
$$Y^i = \min(Z^i, Y^{i-1} + K_1, X^{i+1}) \text{ for } i = 2, \ldots, n-1$$
$$Y^n = \min(Z^n, X^n + K_n, X^{n+1})$$
$$Y^i = \min(Z^i, Y^{i-1} + K_n, X^{i+1}) \text{ for } i = n+1, \ldots, N-1$$
$$Y^N = \min(Z^N, Y^{N-1} + K_n).$$

We call this policy 2-MEBS(K_1, K_n). When we want to explicitly show the dependence on targets Z, we will refer to the policy as 2-MEBS(Z, K_1, K_n).

We observe that *2-MEBS*(∞, ∞) is equivalent to *MEBS*(∞), which is equivalent to the echelon base-stock policy of Clark and Scarf (1960) (i.e., no capacity constraints). The critical factors that differentiate *2-MEBS*(K_1, K_n) from *SF* are as follows:

a. Similarly to *MEBS*, *2-MEBS* is a full information policy and the *2-MEBS* policy does not impose a capacity constraint at any installation but instead imposes that each installation will never *stock* more than K_1, $Y^i \leq Y^{i-1} + K_1$ for $i \in \{2, \ldots, n-1\}$ and the higher installations would not stock more than K_n, $Y^i \leq Y^{i-1} + K_n$ for $i \in \{n+1, \ldots, N\}$, thus introducing two "band" limitations.

b. *SF* is a local information policy and imposes capacity constraints upon the constraining installations (only) $y^1 - x^1 \leq K_1$ and $y^n - x^n \leq K_n$.

Before we show the formal results, let us consider an example to illustrate how demand information is conveyed through the tightest capacity, internal to the channel. Consider the example in Table 4. In this four-stage system, $K_1 = 10, K_2 \geq 10, K_3 = 7,$ and $K_4 \geq 7$. In the upper panel, the *2-MEBS* is exercised over seven periods and in the lower panel, *SF* is exercised. In period 1, each system begins with inventory at the targets ($Z^1 = 15, Z^2 = 25, Z^3 = 36, Z^4 = 43$) and with no backlogs in the *SF* system; these are the requirements of Condition 1. The inventory levels at the start of each period are identical between *2-MEBS* and *SF*, expressed as echelon and installation inventories, respectively. Likewise you can see the order amounts in *2-MEBS* (a_t^i) equal the shipment amounts in *SF* ($q^{i,i+1}$), which will be proven below. A demand of 30 in the first period ($d_1 = 30$) is far larger than either limiting capacity ($K_1 = 10$ and $K_3 = 7$) can handle in one period. The effect of this is that the retailer will censor her orders by K_1 and carry the excess as a backlog. The censored orders are conveyed up the channel (see $a_2^1 = 10$ and $a_2^2 = 10$) to installation 3, but they exceed the limiting capacity $K_3 = 7$. Under *SF*, once the inventory at installation 3 is drawn down faster than it can be replenished, there begins to be a local backlog at installation 3, $B_4^{23} = 2$. Because the mean demand over this time duration is lower than the tightest demand ($30/7 = 4.3 < K_3 = 7$), the system will get back to its initial condition (Condition 1). Although there has been a second censoring of the original demand process, sufficient information is conveyed up the channel and the local backlog at installation 3 ensures that the excess demand is cumulatively retained until it can be eroded through sufficient supply. Imposing a *2-MEBS* policy on this system, the "optimal" target

Table 4. Comparison of 2-*MEBS* and *SF*

2-*MEBS*

t	d_t	X_t^1	Y_t^1	a_t^1	X_t^2	Y_t^2	a_t^2	X_t^3	Y_t^3	a_t^3	X_t^4	Y_t^4	a_t^4
1	30	15	15	0	25	25	0	36	36	0	43	43	0
2	0	−15	−5	10	−5	5	10	6	13	7	13	20	7
3	0	−5	5	10	5	13	8	13	20	7	20	27	7
4	0	5	13	8	13	20	7	20	27	7	27	34	7
5	0	13	15	2	20	25	5	27	34	7	34	41	7
6	0	15	15	0	25	25	0	34	36	2	41	43	2
7	0	15	15	0	25	25	0	36	36	0	43	43	0

SF

t	d_t	x_t^1	q_t^{12}	x_t^2	B_t^{12}	q_t^{23}	x_t^3	B_t^{23}	q_t^{34}	x_t^4	B_t^{34}	a_t^4
1	30	15	0	10	0	0	11	0	0	7	0	0
2	0	−15	10	10	0	10	11	0	7	7	0	7
3	0	−5	10	10	0	8	8	0	7	7	0	7
4	0	5	8	8	0	7	7	2	7	7	0	7
5	0	13	2	7	2	5	7	5	7	7	0	7
6	0	15	0	10	0	0	9	0	2	7	0	2
7	0	15	0	10	0	0	11	0	0	7	0	0

levels[9] at installations 1 and 3 will naturally be higher to compensate for the limited upstream supply.

Below we expand the definition of $d(K_1)$ to $d(K_1, K_n)$ and $Z(K_1)$ to $Z(K_1, K_n)$

Definition 5. *Let d_t be the periodic demand, with $D_t = \sum_{i=1}^{t} d_i$.*

a. *Let a_t^{n-1} be the order from installation $n-1$, with $A_t^{n-1} = \sum_{i=1}^{t} a_i^{n-1}$. We define $a_t^n(K_n)$ inductionally, as $a_t^n(K_n) = (A_t^{n-1} - A_{t-1}^n(K_n)) \wedge K_n$, where $A_t^n(K_n) = \sum_{i=1}^{t} a_i^n(K_n)$.*

b. *Let $d(K_1, K_n)$ denote the application of $d_t(K_1)$ and $a_t^n(K_n)$.*

c. *Let $z^1(K) = Z^1 = Z^1(K)$, $Z^n(K) = Z^n$, $z^n(K) = Z^n - Z^{n-1}(K)$, and $z^i(K) = (Z^i - Z^{i-1}(K)) \wedge K$ where $Z^i(K) = \sum_{j=1}^{i} z^j(K)$ and $K = K_1$ for $i = 2, \ldots, n-1$ and $K = K_n$ for $i = n+1, \ldots, N$.*

Given Lemma 9, without loss of generality, we replace Z with $Z(K_1, K_n)$.

Lemma 10. *For any demand d and $X_1^i \le Z^i(K)$ where $K = K_1$ for $i < n$ and $K = K_n$ for $i > n$, 2-MEBS$(Z, K_1, K_n) \oplus d \equiv$ 2-MEBS$(Z(K), K_1, K_n) \oplus d$.*

The primary result for the serial system with two limiting capacities follows.

Theorem 3. *$SF(K_1, K_n) \oplus d \equiv$ 2-MEBS$(K_1, K_n) \oplus d$.*

This equivalency result follows from the following relationships: $SF(K_1, K_n) \oplus d \stackrel{xR}{\equiv} SF(\infty, \infty) \oplus d(K_1, K_n) \equiv$ 2-$MEBS(\infty, K_n) \oplus d(K_1) \equiv 2 - MEBS(K_1, K_n) \oplus d(K_1) \stackrel{xR}{\equiv}$ 2-$MEBS(K_1, K_n) \oplus d$. We will not onerously duplicate the mirroring results from Section 2, but will prove results that illustrate the distinctions from the single-band system. We will demonstrate results for the first two equivalencies and borrow earlier results for the remaining ones. Proofs appear in Online Appendix A.

Lemma 11. *$SF(K_1, K_n) \oplus d \stackrel{xR}{\equiv} SF(\infty, \infty) \oplus d(K_1, K_n)$.*

Lemma 12. *$SF(\infty, \infty) \oplus d(K_1, K_n) \equiv$ 2-MEBS$(\infty, K_n) \oplus d(K_1)$.*

Lemma 13. 2-*MEBS*$(\infty, K_n) \oplus d(K_1) \equiv$ 2-*MEBS*$(K_1, K_n) \oplus d(K_1)$.

Lemma 14. *2-MEBS$(K_1, K_n) \oplus d(K_1) \stackrel{xR}{\equiv}$ 2-MEBS$(K_1, K_n) \oplus d$.*

Through combining Lemmas 11–14, we achieve Theorem 3. In short, the orders in *SF* result remain sufficient to result in shipments identical to a *MEBS*-like policy. As with the single-band result, the local backlogs capture each installation's (cumulative) inability to immediately satisfy their immediate customer's needs. The difference in this dual-band system is that the orders from installation $n-1$ get censored by installation n's lower capacity (K_n) and thus, there is a cumulative carryover of installation $n-1$'s orders in excess of K_n, reflected in installation n's orders in $a_t^n(K_n)$.

This theorem carries the same significance as Theorem 1, that local information in $SF(K_1, K_n)$ is sufficient to generate shipments identical to 2-*MEBS*(K_1, K_n), the full information policy closest to *MEBS* for a dual-band system. As before, *SF* will naturally limit the size of the shipments without the imposition of shipment constraints.

Although we are unable to numerically test the system with two or more bands, it is easy to see that the same logic as for single band can be used to create LB1 and UB1 (from Section 3.2) for such systems. If $K_n < K_1$, then clearly $(K_1 - \mu)(n-1) + (K_n - \mu)(N - n + 1)$ may be not as tight as for a single band, but the intuition suggesting a good performance of *MEBS* remains. For any portion of the system, where capacity is not tight, we have a policy similar to the optimal policy for an uncapacitated system and the echelon base stock policy works well. For the portions of the system where capacity is tight, it is increasingly likely that the system will operate with full capacity upstream of the bottleneck, which follows the *MEBS* structure.

4.2. Serial System with *m* Bands

Section 4.1 illustrates that in a dual-band capacity-limited system, sufficient demand information is conveyed via the orders to match a dual-band *MEBS* policy. Consider an N-stage system where there are $m > 2$ constraining capacities, such that each successive upstream constraining capacity is lower than the downstream one. The capacities between the constraining capacities are at least as high as the downstream constraining capacity. An example system with $N = 12$ and $m = 3$ would be $(K_1, K_2, K_3, K_4, K_5, K_6, K_7, K_8, K_9, K_{10}, K_{11}, K_{12}) = (15, 20, 20, 20, 14, 20, 20, 20, 10, 20, 20, 20)$ where the constraining capacities are at installations 1, 5, and 9: $K_1 = 15$, $K_5 = 14$, and $K_9 = 10$. Using a 3-band *MEBS* policy is the most analogous policy to *MEBS* and a simple adaptation of Theorem 3 will again ensure the sufficiency of the orders. Extending Lemma 9 suggests $x^j \le K_1$ for $j = 2, 3, 4$, $x^j \le K_5$ for $j = 6, 7, 8$, $x^j \le K_9$ for $j = 10, 11, 12$. The condition that each successive upstream constraining capacity is lower than the downstream one is not limiting but simply illustrates the extent of each band in the channel.

Thus, in this manner any capacity configuration may be accommodated in a serial channel and an appropriate *m-MEBS* policy may be matched by using local information alone. We do not provide any formal proofs for m bands, as dual band provides a complete prototype for m bands. In summary, *MEBS*-type policies can be applied to serial systems with *any* capacity configurations with local information only.

5. Incentives in Decentralized Operations with Capacity Limits

In this section, we address the topic of incentives of locations when operated in a decentralized manner.

To put this question in a broader context, the results of Clark and Scarf (1960) were extended in, at least, three interesting dimensions: (a) decentralized ownership: Cachon and Zipkin (1999) shows that echelon structure holds in decentralized systems, although the targets may be different; (b) information sufficiency: Axsäter and Rosling (1993) show that centralized policy may be "operated" with local information only; and (c) incentives sufficiency: Lee and Whang (1999) show that it is possible to create a performance evaluation system where local managers (divisions) will have incentives to follow the centralized (headquarters) policy. All of these papers deal with uncapacitated systems. For capacitated systems the structure of the optimal policy (*MEBS*) was established for two stages and several papers showed the difficulty in expanding it further, including Janakiraman and Muckstadt (2009) and Huh et al. (2016). (a) The structure of the decentralized policy was shown again for two stages in Parker and Kapuściński (2011), and (b) information sufficiency is the topic of this paper in the previous sections. It is relevant to ask a third question of whether the results (c) incentives sufficiency of Lee and Whang (1999) can be extended to capacitated system. This is the topic of this section.

Lee and Whang (1999) (hereafter "LW99") design an incentive compatibility mechanism for the Clark and Scarf (1960) (hereafter "CS60") system. They suggest a good performance measurement scheme will be one that has:

1. *Cost Conservation*, where the accounting system can trace all costs to the individual sites, not requiring taxes or subsidies from headquarters in any period;

2. *Incentive Compatibility*, where the potential for incentive misalignment is eliminated and the mechanism allows each site to make decisions consistent with the optimal decisions; and

3. *Informational Decentralizability*, which means the decisions can be made with local information alone.

LW99 design a mechanism that satisfies (1) cost conservation and (2) is incentive compatible. Each location will order to their ideal stocking level that coincides with the optimal level for the system (incentive compatibility). The key ingredient of their mechanism is a payment from the supplier to the retailer when he is unable to provide sufficient inventory for the retailer to reach her desired stocking level. It is important to note that Veinott (1966) expresses the idea of CS60 in an elegant manner using stage decomposition, and Federgruen and Zipkin (1984) provides an algorithm to compute up to levels for the infinite-horizon case. LW99's mechanism design is achieved by adopting the induced penalty cost functions of Veinott (1966) and Federgruen and Zipkin (1984), which result from the decomposition of the system's value function.

The underlying intuition of LW99 is relatively simple and follows the structure of CS60. The retailer's and supplier's ordering decisions are actually influencing echelon inventory rather than stage inventory. That is, the supplier's order brings inventory into echelon 2 (consisting of stages 1 and 2), and the retailer brings a portion of that inventory into echelon 1. CS60 effectively shows that the optimal echelon levels are the result of minimizing each echelon costs with some added induced-penalty function (the supplier having too little inventory is penalized for not allowing the retailer to freely increase the retailer's inventory). Of course in practice the supplier does not incur echelon costs. LW99 leverage these two facts by modifying the objective functions of both supplier and retailer and making the supplier incur echelon-2 costs (rather than his own costs) and the retailer incur echelon-1 costs. This is a portion of the transfer mechanism they define. They also reassign the costs across periods, so that the net transfers balance out. And finally LW99 rearrange the decomposed value function so that the induced penalty cost function becomes a *payment* from the supplier to the retailer.

Notably, LW99 consider the decentralized operation of a centrally owned system, referring to different stages as divisions. Also, as is evident from the functional form, the additional payments and credits between supplier and retailer, that reassign echelon-2 costs to the supplier (and the remainder to the retailer) are functions of a given period's inventory levels. To execute these payments, current inventory levels need to be visible to both parties.

We follow the same logic as LW99 to adjust incentives for both supplier and retailer. Specifically, the reassignment of costs to echelons and across time is identical to LW99.

Transfer pricing: in order to remove the profit motive (and translate the problem into cost minimization) the retailer is charged at the marginal cost of the supplier, c_2.

Consignment: The supplier is responsible for the inventory cost at its site as well as at the retailer at the supplier's holding rate $H_2 = h_2$, while retailer is only responsible for the excess of holding costs, $h_1 = H_1 - H_2$ at its own site, thus aligning holding costs with echelon holding costs when the inventory at the retailer is nonnegative.

Additional backlog penalty: consistent with an echelon formulation, the retailer reimburses the supplier for negative inventory (backlog) at its site.[10]

Shortage reimbursement (or induced penalty function): this element matches the induced penalty in CS60, Veinott (1966), and Federgruen and Zipkin (1984). In all these papers, echelon 2 is penalized for not being able to satisfy orders from echelon 1, whereas echelon 1 ignores the possibility of not having enough inventory. LW99 leverage this dependency: transferring the penalty to echelon 1 is the same as reimbursing echelon 1 for any consequences of not getting enough inventory, making echelon 1 behave the same as an unconstrained system in these papers.

Clearly, consignment and additional backlog penalty require visibility of current-period inventory, while shortage reimbursement may be implemented without observability of inventory, simply based on orders placed by echelon 1 not delivered by echelon 2.[11]

We use the first three components and modify the fourth one. Thus, we provide details here for the fourth element, shortage reimbursement. Parker and Kapuściński (2004) (hereafter "PK04") derives induced penalty cost functions for the centralized system: in the two-stage system, the supplier self-imposes a cost for having insufficient stock and the retailer self-imposes a cost when her capacity limits the supplier from reaching his desired stocking level. We rearrange the decomposed value functions from PK04 so that there will be potentially penalties in both directions and then, like LW99, we change them into actual payments.

Specifically, over the finite-time horizon the value function for the system can be decomposed into two value functions dependent on single-echelon variables, delivering two induced penalty functions, as follows:

Setting $V_0(\cdot,\cdot)=0$, we have for $(X^1,X^2)\in\Re^2$

$$V_n(X^1,X^2)=\min_{(Y^1,Y^2)\in\mathcal{A}(X^1,X^2)}\{f_n^1(Y^1)+f_n^2(Y^2)\}$$

where $f_n^i(Y^i)=L_i(Y^i)+\alpha \mathrm{E}V_{n-1}^i(Y^i-D)$ and

$$\mathcal{A}(X^1,X^2)=\{(Y^1,Y^2)\subset\Re^2 \mid X^1\le Y^1\le X^2\le Y^2 \le X^2+K, Y^1\le X^1+K\}.$$

Assuming $X^2-X^1\le K$, based on the results in PK04, we can rephrase as $\mathcal{A}(X^1,X^2)=\{(Y^1,Y^2)\subset\Re^2 \mid X^1\le Y^1\le X^2\le Y^2\le Y^1+K\}$. From Karush's lemma we have:

$$\min_{X^2\le Y^2\le Y^1+K} f_n^2(Y^2)=f_n^{2L}(X^2)+f_n^{2U}(Y^1+K).$$

Defining $\tilde{f}_n^1(Y^1)=f_n^1(Y^1)+f_n^{2U}(Y^1+K)$, Karush's lemma gives us:

$$\min_{X^1\le Y^1\le X^2}\tilde{f}_n^1(Y^1)=\tilde{f}_n^{1L}(X^1)+\tilde{f}_n^{1U}(X^2),$$

where function $\tilde{f}_n^{1U}(X^2)$ is due to limited availability of inventory in echelon 2 to be ordered by echelon 1. This is the exact analog of induced penalty function in CS60, which was used by LW99 as a payment. Function $f_n^{2U}(Y^1+K)$ reflects limiting echelon 2 in how much it can order (not to exceed Y^1+K) and is due to band structure of capacitated problem.

Operationalizing these penalties translates into redefining the value functions for both retailer and supplier and using Karush's lemma for the redefined functions. Although the details are shown in Online Appendix B, the final outcome is that the payment from the supplier to the retailer is as follows:

$$g_n^{1U}(X^2)=\begin{cases}0 & \text{if } X^2\ge S_n^{1*}\\ \tilde{f}_n^1(X^2)-\tilde{f}_n^1(S_n^{1*}) & \text{if } X^2< S_n^{1*}\end{cases}$$

and the payment from the retailer to the supplier is as follows:

$$g_n^{2U}(Y^2\mid Y^1)=\begin{cases}0 & \text{if } Y^1+K\ge S_n^{2*}\\ f_n^2(\max(Y^1+K,Y^2))-f_n^2(S_n^{2*}) & \text{if } Y^1+K< S_n^{2*}\end{cases}$$

We note that there are many forms of penalty functions g_n^{2U} that achieve the condition required for matching decomposed value function at the equilibrium. Including the payments g_n^{1U} and g_n^{2U} into the site value functions endows the system with *cost conservation* and *incentive compatibility*, as per LW99. Complete details of the payment functions derived from these induced penalty functions are also included in Online Appendix B.

Although our approach is the same as in LW99, we note that for unlimited capacity the closed-form expressions of the induced penalty cost functions for the infinite-time horizon are known due to Federgruen and Zipkin (1984), as the future expected economic repercussions of the supplier's insufficient inventory upon the retailer can be brought back to the current period precisely. However, in the capacity limited case, due to the unknown number of time periods over which the system needs to catch up, closed-form expressions are not available.

5.1. Informational Requirements

Recalling the third attribute of a good performance measure scheme, *informational decentralizability*, as we noticed above, the value functions in both LW99 and our modification of LW99 include payments that are dependent on *both* echelon inventory levels. Thus, these schemes do not satisfy the attribute of using only local information for operational purposes.

It is important to recognize that the information decentralizability features of Axsäter and Rosling (1993) and in Section 2 use a *given* set of target inventory levels rather than deriving them. In the context of a headquarters wishing the local sites to operate independently, these target levels could be calculated by headquarters and *given* to the sites with the instructions to follow a policy with these parameters. This provides the opportunity to map out three categories of information:

1. economic parameters (e.g., holding and stockout costs), demand distribution, capacities, system length;
2. policy parameters; and
3. inventory levels and demand realizations.

Typically, the information in category 1 is fixed for a given problem and the information in category 2 can be derived from that in 1. It is possible that individual sites could learn of these static items in 1 over time. The information in category 3, however, is quite dynamic and changes from period to period. Unless intended and agreed on, it would be very difficult for an individual site to be aware of another sites periodic inventory levels. Axsäter and Rosling (1993) and we (in Section 2) use *only* the items in categories 2 and 3 to operate the system. More specifically, each installation only uses their site-specific information to operate according to the given policy: site $i > 1$ uses installation inventory target z^i, their local backlog $B_t^{i,i+1}$, their immediate downstream obligation $B_t^{i-1,i}$, and the immediate downstream order a_t^{i-1} (under *SF*). All of this information would be at hand for a given site and could be easily tracked. In contrast, when implementing the incentive compatible scheme described in LW99 and in this section, the site value functions are dependent on *all the information* in categories 1 and 3 to calculate the items in category 2 and operate the system. Thus, the incentive compatible mechanism and the local operation mechanism have quite different informational requirements.

6. Discussion and Interpretation

The issue of inventory visibility throughout the supply chain is particularly important when considering the practicality of implementing a theoretically optimal policy in reality. Echelon inventory policies are demonstrably optimal for some serial systems (e.g., Clark and Scarf 1960; Parker and Kapuściński 2004) and have the desirable property of ensuring a given quantity of inventory is available to the market within a prescribed number of periods (dictated by the number of downstream installations and their leadtimes). The work of Axsäter and Rosling (1993) illustrate that the echelon policies may be replicated with policies using local information alone for serial systems with no capacity limits. Because, in practical settings, most systems have capacity constraints, the question is whether information sufficiency holds in such systems. We demonstrate that local information is sufficient for serial systems with capacity limits, but with a caveat of using modified echelon base-stock (*MEBS*) policies. *MEBS* have been previously shown to be the optimal or equilibrium policies for some centralized or decentralized systems but under a limiting assumption of "common knowledge."[12] As far as informational requirements, the assumption of common knowledge would appear to be more limiting in a decentralized game where the independent firms would be unlikely to share operational information such as inventories, but is even likely to be a factor in integrated channels where local managers may jealously guard such knowledge from outsiders to the factory. Although ERP systems are intended to share information across locations within a single enterprise, in practice they may be prone to inaccurate reporting of data, delayed entry of data, incompatible systems across locations, legacy systems, all which indicate that an ERP system may not be a panacea for information sharing. *Ex-ante* it is not obvious that the market demand information is conveyed faithfully upstream (in the form of orders) in a channel with capacity limits, which serve to censor demands. In fact, that information is *not* conveyed fully. We show, however, that *sufficient* demand information is conveyed upstream so that the shipping decisions in the channel are replicated despite the installations having access to *local information only*.

We examined two decentralized mechanisms in the capacity-limited serial channel reflecting different timing: Sequential Fast (*SF*) and Sequential Slow (*SS*). The former reflects where the orders arriving from downstream are coming sufficiently fast that they can be incorporated into the current-period decisions whereas the latter is where the orders are arriving more slowly. We illustrate that the Sequential Fast timing regime directly relates to shipments encountered when operating a *MEBS* inventory policy centrally or not (Parker and Kapuściński 2004, 2011). Although *MEBS* is shown to be the optimal/equilibrium policy only for the two-echelon system only, it is a viable and attractive policy for longer chains, too, and our analysis is not limited to two echelons. We demonstrate the Sequential Slow timing corresponds to an installation-based policy, which is a reasonable extension of Federgruen and Zipkin's (1986a, b) modified base-stock policy to a serial system. Then, given a relationship between the desired target levels in *SF* and *SS*, we establish the strong relationship between the inventory levels in analogous systems. Specifically, we show an N-stage *SS* system may be replicated by a $(2N-1)$-stage *SF* system, illustrating that the slowness of the information conveyance upstream in *SS* is equivalent to a longer supply chain where information is conveyed quickly. This is similar to Hariharan and Zipkin's (1995) result where future knowledge of the customer's orders is a direct substitute for a replenishment leadtime, although our equivalence is exclusively on the supply side. Thus, both *SF* and *SS* (via its longer $(2N-1)$-stage *SF* system) have shipments equal to the orders under the *MEBS* policy. The key to this relationship is that under *SF* and *SS*, installations do not limit their order based on the available inventory from their immediate supplier (which is not known when the installation has local information only) but order their desired quantity and are willing to receive whatever the supplier can provide.

It is natural to question whether the same result may extend to serial systems operating under more general capacity configurations. Under a general capacity configuration, operating bands can be identified, defined as a group of neighboring installations where the lowermost installation's capacity is the (weakly) smallest. Under the operation of a corresponding *MEBS* policy, labelled as *m-MEBS*, sufficient demand information will be conveyed up the channel if accessing local information alone. The *m-MEBS* corresponds to *MEBS* because the ending local inventory at every installation above that most constraining installation within that band will be below that capacity level.

Our paper provides an additional justification that *MEBS* policies perform very well in practice and, arguably, they may be considered a "reasonable" class of policies for capacitated systems, especially because the *MEBS* policy can be run with local information only. The justification is based on both numerical experiments as well as formal bounds on the optimal policy and on *MEBS*. Note that *MEBS* policies do not store outrageously large inventories because they are aware of capacity limits downstream which would make the higher inventory useless. The formal difference between UB1 and LB1 further emphasizes that there is actually less room for wrong decisions, especially in systems with high utilizations, as long as *MEBS* policies are used.

We also highlight the difference in informational requirements between (a) operating a multistage system for a specific given *MEBS* policy versus (b) having incentives to maintain the optimal policy. In both uncapacitated and capacitated systems (a) can be achieved with only local information. On the other hand, the mechanism proposed in the literature for uncapacitated systems (LW99), as well as its adaptation to capacitated systems, requires full information about the state of the system.

The focus on the lack of knowledge of nonlocal inventory is deliberate. Inventory levels change so frequently from period to period that they are unlikely to be known to neighboring installations in the channel. Thus, merely by knowing her own inventory (and local backlogs), a local manager can mimic the ordering pattern of an inventory policy, which hitherto required global visibility of inventory levels, resulting in near optimal performance.

Acknowledgments

The authors thank the Area Editor, C.P. Teo, for his support throughout the review process. The authors are enormously grateful for the continuing support from the Associate Editor, and the thoughtful and constructive comments from two anonymous referees. Their suggestions considerably improved the paper.

Endnotes

[1] The assumed one period leadtimes above installation 2 are synonymous with the *MEBS* policy we analyze later.

[2] The *MEBS* policy is formally introduced in Definition 2.

[3] This is different from the usual multiechelon analysis with full information where the available upstream inventory typically limits orders.

[4] This is to reflect the fact that customers likely order products throughout the whole period.

[5] Similarly to Rosling (1989), if the initial inventory is above the target level, it will be brought below those target levels after a few initial periods.

[6] We appreciate this suggestion by an anonymous referee.

[7] Indiana University's Big Red II features a hybrid architecture based on two Cray Inc. supercomputer platforms, comprised 344 XE6 compute notes and 676 XK7 "GPU-accelerated" compute nodes, totaling 1020 compute nodes, 21,824 processor cores, and 43,648 GB of RAM.

[8] The formal justification that this is a lower bound is based on two steps: in step 1, the cost inventory is charged in installations above 1 is lower-bounded, by the installation holding cost multiplied by 1 unit time, while inventory may spend more than a single unit time in any installation. Clearly the average number of units must be μ. Step 2 is based on ignoring the potential unavailability of inventory, which might cause stage 1 to starve.

[9] We accentuate "optimal" for these target levels because they will be optimal within the confines of the *2-MEBS* policy structure.

[10] Using the notation of Parker and Kapuściński (2004): stage inventory levels are x^1 and x^2, whereas echelon levels are X^1 and X^2, the new arrangement requires that echelon 2 pays for all echelon holding costs h_2X^2. To implement this, supplier needs to pay for all inventory at retailer (**consignment**), $h_2(X^1)^+$, and to be reimbursed for negative inventory at the retailer, $h_2(-X^1)^+$, which is the **additional backlog penalty**.

[11] We discuss information aspects further below.

[12] As defined in Fudenberg and Tirole (1991), p. 541.

References

Angelus A, Zhu W (2013) On the structure of capacitated assembly systems. *Oper. Res. Lett.* 41(1):19–26.

Axsäter S, Rosling K (1993) Installation vs. echelon stock policies for multilevel inventory control. *Management Sci.* 39(10): 1274–1280.

Cachon GP, Zipkin PH (1999) Competitive and cooperative inventory policies in a two-stage supply chain. *Management Sci.* 45(7):936–953.

Chen F (1998) Echelon reorder points, installation reorder points, and the value of centralized demand information. *Management Sci.* 44(12):S221–S234.

Chen F (2003) Information sharing and supply chain coordination. de Kok AG, Graves SC, eds. *Handbooks in OR & MS, Volume 11, Supply Chain Management: Design, Coordination and Operation* (Elsevier, Amsterdam), 341–421.

Choi H-P (2010) Information sharing in supply chain management: A literature review on analytical research. *California J. Oper. Management* 8(1):110–116.

Clark AJ, Scarf H (1960) Optimal policies for a multi-echelon inventory problem. *Management Sci.* 6(4):475–490.

Federgruen A, Zipkin PH (1984) Computational issues in an infinite-horizon multiechelon inventory model. *Oper. Res.* 32(4):818–836.

Federgruen A, Zipkin PH (1986a) An inventory model with limited production capacity and uncertain demands I: The average-cost criterion. *Math. Oper. Res.* 11(2):193–207.

Federgruen A, Zipkin PH (1986b) An inventory model with limited production capacity and uncertain demands II: The discounted-cost criterion. *Math. Oper. Res.* 11(2):208–215.

Fudenberg D, Tirole J (1991) *Game Theory* (MIT Press, Cambridge, MA).

Gavirneni S, Kapuściński R, Tayur S (1999) Value of information in capacitated supply chains. *Management Sci.* 45(1):16–24.

Glasserman P (1997) Bounds and asymptotics for planning critical safety stocks. *Oper. Res.* 45(2):244–257.

Glasserman P, Tayur S (1994) The stability of a capacitated, multi-echelon production-inventory system under a base-stock policy. *Oper. Res.* 42(5):913–925.

Glasserman P, Tayur S (1995) Sensitivity analysis for base-stock levels in multiechelon production-inventory systems. *Management Sci.* 41(2):263–281.

Glasserman P, Tayur S (1996) A simple approximation for a multi-stage capacitated production-inventory system. *Naval Res. Logist.* 43:41–58.

Ha AY, Tong S, Zhang H (2011) Sharing demand information in competing supply chains with production diseconomies. *Management Sci.* 57(3):566–581.

Hall G, Rust J (2000) An empirical model of inventory investment by durable commodity intermediaries. *Carnegie-Rochester Conf. Ser. Public Policy* 52(1):171–214.

Hariharan R, Zipkin PH (1995) Customer order information, lead times, and inventories. *Management Sci.* 41(10):1599–1607.

Huh WT, Janakiraman G (2010) Base-stock policies in capacitated assembly systems: Convexity properties. *Naval Res. Logist.* 57(2):109–118.

Huh WT, Janakiraman G, Nagarajan M (2010) Capacitated serial inventory systems: Sample path and stability properties under base-stock policies. *Oper. Res.* 58(4):1017–1022.

Huh WT, Janakiraman G, Nagarajan M (2016) Capacitated serial inventory systems: Policies and bounds. *Manufacturing Service Oper. Management* 18(4):570–584.

Janakiraman G, Muckstadt JA (2009) A decomposition approach for a class of capacitated serial systems. *Oper. Res.* 57(6):1384–1393.

Lau JSK (2007) *Information Sharing in Supply Chains: Improving the Performance of Collaboration* (Erich Schmidt Verlag, Berlin).

Lee HL, Whang S (1999) Decentralized multi-echelon supply chains: Incentives and information. *Management Sci.* 45(5):633–640.

Li L (2002) Information sharing in a supply chain with horizontal competition. *Management Sci.* 48(9):1196–1212.

Li L, Zhang H (2008) Confidentiality and information sharing in supply chain coordination. *Management Sci.* 54(8):1467–1481.

Lippman SA, McCardle KF (1997) The competitive newsboy. *Oper. Res.* 45(1):54–65.

Nagarajan M, Rajagopalan S (2009) A multiperiod model of inventory competition. *Oper. Res.* 57(3):785–790.

Netessine S, Rudi N (2003) Centralized and competitive inventory models with demand substitution. *Oper. Res.* 51(2): 329–335.

Parlar M (1988) Game theoretic analysis of the substitutable product inventory problem with random demands. *Naval Res. Logist.* 35: 397–409.

Parker RP, Kapuściński R (2004) Optimal policies for a capacitated two-echelon inventory system. *Oper. Res.* 52(5):739–755.

Parker RP, Kapuściński R (2011) Managing a non-cooperative supply chain with limited capacity. *Oper. Res.* 59(4):866–881.

Rosling K (1989) Optimal inventory policies for assembly systems under random demands. *Oper. Res.* 37:565–579.

Shang KH, Song J-S, Zipkin PH (2009) Coordination mechanisms in decentralized serial inventory systems with batch ordering. *Management Sci.* 55(4):685–695.

Speck CJ, van der Wal J (1991) The capacitated multi-echelon inventory system with serial structure: 1. The 'push ahead'-effect. Working paper, Eindhoven University of Technology, Netherlands.

Tayur S (1993) Computing the optimal policy for capacitated inventory models. *Comm. Statist. Stoch. Models* 9(4):585–598.

Veinott AF (1966) The status of mathematical inventory theory. *Management Sci.* 12(11):745–777.

Roman Kapuściński is the John Psarouthakis Research Professor of Manufacturing Management, Technology, and Operations Management at the University of Michigan Ross School of Business. His primary research interests include stochastic modeling with applications to supply chain management, revenue management, with recent emphasis on energy markets.

Rodney P. Parker is associate professor of operations management and Fettig/Whirlpool Faculty Fellow at the Indiana University Kelley School of Business. He received his PhD from the University of Michigan and has held prior faculty positions at Yale University, Cornell University, and the University of Chicago. His research interests include supply chain management, inventory theory particularly with capacity limits, healthcare operations management, industrial organization, finance-operations interface, and Markov games.