1. Introduction

Managing operations in today’s competitive marketplace poses significant challenges. The traditional thought was that there were so many conflicts in the multiple demands on the operations function that trade-offs were made in achieving excellence in one or more of these dimensions. Thus, for example, if a firm competed on the dimensions of flexibility and quality, cost was of secondary concern. This type of thinking was motivated by the fact that customers requiring product variety and high quality were typically willing to pay a premium price. Similarly, if the competitive pressures required the supply of low cost products (i.e., price was the major competitive force), then flexibility in terms of new product variety was not necessarily a major priority for the firm and hence, was sacrificed. This type of analysis forced the traditional operations manager to visualize the trade-offs between competitive priorities were a way of life.

However, the quality revolution and increased efficiency of foreign manufacturers in the late 1970s and early 1980s changed our thinking. Thus, for example, regardless of the type of product, flawless quality was required at extremely competitive prices. In fact, the competitive realities of the current marketplace and the impact on the contemporary operations manager are aptly summarized below (Skinner, 1985):

Make an increasing variety of products, on shorter lead times with smaller runs, and flawless quality. Improve the ROI by automating and introducing new technology in processes and materials so that prices can be reduced to meet local and foreign competition. Mechanize – but keep the schedules flexible, inventories low, capital costs minimal, and the work force contented.

From a strategic perspective, this indicates that the dimensions of cost, flexibility, quality and delivery are not to be traded-off against one another but need to be simultaneously prioritized. In response to this, operations managers have focused their attention on a multitude of approaches. In our opinion, the advent of Total Quality Management (TQM) to achieve flawless quality, Flexible Manufacturing Systems (FMS)
to achieve quick response and agile manufacturing at low cost, and effective Supply Chain Management (SCM) mechanisms to deliver products quickly with low inventories can all be regarded as responses to these new competitive pressures. The focus of this invited review is on the effective management of operations using one of these approaches – SCM.

The remainder of this paper is structured as follows. In the next section, we propose a taxonomical framework for analyzing supply chains from an operational perspective. Section 3 identifies the relevant decisions at each stage of the supply chain. Subsequently Sections 4–6 reviews and critically evaluates the relevant literature on production/distribution planning at each stage in the chain. Finally, Section 7 concludes with directions for future research.

2. Supply chains – An operational perspective

In order to analyze supply chains, we first need to clearly define the domain of the area. From an operational perspective two features are of interest when dealing with a supply chain: (a) the supply chain network; and (b) the nature of the relationship between each stage in the network.

2.1. Supply chain network

From an overall perspective, our analysis of a supply chain network focuses on three distinct stages. The first stage is the supply network consists of \( M \) suppliers providing goods and services to several plants. This network consists of all suppliers of raw materials, fabricated parts, service parts and any other supplies to the plants. The second stage includes \( N \) plants where the actual transformation process occurs and the product/service is created. The third and final stage is the distribution network consisting of \( K \) distributors and this stage generates the demand for the product or service.

The first stage in our analysis of supply chains is more focused than that of Choi et al. (1996) who contend that the supply network consists of all upstream suppliers who provide inputs to those that precede the plants. There are three major reasons for this narrower focus. First, it facilitates an emphasis on suppliers who provide direct inputs to each plant. This, in turn, fosters the development of long-term relations and the establishment of supply contracts with its direct suppliers. Second, this type of configuration is more in line with contemporary business practices. For example, Lee et al. (1993) reports on the strategic analysis of supply chains by Hewlett–Packard which primarily focuses on the immediate suppliers to the plants operated by the company. A third and final reason is motivated by the fact that operational decisions (such as location and inventory management) are directly impacted by the immediate suppliers, and hence, such a focus is more appropriate.

At the second stage of the supply chain, the direct inputs from the suppliers are transformed into the final product/service. Two operational issues are of relevance when analyzing this stage of the supply chain. First, plants at this stage could be organized into multiple sets (e.g., fabrication plants, subassembly plants, and final assembly plants) where each set of plants are the primary input source for the downstream set of plants. Alternatively, all plants could be grouped into a common set where each plant is organized to completely manufacture a set of products/services. Second, regardless of the type of plant setup just described, another feature of the second stage in the supply chain is whether there are between plant materials flows. Here we are explicitly referring to material flows between plants engaged in the same activity set, i.e., between all fabrication plants, between subassembly plants, etc. Depending upon the plant setup configurations and the flow of materials, operational planning will obviously be affected.
The final stage of the supply chain focuses on the distribution network which generates the final demand for the product/service. In practical terms this network could be a set of customers, a set of distributors, a set of product dealers, etc. Although managerial issues in controlling and managing the supplier and distributor networks differ to some extent, any operational analysis of the supply chain needs to integrate both these networks simultaneously.

2.2. Nature of relationship

A second operational feature that impacts our analysis is the nature of the relationship between the three stages of the supply chain network discussed above. Two aspects play a role in determining the type of relationship: competitive position and type of relationship. If we assume that a single dominant firm controls one or more of the stages in the chain, then this firm could, in the short run, simply optimize its own operational decisions regardless of the impact of such decisions on the other stage(s) of the chain. In some cases, however, dominant firms have actually been in a position to foster more cooperative agreements in the chain. The case Wal–Mart implementing the concept of Vendor Managed Inventory with its suppliers is a case in point.

On the other hand, if the stages of the supply chain are controlled by multiple firms none of which has a clear competitive advantage, then independent decision making would probably be easier. However, these firms may form alliances to better compete in the domestic and international markets. In such cases, they may also be motivated to cooperatively make operational decisions to jointly optimize the entire supply chain. For example, Weiser Lock, facing increasing competition from East Asian manufacturers, decided to implement just-in-time manufacturing within its own plant so as to reduce inventory costs and decrease cycle time. In order to successfully carry out this initiative, the company personnel actively worked with its Canadian suppliers so that they could ship supplies as and when required. In sum, it appears that regardless of the competitive position of the companies at different stages in the supply chain, cooperative decision making is the more preferred mode of operation. This leads us to focus on integrative operational decision making in the supply chain in this invented review. Further, the focus is on production/distribution planning decisions within the entire supply chain network.

3. Production/distribution planning

Given the previous discussion of the supply chain network, we now proceed to identify the relevant decisions that need to be considered in jointly optimizing production/distribution planning decision in the entire network.

Supplier stage: The major decisions in this stage of the network are:

- How are the suppliers selected? What criteria are used in the selection of suppliers?
- How many suppliers for each category or set of materials should there be?
- What kind of relationship should be established with each supplier?
- What is the volume and frequency of shipments from each supplier? Shipments from suppliers can be made to directly to the plants or to intermediate stocking points to which we will refer to as Supply Distribution Centers (SDC’s). How should information be shared between suppliers and SDC’s/plants. What is the amount of inventory of each material that should be maintained at each SDC? What decision rules should be used for controlling inventory at each SDC?
- Distribution network: What is the configuration of the suppliers-to-plants distribution network? Should supplier shipments be sent directly to a plant or should there be intermediate stocking points between the suppliers and the plants. The advantage of the latter is that there can be a central location where the qual-
ity of inputs to each plant can be monitored and transportation costs can be reduced due to economies of scale. However, it also adds another link in the supply chain and could lead to a slower material flow.

- **Location/Allocation decision**: How many SDC's should be operated? Where should they be located? Which suppliers should ship in what quantities to which SDC's and/or which plants?

- **Plant stage**: Major decisions in this stage of the network are:
  - **Transformation network**: What is the network configuration of the transformation processes? What is the nature of the capacity constraints binding the plants in this network?
  - **Location/Allocation decisions**: Where should transformation processes be located? Does the flow of inventory from feeder plants (i.e. fabrication plants supplying assembly plants) cross organizational boundaries?

- **Inventory**: What effect do inventory practices at one plant have on other plants in the transformation network? What are the effects of centralized versus decentralized policies concerning inventory?

- **Distribution stage**: The major issues in this stage of the supply chain network are:
  - **Distribution network**: What is the network configuration of the distribution channel? Are there to be intermediate stocking points between the distribution centers and the supplier (plant)? How are the final customer demands satisfied? Do distribution centers supply customers directly or are demands satisfied through retail outlets?
  - **Location/Allocation decision**: How many distribution centers should be operated? Where should they be located? What is the aggregate capacity of each center? Which product demands should be handled by each distribution center?
  - **Inventory**: What is the amount of inventory of each product that should be maintained at each distribution center? What decision rules should be implemented for controlling inventory at each center? What is the role of information in the distribution network? Should there be complete or partial sharing of information within the entire channel for better management of inventory costs?

Each of these issues are discussed below.

### 4. Supplier stage

In this section we will specifically address the first three and the first part of the fourth major supplier stage decisions indicated in section three. The remaining issues are similar to those that are discussed in the distribution stage (Section 6). Supplier, SDC and plant in the supplier stage play the roles of plant, distribution center and customer, respectively in the distribution stage.

A typical manufacturer spends 60% of its total sales on purchased items such as raw materials, parts, subassemblies and components (Krajewski and Ritzman, 1996). Therefore, the supplier link in the supply chain appears to have significant cost-cutting opportunities. For a manufacturer, to successfully offer low-cost, high-quality products and deliver them to its customers in sufficiently short time, it is imperative that it has reliable suppliers that can deliver quality materials in the right quantities with short delivery times. As implied by the above statement, quality, cost and on-time delivery are the three most important criteria in the evaluation of supplier performance. In their survey Weber et al. (1991) identified these as the most important supplier selection criteria. Production facilities and capacity, technical capability and geographical location were also identified as important criteria. Geographical location had a higher ranking in Just-In-Time (JIT) environments. This, of course, is not surprising given the emphasis on nearby suppliers in JIT systems as demonstrated by Toyota City in Japan and Buick City in the US. In their study Weber et al. also discuss linear weighting models, mathematical programming models and statistical models as quantitative approaches to supplier selection.

Benton and Krajewski (1990) studied, through simulation, how poor supplier performance impacts on the quantity and the distribution of inventories and customer service in a variety of manufacturing envi-
environments. Supplier performance is measured in two dimensions: lead time uncertainty and the quality of materials supplied. Their conclusions include: (a) presence of intermediate stocking points in the manufacturing process makes the total inventories held and customer service, measured by average amount of backorders, less sensitive to lead time uncertainty; (b) for all the manufacturing environments considered, poor quality has a minimal effect on total inventories, yet its effect on backorders is rather pronounced; (c) component commonality (among the products manufactured) helps reduce the backlog levels at the expense of increased intermediate inventories and; (d) component commonality dampens the effects of lead time uncertainty but amplifies the effects of poor material quality.

Today firms are demanding a high level of performance from their suppliers while at the same time reducing their numbers. It is reported that the American auto industry operated with over 30,000 suppliers in the early 1980s, today that number is less than 4000 and by the end of the decade it is expected to dwindle down to less than 3000. As a natural consequence of the reduction in the number of suppliers, firms are working more closely with their suppliers. Furthermore, firms look at some of their major suppliers as partners which, in some cases take part in strategic activities such as product and process design and development and in many cases have complete responsibility for component testing and quality control (Kamath and Liker, 1994). Supplier performance and relations with suppliers have become critical components in JIT environments. A JIT manufacturer can significantly improve its competitive standing by also becoming a JIT purchaser. Following are some typical characteristics of successful JIT manufacturing/purchasing environments (see, for example, Arnold, 1998; Hahn et al., 1983; Manoocheri, 1984; Ansari and Modarress, 1988; O’Neal, 1987).

- The supplier makes frequent deliveries. This usually forces the supplier to become a JIT manufacturer itself.
- The supplier works with the buyer to improve cost, quality and delivery lead time.
- Through the suppliers quality improvement programs, quality is controlled at the source. Therefore, the buyer need not inspect the incoming materials.
- Both the buyer’s and the suppliers inventory holdings are reduced.
- The buyer works with fewer suppliers so that each supplier has a greater share of the business.
- The buyer and the supplier share information that will affect their respective businesses.
- The supplier is able to plan more effectively.
- The supplier becomes more competitive as a JIT supplier.

The above list indicates both the supplier and the buyer benefit from the successful implementation of a JIT manufacturing/purchasing system. Yet, it should be noted that in situations where there is significant imbalance of power between the supplier and the buyer, the more powerful party tries to push most of the implementation cost onto the other party. In a survey conducted in 1993, more than half of the suppliers to the automakers in the United States and one third of the suppliers to automakers in Japan agreed with the statement “JIT only transfers inventory responsibility from customers to suppliers” (Helper and Sako, 1995). This was due to the fact that JIT deliveries were not matched with JIT production and consequently suppliers ended up stockpiling inventory to meet their customers’ JIT delivery demands. There is also some research addressing the ordering policies (volume and frequency) that are jointly optimal for the buyer and the supplier. Most of the models that were developed to address this issue are EOQ based (Thomas and Griffin, 1996).

An important factor that facilitates closer relationships and information sharing between suppliers and their customers is the advances in information and communication technology such as electronic data interchange (EDI), Internet and Enterprise Resource Planning (ERP) systems such as the R/3 system of SAP. There are many instances where effective use of these technology tools resulted in significant savings throughout the supply chain. For example, through effective implementation of EDI technology Campbell Soup was able to realize significant cost reductions for itself and its retailers (Fisher, 1997). Suppliers that are users of the R/3 system can have their customers use the Internet to trigger certain business processes in
their R/3 systems. A customer can submit its purchase order and immediately receive information about availability, prices and delivery dates. Also, it is reported that SAP is in the process of adding supply chain functionality to its R/3 ERP system.

5. Plant stage

5.1. Transformation network

A transformation network links production facilities, conducting work-in-process inventories through the supply chain. Whether considered alone or modeled in conjunction with supporting suppliers and distribution systems, this network can be described by the multiple stage inventory problem, in which a set of inventoriable stocks is linked by necessary predecessor relationships. A general multi-product, multi-stage formulation assuming fixed costs of replenishment, linear inventory holding costs, and no capacity constraints is as follows:

$$\min_Z \sum_{p,t} h_{pt} I_{pt} + \sum_{p,t} s_{pt} Z_{pt}$$

subject to

$$I_{p,t-1} + X_{p,t-j(p)} - \sum_{q \in Q(p)} X_{pt} - I_{pt} = d_{pt} \quad \forall p,t,$$

$$X_{pt} - MZ_{pt} \leq 0 \quad \forall p,t,$$

$$X_{pt}, I_{pt} \geq 0 \quad \forall p,t,$$

$$Z_{pt} \in \{0,1\} \quad \forall p,t.$$

In this model, $p$ and $t$ are indices for products (stages, or stock keeping units) and time periods, respectively. The decision of how much of what to produce each period ($X_{pt}$) then determines inventory levels ($I_{pt}$) and the timing of replenishment ($Z_{pt}$) required each period. The model implies necessary predecessor relationships among stages, in that $Q(p)$ is the set of all immediate successor stages of $p$. In this environment, some products can represent components or work-in-process inventory, required to support the production of parent products, or finished goods. This comprises the essence of a multiple stage inventory system. The objective function minimizes the total combined linear inventory holding costs ($h_{pt}$) and fixed replenishment costs ($s_{pt}$) of the production schedule, where replenishment is associated with the initiation of production that may lag availability by a fixed lead time of $j(p)$. Eq. (2) balances inventory available against successor stages’ consumption and external demand ($d_{pt}$). Eq. (3) ensures that, given production/replenishment is scheduled for some stage $p$, production/replenishment at that same stage in that same period has been initiated, where $M$ is some number larger than any $X_{pt}$. Finally, the technological constraints on the decision variables are given in Eqs. (4) and (5).

A substantial body of research exists which addresses multiple stage inventory systems, primarily framing the problem as one of internal coordination within a plant. A transformation network housed within a single facility is may be known more familiarly as a material requirements planning environment. Earlier work on this type of inventory problem often focused on optimization. Confining the problem to the constant demand case, Crowston et al. (1973) used a dynamic programming approach for assembly structures, while Schwartz and Schrage (1975) developed a branch and bound methodology. Relaxing the constant demand assumption in favor of time variant demand such as modeled in the previous formulation, solution procedures using dynamic programming (Crowston et al., 1973) and Lagrangian relaxation with branch and bound (Afentakis et al., 1984; Afentakis and Gavish, 1986) were then developed. The com-
putational burden of optimization associated with the multiple stage inventory problem provided a motivating factor for the examination of simpler heuristic techniques. The focus of these investigations has typically been on the application or adaptation of single product lot sizing algorithms to this dependent demand environment (e.g. Graves, 1981; Blackburn and Millen, 1982; Veral and Laforge, 1985; Bookbinder and Koch, 1990; Sum et al., 1993; Roundy, 1993).

One important characteristic shared by the research previously mentioned is the assumption of infinite production capacity. Extending this model to finite capacity conditions introduces the issue of the nature of those constraints. For instance, should the stages of a transformation network share a single resource, such as limited labor hours, this limitation could be modeled by appending this expression to the previous formulation:

$$\sum_p a_p X_{pt} \leq C_t \quad \forall t. \tag{6}$$

Under these conditions, a single resource of level $C_t$ in period $t$ is limiting total combined production in that period, where $a_p$ is the absorption coefficient of product $p$, or how much of the resource is consumed in production of one unit of that product. Many past studies (as examples, Dixon and Silver, 1981; Barany et al., 1984; Maloney and Klein, 1993; Brethauer et al., 1994; Gallego et al., 1996; Maes and Van Wassenhove, 1996) frame the issue of production capacity in this fashion, although not always including provisions for time variant demand or necessary predecessor relationships. Eq. (6) does not require that all products consume the single limited resource, in that the absorption coefficient of a product may be set to zero if its production or replenishment does not stress the network’s capacity. In this case, the single resource is usually some form of bottleneck workcenter within the transformation network, introducing the issue of set-up times as well as variable capacity absorption rates. Fixed set-up times ($b_p$) can be included in the formulation by expanding Eq. (6) as follows:

$$\sum_p (a_p X_{pt} + b_p Z_{pt}) \leq C_t \quad \forall t. \tag{7}$$

Trigeiro et al. (1989) study the effect of set-up times on this problem, in the single machine (multiple products, single stage network) case. With the inclusion of set-up times, the problem of finding a feasible solution (not necessarily an economical or optimal solution) becomes NP hard. A transformation network that spans multiple facilities and/or involves several heavily loaded processes may have more than one limitation on its output. Provisions for multiple resources can be include in this formulation by expanding Eq. (7) as follows:

$$\sum_k \sum_p (a_{pk} X_{pk} + b_{pk} Z_{pk}) \leq C_{kt} \quad \forall t, k. \tag{8}$$

where $k$ is an index of the set of limited resources of which production at various stages in the network may consume fixed and variable amounts. A smaller group of studies frame the issue of capacity in this broader fashion, among them the heuristic approaches of Maes et al. (1991) and Tempelmeier and Derstroff (1996). The initial feasibility problem encountered with the inclusion of set-up times can be eased somewhat if opportunities exist for overtime production in the network. This can be reflected by revising Eq. (8):

$$\sum_k \sum_p (a_{pk} X_{pk} + b_{pk} Z_{pk}) + U_{kt} - O_{kt} = C_{kt} \quad \forall t, k, \tag{9}$$

where overtime ($O_{kt}$) and undertime ($U_{kt}$) combine with actual consumption to balance the amount of resource $k$ available in period $t$, as in the earlier formulations of Billington et al. (1983). Provisions for the
backlogging of product demand (for example, Gupta and Brennan, 1992) implies similar benefits concerning solution feasibility.

5.2. Transformation/allocation decisions

Given that a transformation network can span several facilities, adopting an integrative view of the system can challenge the concept of location planning. Previously, location planning for manufacturing has been synonymous with plant location. Kogut (1985) discusses the strategic advantages of a firm with plant locations internationally, developing the distinction between comparative (locational) advantages and competitive (firm-specific) advantages. Kogut and Kulatilaka (1994) model the shifting of production between two geographically dispersed plants, treating the value of such flexibility as dependent on the real exchange rate.

However, Lee et al. (1993) reported an 18% reduction in inventory at Hewlett-Packard realized by revising the supply chain network of a particular product to reflect delayed customization. Also known as design for localization, this revision did not involve the relocation of facilities, but rather the transfer of particular operations between facilities, such that the later transformation stages were placed in existing distribution centers closer to market. The advantages of and opportunities for delayed customization are not apparent if the transformation activities within each facility are modeled as inherent in the definition of the facility.

When work-in-process flows between facilities in a transformation network, barriers may exist against efficient planning or optimization. Lee and Billington (1992, 1993) discuss poor coordination, discrimination against internal customers, and the creation of counter incentives through independent performance measures as potential problems within a given network. Foster (1996b) notes that inbound and outbound logistics may not be linked within a network, failing to capitalize on the savings that can be gained through two-way movement of freight. Similarly, many companies with multiple production facilities and one supplier fail to consider the total material requirements for all facilities.

5.3. Inventory

Anecdotal evidence suggests that transformation networks may be subject to greater opportunity losses when inventory flows cross the boundaries between firms. For example, Cottril (1996) reports surprising disarray at the component level of a particular automotive supply chain, stemming from the Just-In-Time practices of the assembly plants. Sometimes known as whiplash (Johnson, 1997), this can be interpreted as an inter-firm manifestation of production schedule nervousness, often studied (Kropp et al., 1983; Blackburn et al., 1986; Ho, 1993; Sridharan and Laforge, 1995) as an internal problem in material requirements planning. An interesting development in this area is JIT-II, in which several large manufacturing firms are encouraging suppliers to place employees on-site within the client facility (Dysart, 1993a, b). Arguably, this addresses those inefficiencies associated with the flow of inventory between firms in a transformation network, similar to those benefits gained through vendor managed inventory (VMI) typically seen at retail facilities (for example, see Devenyi, 1998).

Analysis of appropriate inventory policies within a transformation network raises the issue of centralized versus decentralized supply chain planning. As suggested in the previous section and discussed by Lee and Billington (1992), simplistic inventory stocking policies and poor coordination within a network can lead to substantial inefficiencies. Supply chain optimization demands centralized control, but can yield planning problems of great complexity. As an example, Martin et al. (1993) report successful application of mathematical programming to a combined production/distribution system consisting of four plants.
demand centers, and 200 products. Recent articles in practitioner's literature, such as Foster (1996a) or Burton (1996), suggest that one response to the increasing complexity of material planning in a supply chain framework will be to outsource that planning to third party firms.

6. Distribution stage

6.1. Distribution network

The type of distribution network to be implemented is a strategic level decision. In general, there are three generic types of networks which a supplier (manufacturer) can choose between in order to deliver the products to the marketplace. The first type of network consists of each manufacturer using a single intermediary stage (typically owned and operated by the manufacturer) where products are stocked and shipped to or purchased directly by the customers. This is typical of the network utilized by a majority of appliance manufacturer when supplying service parts (e.g., the distribution centers operated by Sears Roebuck where consumers can directly purchase service parts for washers, dryers, vacuum cleaners, etc.). This type of configuration is also used in the apparel industry which sets up company outlets in select locations (e.g., outlet stores for Liz Claiborne apparel).

In contrast to this first type of network configuration, the other two types include multiple manufacturers (as a supply point) and multiple independent retailers (where customer purchases take place). The primary rationale for these types of configurations is based on the fact that independent retailers tend to focus on satisfying demands for a large group of customers and hence, stock products from multiple manufacturers. The major difference between these two types of network configurations is whether there exists an intermediary in the distribution channel. Thus, a second type of network configuration is one in which multiple manufacturers supply products directly to the retail outlets while the third type is one with the same structure except that it also contains an intermediary, traditionally referred to as a wholesaler. The choice between these two types of network configurations is driven by the costs/benefits of centralization versus decentralization. Traditionally, it has been argued that an intermediary is preferred since the efficiency gains from centralization outweigh higher effectiveness from decentralization. On the other hand, recent arguments support the removal of an intermediary since it allows quicker response to changing market conditions and also more effective management of inventories.

There is no reported study in the literature which develops models or methods for identifying conditions under which one type of distribution network is preferred over another. Instead the focus appears to have been on managing the distribution process better given a network configuration. One potential explanation for this fact is that companies tend to use hybrid distribution networks in practice. For example, Smykay and Higby (1981) map the distribution network for a producer of canned fruits. In this case, the network configuration used is a function of the market segment served. For example, large retail grocery outlets (e.g., Supermarkets and Discount Hyperstores) receive their supplies from their own distribution centers; the smaller grocery outlets receive their supplies from cash-and-carry wholesalers; institutional customers (restaurants, schools, etc.) receive their supplies from an institutional wholesaler; and local bakeries receive their supplies from a bakery wholesaler. Another issue in network design is the lack of operational considerations in prior research. As the prior example indicates, networks have been organized around market segments but have ignored operational issues such as capacity, inventory and delivery times. Rather, it seems that these operational issues are a function of the marketing strategy adopted by the company. Given the current competitive pressures which have forced companies to reexamine the supply chain, our opinion is that there is a need to consider network design issues by integrating marketing and operations issues. Thus, for example, market segments could be identified based not only on customer attributes such as similar consumption patterns but also by simultaneously considering operational attributes such as location.
(which impacts supply lead times) and demand volumes (which impacts economies of scale and capacity). This would also be in line with current industry trends where the choice of distribution networks is made by adopting an integrated perspective.

6.2. Location/allocation decision

While the distribution network design is more strategic in nature, the issues of location/allocation can be regarded as tactical. The location decision addresses issues related to which set of distribution center (DC) sites should be selected while the allocation decision focuses on the potential service zones (in terms of products, customers, volumes, etc.) which should be served by each selected DC site. In general, there are two distinct set of approaches which could be used for addressing these decisions. One approach is to first address the location decision and subsequently the allocation decision while the second approach is to address both decisions simultaneously. From an application perspective, the former approach is easier to address since we are decomposing the overall problem into simpler optimally solvable individual problems (e.g., the location decision could be addressed using a knapsack approach while the allocation decision could be addressed using a transportation approach). On the other hand, given the interrelationship between the two decisions (e.g., costs of supplying customers are a function of which DC sites are chosen), it would be preferable to adopt the simultaneous location and allocation approach. Thus, our analysis is restricted to models/methods which can be used to analyze this decisions simultaneously.

Examining the distribution networks described in the earlier section, it is obvious that what we are interested in is determining the optimal locations of DC’s to serve customers (i.e., directly or through the retail outlets). A general multi-product formulation for this problem assuming that we have a single manufacturing plant is as follows: ¹

\[
\min Z = \sum_{p,k,l} C_{pkl} Y_{pkl} + \sum_k f_k z_k
\]

subject to

\[
\sum_l Y_{pkl} \leq S_{pk} z_k \quad \forall p, k,
\]

\[
\sum_p Y_{pkl} = d_{pl} \quad \forall p, l,
\]

\[
z_k \in \{0, 1\} \quad \forall k,
\]

\[
Y_{pkl} \geq 0 \quad \forall p, k, l.
\]

In this model, \(p\), \(k\), and \(l\) are the indices for products, DC sites, and customers, respectively. The two decisions are whether we should locate a DC at site \(k\) (\(z_k\)) and how much of a product \(p\) is supplied by a DC at site \(k\) to customer \(i\) (\(Y_{pkl}\)). In making these two decisions, the objective function minimizes the total linear transportation costs of supplying a product from a DC to a customer (\(C_{pkl}\)) as well as the total fixed costs of setting up a DC at a site (\(f_k\)). Assuming that we know the potential capacities of each DC site for each product (\(S_{pk}\)) and the demands for each product at each customer zone (\(d_{pl}\)), the first set of constraints (Eq. (11)) ensures that we do not exceed capacity limitations at each DC, while the second set of constraints (Eq. (12)) ensures that demand at each customer zone are met. Finally, technological constraints on the decisions are enforced through constraint sets (13) and (14). Note that logical relationships between the

¹ The proposed model implicitly assumes that plant capacity is adequate to supply all customer demands.
location and allocation decision variables (i.e., \( z_k \) and \( Y_{pkl} \)) are enforced implicitly through the capacity constraint.

This approach is an extended version of the capacitated facility location problem (the extension is in the context of the multiple product scenario) and hence, can be solved by modifying existing approaches for this problem (e.g., branch and bound approach of Akinc and Khumawala (1977); subgradient optimization algorithm developed by Geoffrion and McBride (1978); Lagrangian/benders’ decomposition scheme of Van Roy (1986); and the cutting plane algorithm developed by Leung and Magnanti (1986)).

A generalization of this model in the context of multiple manufacturing plants was developed by Geoffrion and Graves (1974). In their model, production capacity at each plant for each product is known and fixed. The product demands at multiple customer zones are known and fixed. These demands are satisfied by shipping products through distribution centers (DC) with each customer zone being exclusively assigned to a single DC. Potential site locations for DC’s are known, but the particular sites to be used are selected based on minimizing total network costs. These costs consist of fixed charges (for operating a DC), a variable operating cost (based on the amount shipped through a DC), and total transportation costs for routing products from a plant to a DC and eventually to a customer zone. This model formulation is as follows:

Minimize \( Z = \sum_{p,j,k,l} c_{pjk} x_{pjk} + \sum_k \left[ f_k z_k + v_k \sum_{p,l} d_{pl} y_{kl} \right] \) \hspace{1cm} (15)

subject to

\[ \sum_{k,l} x_{pjk} \leq s_{pj} \quad \forall p,j, \] \hspace{1cm} (16)

\[ \sum_j x_{pjk} = d_{pl} y_{kl} \quad \forall p,k,l, \] \hspace{1cm} (17)

\[ \sum_k y_{kl} = 1 \quad \forall l, \] \hspace{1cm} (18)

\[ \underline{V}_k z_k \leq \sum_{p,l} d_{pl} y_{kl} \leq \overline{V}_k z_k \quad \forall k, \] \hspace{1cm} (19)

\[ z_k \in \{0,1\} \quad \forall k, \] \hspace{1cm} (20)

\[ y_{kl} \in \{0,1\} \quad \forall k,l, \] \hspace{1cm} (21)

\[ x_{ijkl} \geq 0 \quad \forall i,j,k,l. \] \hspace{1cm} (22)

In this model, \( p, j, k, \) and \( l \) represent the indices for products, plants, distribution center sites, and customer zones, respectively. The objective function (Eq. (15)) specifies the transportation costs of each product to each customer zone (where \( c_{pjk} \) is the linear transportation cost and \( x_{pjk} \) is the amount shipped) and the total distribution center operating costs which are \( f_k \) the fixed cost (incurred if \( z_k \) is 1, i.e., site \( k \) is chosen as a DC location) and \( v_k \) the variable cost (incurred if \( y_{kl} \) is 1, i.e., product \( p \) demands are supplied to customer zone \( l \) through distribution center \( k \)). Constraints in the model are related to capacity limits for each plant (Eq. (16)); demands for a customer zone (Eq. (17)); each customer zone is supplied by a single DC (Eq. (18)); the total throughput for a DC is constrained between given lower (\( \underline{V}_k \)) and upper bounds (\( \overline{V}_k \)) (Eq. (19)); and technological constraints on the decision variables (Eqs. (20)–(22)). Note that the constraints given in Eq. (19) also help to maintain the logical relationship between the \( z_k \) and \( y_{kl} \) decision variables (i.e., any customer zone \( l \) can only be supplied by a DC site \( k \) provided a DC is located at that site). Although the model is computationally complex, the authors are able to develop an efficient solution procedure based on Benders’ decomposition. There are several features of this approach that can handle general types of location/allocation decisions for different types of distribution networks. For example,
general service time constraints (i.e., there is a maximum limit on the required time/distance between a customer zone and a DC) can be easily incorporated. Further, by setting up a dummy DC node with no lower and upper limits on throughput as well as zero fixed and variable costs, the user can also investigate whether it might be feasible to service all (or part) of the customer zones directly from the plants. In this latter case, the parameters $c_{pjkl}$ would simply reflect the direct transportation cost of supplying product $p$ to a customer zone $l$ directly from a plant $j$.\footnote{We also need to include constraints which link the $z_k$ and $y_{kl}$ variables for this dummy DC.}

More recently, it has been observed that the fixed costs of locating a facility at a site are dictated by the level of capacity installed. This is particularly appropriate in the context of distribution location/allocation decisions since it allows DC size (in terms of capacity) to be, for example, small, medium and large. Thus, instead of a single fixed cost of opening a DC at a site, these costs are a function of the size of the DC. A second feature which is also of relevance in terms of operating costs of a DC is whether it has been equipped to handle a particular product. The relevance of this feature stems from the fact that the variable costs of operating a DC could be product specific since different type of materials handling equipment may be required depending upon the product set handled by a DC. Lee, 1991, 1993 and Mazzola and Neebe (1998) analyze such a problem and refer to it as the Multiproduct capacitated facility location problem by facility type. Efficient solution procedures for this problem using cross decomposition (Lee, 1993) and lagrangian relaxation (Mazzola and Neebe, 1998) are developed.

### 6.3. Inventory

Analyzing the inventory decision in the distribution network has probably been the most extensively researched area. Traditionally, the focus has been on understanding the structure of the optimal decision (in terms of ordering quantities or review levels) assuming that demand patterns are known and stationary, and that the objective is to minimize ordering and holding costs. The general framework we use for examining inventory decisions in the distribution system is as follows. There are multiple supply sources (i.e., plants) each with a fixed capacity for each of several products and each source is required to supply a dedicated set of DC’s. In turn, each DC is required to supply a specified set of customers and there are aggregate stocking capacity limits for each product at each DC. Shipments between the DC’s and customers for each product occur continuously over a discrete time period. Thus, the DC is required to have adequate inventory of each product at the beginning of each time period. We implicitly assume that: (i) backordering of demand is not an option although it could be easily accommodated; and (ii) lead times between the plants and DC’s as well as between the DC’s and the customer are zero. In line with traditional approaches, the inventory decisions made by each DC and each customer focus on determining the order quantities by trading-off ordering and holding costs. A general model for analyzing the inventory decision for the entire distribution network is as follows:

$$\text{minimize } Z = \sum_{p,k,t} [s_{p,k,t}X_{p,k,t} + h_{p,k,t}I_{p,k,t}] + \sum_{p,l,t} [S_{p,l,t}Y_{p,l,t} + H_{p,l,t}B_{p,l,t}]$$

subject to

$$I_{p,k,t-1} + \sum_{j \in K_k} q_{p,j,k,t} - I_{p,k,t} = \sum_{l \in L_k} d_{p,l,t} \quad \forall p, k, t,$$

$$B_{p,l,t-1} + \sum_{k \in D_l} Q_{p,k,l,t} - B_{p,l,t} = d_{p,l,t} \quad \forall p, l, t,$$

$$23$$
In this model, $p, j, k,$ and $t$ are the indices for products, plants, DC’s, and time periods, respectively. Further, the sets $K_k, L_k, D_l,$ and $J_j$ are the set of plants from which DC $k$ obtains all products, the set of customers to which DC $k$ supplies products, the set of DC’s from which customer $l$ purchases products, and the set of DC’s to which each plant $j$ supplies products, respectively. The objective function minimizes the total ordering and holding costs for all DC’s and all customers in the system. Constraints in the model include inventory balance for each distributor and each retailer (24) and (25); capacity limits for each plant (26) where $P_pjt$ is the production capacity of plant $j$ in period $t$ for product $p$; capacity limits for the amount stocked at each DC (27) where $Ppklt$ is the stocking capacity of DC $k$ in period $t$ for product $i$; the linkage constraints (28) and (29) which indicate that the ordering cost is incurred only when an order is placed either at the DC or customer level; and the technological constraints (30), (31) and (32) on the decision variables. Note that the total demand for a product at a particular DC in a time period is simply an aggregation of the total demands for all customers which are to be supplied by the DC. If we needed to include a lead time of delivery to the customers from the DC, we can do so by defining the demands faced by a DC in a particular period as an aggregate quantity of the demands for each customer lagged by the delivery lead time (i.e., demand for product $p$ in period $t$ for DC $k$ is the sum of the aggregate demands for all customers in future period $tl$ where $tl = t$ represents the delivery lead time for each customer). The model can also be modified to account for lead times of supply from the plant to the DC’s by defining lagged availability for products in the inventory balance constraint (i.e., the demand in period $t$ is satisfied through the inventory at the beginning of the period plus the amount ordered for the product $t - L$ periods earlier where $L$ is the lead time).

To our knowledge, there are no existing methods which have been developed to solve such an integrated inventory problem with capacity constraints at the plant and DC levels. Given the complexity of the model, this is not surprising. Thus, prior work has focused on solving partial problems in this context. For example, several researchers have examined the single product, one DC and multiple customer problem (e.g., Veinott, 1969; Kalymon, 1972; Schwarz, 1973; Williams, 1981; Roundy, 1985; Maxwell and Muckstadt, 1985; McGavin et al., 1993; Lu and Posner, 1994) while others have examined the same problem with multiple products (e.g., Muckstadt and Roundy, 1987; Atkins and Iyogun, 1993; Iyogun and Atkins, 1993). Other work in this area has investigated the impact of logistics issues (Anily and Federgruen, 1990; Hodder and Jucker, 1985; Burns et al., 1985; Cohen and Lee, 1988, 1989; Cohen and Pyke, 1990), price discounts (Aull-Hyde, 1992; Benton and Rubin, 1993), fill rate optimization (Badinelli et al., 1985); non-identical retailers (Axsäter, 1995); dealer inventory systems (Cohen and Ernst, 1993); impact of lost sales (Nahmias and Smith, 1994); substitutable commodities (Pasternack and Drezner, 1991); impact of small lot ordering (Moinzadeh et al., 1997); integrated pricing and ordering decisions (Weng, 1997); use of scheduled ordering policies to manage demand variability (Cachon, 1997); vendor managed inventory in a promotional retail...
An emerging stream of research in distribution system inventory is on the impact of information flows between channel members. In general, this stream is motivated through empirical observations which suggest that channel intermediaries (e.g., wholesalers, distributors, retailers, etc.) have effectively developed coordination mechanisms for addressing inventory decisions through the efficient flow of information. Gavirneni et al. (1996) examine the value of information for managing echelon inventories in capacitated supply chains. They find that upstream firm costs decreased by up to 35% when capacity is not binding. Bourland et al. (1996) find that information sharing reduced upstream firm inventories by up to 62% but increased downstream inventories by up to 4.2% while Chen (1996) finds that information sharing reduces inventory system costs by up to 9%. Lee et al. (1997) shows that information sharing reduces the supplier’s demand variance while Chen and Zheng (1995) and Cachon and Fisher (1997) show that an inventory policy which implements shared information is close to optimal. Other work in this vein explores the issues of different inventory allocation rules in the case of restrictive capacity upstream (e.g., Cachon, 1997; Cachon and Lariviere, 1997; Graves, 1996; and Cachon and Zipkin, 1997).

Corbett (1997) examines cycle stocks, safety stocks, and consignment stocks in a two stage supply chain with asymmetric information. He shows that consignment stocks help to reduce cycle stocks but simultaneously, through incentives, the buyer is induced to increase safety stocks. Hsu and Zeng (1997) examine lead time-cost trade-offs in managing inventories and develop simple (but not optimal) heuristics for determining the order quantity for different lead time options. Chen et al. (1997) describe alternative coordination mechanisms for decentralized distribution systems. They show that coordination always increases channel profits and using cooperative game theory concepts, they attempt to formulate rules for sharing these profits and using cooperative game theory concepts, they attempt to formulate rules for sharing these profits among channel members. Seidmann and Soundararajan (1997) analyze four types of empirical inventory models for sharing logistics information: EDI, VMI, CR, and CM. They describe how competition, contracting and coordination costs affect the nature of value sharing in each of these models. They conclude by pointing out that the nature of competition among buyers and suppliers is the most critical element in understanding and evaluating information sharing arrangements. Anupindi et al. (1997) examine several profit sharing/revenue mechanisms for distributed distribution systems. They show the existence and uniqueness of the Nash equilibrium and compare total stocks under various mechanisms.

7. Directions for future research

Based on our review there are several future research directions for this important area. First, all three stages of the current supply chain are rich with opportunities for future research. Reassessing the traditional multiple stage inventory problem as a system that may span several organizations yields questions such as the contrast between a global optimal solution and locally optimal solutions developed politically within a given supply chain. When modeling the role of echelon inventory within a network, factors such as the ownership of the inventory have seldom been considered, an element that JIT and Vendor Managed Inventory (VMI) have rendered increasing significant in practice. Even if a given supply chain network is under a single ownership and centralized control, the integrative view of supply chain management obligates its managers to solve multiple stage inventory problems of a scale for which researchers have yet to identify consistently efficient solution procedures. Issues such as capacity, commonality, schedule nervousness and leadtime uncertainty, once considered internal factors within a single facility, can now be studied in the broader context of supply chains.

Second, integrated approaches to managing inventory decisions at all stages of the supply chain need to be developed. Current work has primarily focused on single stage or at most, two stage models and has
indicated that there are substantial benefits which can be achieved by coordinating inventory decisions in light of demand uncertainties as well as capacity constraints. However, there is a lack of approaches which explore these decisions simultaneously at all three stages. Such approaches can explore the impact of shifting capacity within the stages of the chain as well as how an uninterrupted flow of materials (due to better management of inventories) could lead to cost reductions for the entire chain.

Third, the use of information sharing in a multi-partner supply chain appears to be an accepted norm of operation. However, there are two issues which have received limited attention. The first deals with the determining contractual agreements between members of the supply chain. This is particularly important since the structure of such agreements will determine to a large extent whether each component of the chain will be motivated to expend the required effort to achieve the cost reductions by integrating decisions across the chain. The second issue deals with how supply chain partners can be incentivized to actually share information. From an individual partner perspective, there may be a hesitancy to share certain information (e.g., in terms of costs) since this could lead to a loss of competitive advantage for the partner. However, to efficiently manage decisions in an integrated chain, such information is essential. Thus, there is a need to develop mechanisms to incentivize the revelation of such information from each partner of the chain. To address both these research issues, the abundant literature in economics on contracts and mechanism design could be adapted to the supply chain setting.

Finally, as noted earlier analytical and simulation models that integrate the three major stages of supply chains is an important future direction of research in this area. In such models, however, it is almost always (at least implicitly) assumed that information flow is smooth and uninterrupted and the required information is available at the right time at the stage requiring it. Although rather comforting, this appears to be too big of an assumption. As the supply chains get longer and go beyond national boundaries, communication and information systems infrastructures to support the efficient operation of such complex systems become essential. ERP such as the R/3 system of SAP represents one major effort in that direction. Using ERP systems software to support academic research appears to be a promising research avenue.

References


Foster, T., 1996b. Supply chain black holes, Distribution, 63.


