Long-term capacity management: Linking the perspectives from manufacturing strategy and sales and operations planning

Jan Olhager*, Martin Rudberg, Joakim Wikner

Department of Production Economics, IMIE, Linköping Institute of Technology, SE-58183 Linköping, Sweden

Received 2 April 1998; accepted 2 February 1999

Abstract

Efficient long-term capacity management is vital to any manufacturing firm. It has implications on competitive performance in terms of cost, delivery speed, dependability and flexibility. In a manufacturing strategy, capacity is a structural decision category, dealing with dynamic capacity expansion and reduction relative to the long-term changes in demand levels. Sales and operations planning (S&OP) is the long-term planning of production levels relative to sales within the framework of a manufacturing planning and control system. Within the S&OP, resource planning is used for determining the appropriate capacity levels in order to support the production plan. Manufacturing strategy and sales and operations planning provide two perspectives on long-term capacity management, raising and treating different issues. In this paper, we compare and link them in a framework for long-term capacity management. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Capacity management; Manufacturing strategy; Sales and operations planning; Production planning; Resource planning

1. Introduction

The management of capacity in a manufacturing firm is often divided into three or four stages, ranging from long-term capacity planning to short-term capacity control and execution. Intermediate capacity management is related to rough-cut capacity planning, linked to the master schedule, and capacity requirements planning, linked to the material requirements plan. In this paper, we deal with the longest-term perspective. The relevant issues treated at the long-term capacity management level are related to determining when and by how much the capacity levels should change. Capacity is most often treated at an aggregate level, dealing with key work centres rather than all individual resources and based on forecasts of product families rather than of individual products. More specifically, long-term capacity management is most interested with the capacities that take a long time to change, either to acquire new capacity or to reduce capacity levels. Typically, the planning horizon is 1–5 years and the planning period is a month, at least for the first year, and then possibly quarters or even longer periods.

The input to long-term capacity management is a sales plan, based on a demand forecast. Such a sales plan must at least cover the time perspective for acquiring new capacity or reducing the relevant capacity. The sales plan can be translated into a corresponding capacity plan. However, decisions
regarding the production plan, in terms of e.g. production smoothing, means that production will not be identical to the sales plan. Furthermore, capacity may be acquired or reduced at times or quantities other than those required by the sales plan. Such issues are treated from two separate perspectives, on the one hand the manufacturing strategy perspective and on the other the sales and operations planning perspective, representing the highest planning level in a manufacturing planning and control system.

In a manufacturing strategy, capacity is considered as one of approximately seven decision categories, for which the manufacturing firm must have some policies. The other decision categories are facilities, production process, vertical integration, quality, organisation and personnel, and finally information or planning and control systems (see e.g. [1,2]). Capacity, facilities, production process and vertical integration are considered structural decision categories, dealing with the long-term manufacturing operations structure. The latter three decision categories deal with the manufacturing infrastructure. The capacity issues are related to the strategic relationship between capacity and demand levels, specifically translated into capacity expansion or reduction strategies. The basic foundation for such a strategy is that capacity comes in large, discrete steps rather than in small increments. Therefore, it is of strategic importance to decide whether capacity should come first, i.e. prior to expected changes in demand, or if capacity should be acquired first when the corresponding level of demand has been acknowledged. There are three different strategies in principle: lead, lag or track. Lead means that capacity is added in anticipation of increasing demand, whereas lag means the opposite. Track is a switching strategy, where the differences between capacity and demand levels are kept to a minimum. There is a corresponding discussion for capacity reduction. Thus, the main focus on capacity management from a manufacturing strategy perspective is the timing of capacity changes.

Sales and operations planning (S&OP) is the longest-term planning level in a manufacturing planning and control (MPC) system. Thus, S&OP belongs to a manufacturing infrastructure decision category from a manufacturing strategy perspective. Nowadays, many MPC systems support the MRPII structure, i.e. manufacturing resource planning, from where the term sales and operations planning originated. In early MRPII terminology, the term production planning was used instead of S&OP, but today the latter is being used to a larger extent. At the S&OP level, a production plan is developed based on a sales plan. Here the issue is related to the production level relative to the demand level in various periods. The principal options available are level, chase and mix (or combination). These are sometimes called planning strategies. Level means that a production rate is established over the planning horizon, whereas chase implies that production matches demand in a manner such that all demand in a period is produced in the same period (typically month). In the mix option a production rate is used for a few periods and then changed. Any difference between the sales plan and the production plan will result in a corresponding inventory or backlog plan — an inventory plan in a make-to-stock situation and a backlog plan for make-to-order or engineer-to-order environments. The production plan is translated into a capacity requirements plan in terms of aggregate resources in the so-called resource planning. Then, under- and over-capacities are identified in the resulting capacity plan. Thus, the focus of capacity management from an S&OP perspective is on the rate of production relative sales.

Both perspectives deal with the long-term management of capacity, i.e. how to best utilise ‘slow-moving’ resources for manufacturing operations. Still, the issues, objectives and options differ. According to the decision categories identified by Hayes and Wheelwright [1] we have structural decisions related to capacity (expansion or reduction strategies) and infrastructural decisions related to MPC systems (e.g. sales and operations planning). This paper links these two decision categories traditionally treated as belonging to different groups of categories. This raises interesting questions regarding the integration of capacity issues and whether they should be treated in sequence or not. The purpose of this paper is to discuss the relationships between the two perspectives, in order to integrate strategic and planning/control issues
into a framework for long-term capacity management. To start with, we treat each respective perspective, first the manufacturing strategy view and then the S&OP view. Thereafter, we study the relationship and the interaction between the two perspectives, and finally we propose a framework for how to link them.

2. Manufacturing strategy perspective

Hayes and Wheelwright [1] use three variables to describe a capacity strategy: the type of capacity needed, the amount of capacity that should be added (or reduced), and the timing of capacity changes. Since the type of capacity strongly influences the amount that is to be added or reduced, the first two are normally discussed together in the so-called sizing problem.

2.1. The sizing of capacity changes

A central part of the sizing problem is scale considerations, where economies as well as dis-economies of scale are weighted against each other, which leads to concepts such as optimal step change, optimal plant size, etc. Due to the inherent properties of most resources, capacity can normally only be changed in discrete steps with a considerable lead-time. Adding a new machine or facility would probably mean a significant capacity expansion making the capacity changing step-wise as illustrated in Figs. 1–3. Assuming the type and amount of capacity changes as given (which it typically is), the timing of capacity changes remains to be decided upon.

2.2. The timing of capacity changes

The timing variable in a capacity strategy is concerned with the balance between the (forecasted) demand for capacity and the supply of capacity. If there is a capacity demand surplus the utilisation will be high, thus enabling a low cost profile, but there is also a risk of loosing customers due to e.g. long delivery lead times. A capacity supply surplus on the other hand creates a higher cost profile but due to the surplus capacity it is easier to maintain high delivery reliability and flexibility. The capacity strategy can thus be expressed as a trade-off between high utilisation (low cost profile) and maintaining a capacity cushion (flexibility). Based on this, two ‘pure’ types of capacity strategies can be identified, usually referred to as leading demand (capacity supply surplus) and lagging demand (capacity demand surplus) in line with the framework in [1]. In their framework capacity strategy is, however, only defined for increasing demand. Our framework contains both increasing and decreasing demand patterns. In many cases it is not possible (nor desirable) to maintain a pure strategy and a middle way is chosen which contains aspects of both leading demand and lagging demand. This approach aims at finding an efficient trade-off between the pure strategies and is here referred to as tracking demand.

![Fig. 1. Capacity leading demand (capacity supply surplus).](image_url)
The objective of the lead strategy approach is to maintain a cushion of capacity that e.g. can be used to support volume flexibility and reliable lead times. If there is a positive trend in demand, capacity should be added in anticipation of demand, as shown in Fig. 1. A corresponding negative trend in demand would mean that capacity should be decreased to the demand level when eliminating a 'step' of capacity. The guiding rule in this case is that the capacity always should be larger than or equal to demand, thereby supporting the same order winning characteristics both in positive and negative trends of demand. Traditionally, this approach has been referred to as a lead strategy since capacity leads the increase in demand to maintain a capacity cushion. But, as can be seen in Fig. 1, when there is a decreasing demand pattern the objective of keeping a capacity cushion by necessity creates a lagging behaviour, not reducing capacity until a reduction can be conducted without endangering the maintenance of a capacity cushion. Therefore, the label lead strategy may be misleading, but it is used below anyhow since it is well established in the literature.

The lag strategy is based on the objective of maintaining a high utilisation of resources. This is of particular interest in environments where price is an order winner and thus the focus on a low cost per unit is central. The basic principle is to produce as much as possible and still maintain full capacity utilisation. Capacity should thus be added in reaction to increasing demand. The lag strategy is, however, challenging when demand is declining since the decision to reduce capacity should be taken when the utilisation is still high. The guiding rule in this case is that capacity should never exceed demand, as illustrated in Fig. 2. Again, the reason is to support the same order winning characteristics independently of the demand pattern. As discussed above in the case of the lead strategy, the lag label is also somewhat misleading. In a situation where demand is declining, the decision to decrease capacity must be taken when the utilisation is still high and thus capacity is leading demand.

In some cases one of the two extreme strategies described above may be desirable and achievable, but in most cases there must be a trade-off due to e.g. difficulties in forecasting the future demand. This combined approach is usually labelled tracking strategy. The objective is to track the demand as close as possible, hence putting more emphasis on the sizing problem. Reducing the size of the step changes facilitates a track strategy, which consequently minimises the deviations between demand and capacity. An example of the tracking strategy is shown in Fig. 3 indicating the presence of both under- and over-capacity under this strategy.

The three capacity strategies discussed above can be described as different positions on a spectrum. The end-points of the spectrum would at one end correspond to a situation where low cost, through high utilisation, is the norm, and at the other end to a situation where flexibility, based on over-capacity, prevails. In between the end-points combinations, based on trade-offs between the end-points, are found. Below, we discuss a corresponding spectrum from a manufacturing planning and control perspective.
3. Sales and operations planning perspective

In recent years the term sales and operations planning (S&OP) has been more frequently used in articles and textbooks, although mostly related to authors discussing MRPII or similar systems. Some authors use S&OP interchangeably with the terms aggregate planning or production planning. However, we want to distinguish S&OP from the others, where we see aggregate production planning (APP) as a part of the S&OP process, which is the long-term planning of production and sales relative the forecasted demand and the supply of capacity.

3.1. The S&OP process

S&OP is often referred to as a fundamental that maintains the balance between aggregate supply and aggregate demand, through monthly updates of the annual business plan, see e.g. [3]. Roughly, S&OP can be divided into a sales plan (based on forecasted demand) and a production plan, which of course both affects inventory and/or order back-log, and capacity requirements. The S&OP process is the forum where different functional strategies meet for establishing a production plan that economically serves the needs of the market, while supporting both the strategic and financial plans of the firm.

One of the most interesting features of S&OP is its part strategic and part tactical nature [4]. S&OP is on the one hand constrained by the capacity strategy, but on the other also influences the long-term capacity planning via feedback from the execution of the sales plan and production plan. Within the S&OP process the resource planning translates the production plan into a resource requirements plan, thereby reviewing the required capacity strategy for conformance. As tactical and operational decisions are made each month, the capacity plan gets executed or the need to review the capacity strategy becomes apparent.

The issue in the S&OP process is to create a balance between the sales plan and the production plan. In doing this, planners must make decisions on marketing activities, output rates, employment levels, inventory levels, backlogs, subcontracting, etc. The decision options can roughly be divided into two types; those trying to modify demand to match the production constraints, and those modifying supply to match the sales plan (sometimes referred to as aggressive and reactive alternatives, see e.g. [5]). The supply option is normally further derived into the well-known methods where production is trying to either level production or chase sales (i.e. forecasted demand), or some kind of combination or mix of the two (see Fig. 4). We label these methods planning strategies.

![Fig. 3. Capacity tracking demand.](image)

![Fig. 4. Tactical decisions in the S&OP process.](image)
3.2. Modify demand

Modifying demand will not be discussed in any detail in this paper. Nevertheless, using a set of marketing tools to influence the sales pattern can be seen as a way to achieve a more levelled production. Such tools can for instance be campaigns, new product introduction, complementary products, pricing, etc.

3.3. Modify supply

The supply option can be divided into three categories – level, chase and combination/mix planning strategies. The two extremes (level and chase) are seldom used in isolation. The process industry, however, is normally using a pure level strategy, and a pure chase strategy can for instance be found in engineer-to-order environments. When discussing planning strategies, it is of great importance to distinguish between different market environments: engineer-to-order (ETO), make-to-order (MTO), assemble-to-order (ATO) (sometimes referred to as finish-to-order, FTO) and make-to-stock (MTS) all call for different approaches when the supply is modified. In the following, we let the MTO case to also cover for the ETO case.

When production is levelled, constant output rates are maintained during the planning horizon. Buxey [6] defines level as effectively decoupling supply from demand. A similar definition is also found in Johansen and Riis [7], where they state that production is decoupled from forecasts and actual market demand for a period of several months. The aim is to achieve a uniform and high utilisation of production resources, including a minimisation of costs related to changes in production rates. In an MTS environment, the tactics can roughly be described as maintaining a steady production (output) rate throughout the planning horizon, absorbing fluctuations in the sales plan by changes in inventory level. In MTO environments fluctuations are managed by changes in the order backlog, i.e. increasing or decreasing the backlog, which of course affects the delivery lead times and reliability (see Fig. 5 for an example of a level strategy). Note that the backlog can be seen as ‘negative inventory’, and that the backlog curve in this sense is the inverse of the inventory curve. However, fluctuations in sales can also be managed by in- and outsourcing, subcontracting, etc [8], without changing the internal production rates.

When the sales plan is chased, output rates are adjusted to match sales each period during the planning horizon. Buxey [6] defines chase as entailing exact synchronisation between (forecasted) demand and supply, while Pan and Kleiner [9] define it as matching forecasted demand period by period. The aim is to minimise investment in inventories and/or the order backlog by producing to the actual sales plan or customer orders. In the ideal case, with very short lead-times, the MTS and the MTO environments collapse into a ‘make-to-requirements’ (MTR) environment without any inventory or order backlog. In the normal case inventory and/or backlog do exist, but are kept on a stable level. Flexibility and adaptability are achieved at the expense of low utilisation of production resources and at higher costs associated with changes in output rates. The literature suggests numerous ways of adopting the chase option; changing employment levels (hire/fire), using overtime or idle-time (excess capacity), flexible working hours, etc.

As mentioned above a combination of the two extremes will commonly be observed in practice. In this combination/mix strategy, output rates are changed during the planning horizon, but not as frequent as when the sales plan is chased. Production can for instance be levelled over a couple of periods, followed by a chase strategy, then altered back to a level strategy, and so forth (see Fig. 6, showing an example of the combination/mix strategy). The combination/mix option requires some anticipation inventory build-up during slack periods (in an MTS environment), or increases in the
order backlog during peak periods (in MTO environments), but only a few changes of output rates are necessary. This planning strategy thereby enables a better correspondence between sales and production output than does a level strategy, moreover it creates a more stable production environment than does a chase strategy. Decisions regarding how often production rates shall be altered during the planning horizon, is based on the trade-off between costs associated with maintaining a steady output rate and costs associated with changes in output rates.

As in the case of capacity strategies, the planning strategies can be described as different positions on a spectrum, with level and chase as the endpoints. Level would correspond to a situation where uniform and high utilisation is the norm, while chase would correspond to a situation where (volume) flexibility is the norm. Along the spectrum different types of combination/mix planning strategies are to be found, where the choice of strategy is based on the trade-off between the endpoints. After a short description of models and methodologies used in S&OP, we then turn to discuss how the two perspectives (capacity strategy and S&OP) interact with each other.

3.4. Models and methodologies

Johansen and Riis [7] provide a framework for choosing appropriate demand management depending on the strategy focus. In their framework a number of company specific factors (like forecast accuracy, intensity of season, process complexity, etc.) and political and social factors (like interest rates, rate of unemployment, etc.) are used to guide managers in the S&OP process. The framework aims at helping managers not only to choose the right option, but also suggests a couple of supportive methods that can be used in connection with these. A more hands-on tool is the S&OP reports (see e.g. [3]), where past and current performance as well as future plans are easily visualised. However, the cost trade-off mentioned above has been an object of research for several decades, wherefore the largest portions of articles in the S&OP/APP area focus on mathematical models trying to optimise the choice of planning strategy. The basic mathematical models can be described as follows: given a set of forecasts \( F_t \), determine production, inventory/backlog, and workforce levels \( P_t, I_t, \) and \( W_t \), respectively, \( t = 1, 2, \ldots, N \), which minimises cost subject to appropriate constraints. \( N \) is the time horizon, normally reaching over 6–18 months [9]. In choosing \( P_t, I_t, \) and \( W_t \) relative to \( F_t \), the decision-makers choose the desired trade-off between the important cost parameters. Nam and Logendran [10] made a comprehensive survey of APP models containing 140 journal articles (from 17 journals) and 14 textbooks, ranging from the early 1950s to 1990. However, all mathematical models for aggregate planning depend on the estimation of a number of relatively uncertain cost parameters, and despite the numerous approaches and techniques covered in the literature, only a few have been implemented in an industrial situation [10].

4. Interaction of the two perspectives

As mentioned in the introduction the capacity strategy mainly focuses on the timing of capacity changes, whereas the S&OP’s primary focus is on the rate of production. Both perspectives do deal with long-term capacity management, but since they belong to different decision categories they are traditionally treated in sequence, where the S&OP has to work within the frames set by the capacity strategy. Yet, there is a strong interrelationship between them, which is discussed below.

4.1. Manufacturing strategy’s impact on sales and operations planning

The decision on when and how to acquire new capacity strongly influences the action space for
S&OP. This strategic manufacturing decision determines whether capacity is adjusted before a change in demand, or only once a demand change has been experienced. Consequently, the planning strategy of S&OP will be more or less forced to work within the capacity levels set by the capacity expansion or reduction strategy. Still, should a production plan be chosen that will need extra capacity, then that decision will implicitly imply that this extra capacity is of temporary nature, and not a new capacity acquisition.

If a lead strategy is followed, then S&OP is left with much freedom. However, using a lag strategy will force the S&OP to act within much tighter capacity levels, restricting the use of a chase planning strategy. If a chase strategy is desirable, then there is a need for extra capacity in terms of sub-contracting, short-term overtime, or capacity flexibility. A track strategy usually means that capacity can be acquired in smaller steps, allowing for a closer fit between capacity and demand. Such a capacity expansion strategy would still imply that the planning strategy will use sub-contracting or overtime at some instances and have over-capacity at others, typically for MTO situations, or build inventory in MTS environments.

4.2. Sales and operations planning’s impact on manufacturing strategy

The decision on how to produce also influences the need for new capacity. Production smoothing will keep the maximum capacity requirements to a minimum, whereas more capacity is needed in one form or another if production volumes vary a lot between periods. Also, in capacity reduction situations a level production rate will facilitate the decision on capacity levels.

If products can be produced in a long-term stable production rate and there is an upward demand trend, then new capacity can be postponed. The option to use a level production rate is usually quite good in an MTS environment. On the other hand, if products are made or engineered to order and a chase planning strategy is pursued, then new capacity will be needed earlier as demand increases.
relative to an MTS situation. A mix or combination strategy would lead to a situation between these two other (extreme) planning strategies.

Even in MTO situations, new capacity investments can be delayed if production can be levelled rather than be chasing demand. The reason is that the maximum capacity requirements are reduced, thereby reducing the need for extra capacity until the average capacity requirements indicate such a need. In one case, where one of the authors was involved, a company would have been able to delay a new investment by approximately one year, should they have adopted a production smoothing option.

4.3. A framework for combining the two perspectives

The product/process matrix [1] can be used to illustrate the correspondence between capacity strategies and planning strategies (see Fig. 7). Typically, manufacturing firms are found along the diagonal from the upper left-hand corner to the lower right-hand corner that is roughly represented by positions 1–4. Low-volume manufacturing of non-standard, one-of-a-kind products is typically performed in a job shop. In such an ETO environment (see e.g. [11]), a typical order winner is flexibility (see e.g. [12]). Then, a lead strategy (according to our definition of capacity supply surplus) is preferable, to allow for volume flexibility. Also, due to the inherent nature of ETO not to manufacture to a forecast, a chase strategy is preferable. At the other end, high-volume, standard commodity items are typically manufactured in continuous processes. Such items are typically made-to-stock (see e.g. [11]) and the predominant order winner is price (see e.g. [12]). Due to the focus on price, resource utilisation is most important leading to a lag strategy (capacity demand surplus), in order to avoid over-capacity. Also, the MTS situation and the importance of cost-efficiency lead to a level production rate. The main reason is again to avoid unnecessary temporary over-capacity. In this paper, we add capacity strategy and planning strategy to the list of characteristics related to the integrated choice of product/process environment.

As can be seen in Fig. 7, there is a correspondence between, on the one hand, lead and chase, and on the other, between lag and level. In this framework, it is necessary to treat the lead strategy as corresponding to a capacity supply surplus situation, and the lag strategy must correspond to capacity demand surplus. Hereby, the capacity strategies support the same order winning characteristics both for expanding and declining demand patterns.

If capacity resources are acquired in large discrete steps, then it is important to distinguish between capacity acquisition (or reduction) and capacity control. The former is dealt with from a manufacturing strategy perspective, whereas the latter is treated in the sales and operations plan. However, these decisions need to be taken in an integrated fashion, and not in sequence. The reason is that there is interdependency between these decisions, such that both influence the decision space of the other.

The main results from linking the perspectives of capacity expansion strategies and planning strategies are summarised in Fig. 8.

The upper left-hand corner and the lower right-hand corner both indicate situations where the two perspectives are mutually supportive. Lead and chase focus on resource availability and flexibility to provide order winners such as flexibility, design and quality. Lag and level combine their respective focus on resource utilisation, in order to support competition on price. The lower left-hand corner illustrates a situation that is neither supportive nor conflicting. Using a combination of lead and level means that production rate changes can be accommodated if needed, i.e. a possibility to move towards a mix or chase planning strategy. Also, using a level strategy means that the maximum capacity requirements are kept to a minimum and thereby new capacity acquisitions can be postponed. The real problem is the upper right-hand corner where the strategies are conflicting. Using capacity demand surplus means that there are very limited opportunities to execute a chase strategy. Instead, the company must rely on other, external sources for additional capacity when needed to support a chase plan. Still, the company is likely to face frequent overloads and subsequently major delivery problems. The positions for track and mix/combination can be viewed as intermediate relative to the other extreme compositions.
5. Summary and extensions

In this paper we have linked a structural and an infrastructural decision category from the manufacturing strategy framework. The structural perspective deals with capacity in terms of capacity levels and expansion/reduction strategies. The infrastructural perspective deals with S&OP in terms of planning strategies for production relative to sales, inventory and/or backlogs. Together they provide two perspectives on long-term capacity management. The decisions related to capacity acquisition/reduction and planning/control are interrelated in such a way that they must be integrated and cannot be dealt with in sequence. We have plotted these two characteristics relative to the product/process matrix [1] and found a strong relationship. Typically, lead and chase are related because they both put a strong focus on resource availability and flexibility. Moreover, lag and level both strongly focus on resource utilisation. When discussing capacity reduction strategies relative to the lead, track and lag strategies suggested by Hayes and Wheelwright [1] for capacity expansion strategies, we suggest that these three strategies, related to the capacity level relative sales, should be interpreted in the following manner. A lead strategy is synonymous to capacity supply surplus, whereas a lag strategy assumes capacity demand surplus.

An extension to this research is to relate the long-term capacity management issues to the product life cycle, to investigate the sensitivity relative to the various life cycle stages, e.g. taking market growth rate into account. Furthermore, the design of decision support systems for long-term capacity management, i.e. how to combine and integrate these issues, needs further attention.

Acknowledgements

The research is supported by grants from the Swedish Foundation for Strategic Research, Ericsson Radio Systems AB and Volvo Research Foundation, Volvo Educational Foundation and Dr Pehr G Gyllenhammar Research Foundation.

References