Hierarchical planning in a single stage system

Fransoo, J.C.

Published: 01/01/1993

Document Version
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Hierarchical planning in a single stage system

Jan C. Fransoo

Research Report TUE/BDK/LBS/93-23
October 1993

Eindhoven University of Technology
Graduate School of Industrial Engineering and Management Science
P.O. Box 513, Paviljoen F12
NL-5600 MB Eindhoven
The Netherlands
Phone: +31.40.472681
E-mail: BDKBJF@urc.tue.nl

This paper should not be quoted or referred to without the prior written permission of the author.

This research has been sponsored by the Netherlands Organization for Applied Scientific Research, Institute for Production and Logistics (IPL-TNO). The current research of Dr. Fransoo has been made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences.
Hierarchical planning in a single stage system

Jan C. Fransoo
Eindhoven University of Technology
The Netherlands

Abstract
Single stage systems with high set-up times and high utilization levels occur in flow process industries. In this paper a two-tiered hierarchical model is developed for such a system. At the top level, the optimal value of the control parameters is determined, while the operational scheduling function is performed at the bottom level. The conceptual aggregation approach used in this model is compared to the aggregation approach used in classical hierarchical approaches.

1. Introduction

Single stage systems have been the object of analysis in many studies in production management. Specifically in the scheduling area of research many publications have appeared. On the one hand, these studies focus on the optimal sequence of a number of jobs on a single installation (for a review, see e.g. Gupta and Kyparisis 1987). On the other hand, there is the vast body of literature concerned with lot-sizing problems (for a review, see e.g. Elmaghraby 1978 or Salomon 1990). Single stage systems find their practical applicability especially in flow process industries, where often a single installation determines the major part of the added value of the products manufactured. In this respect, one could think of industries such as bulk chemicals, glass manufacturers, and steel and paper producers.

There are some typical characteristics of these single stage production systems in practice, which we will consider in this paper. These characteristics include the following:
- the utilization of capacity is very high. This is due to the fact that the installations usually are expensive, while the added value of the products is rather low. This requires a very high production speed and utilization of
The production control problem that we study is essentially when to produce which product for how long (or how many). Additionally, we include the demand management decision: which product demand to fill? For ease of explanation, we assume that all product demand has to be met from stock. Any demand which is not delivered from the finished goods inventory, gets lost.

In this paper, we will develop a hierarchical production planning model for the single stage system described above. The model, which is introduced in Section 2 of this paper, consists of two levels: long-term capacity coordination and short-term production order scheduling. The two levels will be addressed in Sections 3 and 4, respectively. In Section 5, the hierarchical concept developed in this work will be analyzed methodologically, comparing it with the well-known hierarchical approach introduced by Hax and Bitran (Bitran and Hax 1977). Finally, we will draw some conclusions as to the use of hierarchical models in single-stage systems.

### 2. A Hierarchical Model

In the literature, three reasons can be found for using a hierarchical approach. The first refers to the uncertainty which exists in the long run. This uncertainty makes it impossible to set up a detailed plan for the long term. Therefore, hierarchical levels are introduced. Each of the hierarchical levels deals with a different level of uncertainty. At a lower hierarchical level, new information becomes available and uncertainty is reduced. This is in line with the work of Bitran and Hax (1977). However, this reason only makes sense

---

*Since the added value is presumed to be quite low, profit will mainly be realized and increased by maximizing the throughput of the system. This means that any loss of capacity will have its immediate consequences for the profit of the system. If the added value would be higher, then the direct consequences of throughput loss due to change-overs are smaller.*
if the aggregate (upper level) decision has to be taken earlier than the detailed (Lower level) decision. This may be necessary because the variables decided upon at the aggregate level cannot be determined as frequent as the lower level variables, or because it takes some time before such an aggregate decision takes effect in the (physical) object system. The second reason to choose for a hierarchical approach is the complexity of a monolithic model. A monolithic model requires detailed scheduling information to be projected out to a long-term horizon. This may cause severe computational and data input problems (Silver and Peterson 1985, p. 515). The third reason is provided by Meal (1984), who states that the production control system should fit the organizational structure. In this respect, it should be stressed that a decision is not taken by an algorithm, but by an officer responsible for this decision in the organization. It is clear that different organizational levels decide on long-term and short-term issues. This should be reflected in the decision structure. Additionally, the production planning and control structure is by and large not the only determinator for the organizational design. Many other arguments influence the organizational design and in many cases the decision structure for production planning just has to be fitted into the existing organizational design.

We would like to add that there exists a difference in the nature of the models at the longer term and the short term. Departing from the problem, we argue that (in line with the organizational aspects mentioned by Meal), the nature of a long-term decision function is distinct from the nature of a short-term decision function. The objective of the long-term decision function is to ensure that the production system will be able to reach its objectives over more than one time period, while the objective of the short-term function is to maintain control within the long-term objectives. So the top level of the hierarchical model is focused on setting the control parameters in such a way that the objective (profit) is maximized, while the bottom level of the model aims at managing the realization of the control parameters and keeping them at the target levels set by the top level.

The control parameters are the parameters which determine the logistic performance of the production system. This system performance is measured in two ways. The first performance measure is the relevant profit of the system. This is the part of the profit that is influenced by the production planning decisions. It is the objective of the system to
maximize this profit. This objective is formulated at the model's upper level. The second performance measure is the *controllability* of the system. If the system is controlled by the planning hierarchy, this means its performance can be influenced. The controllability is measured by the comparison of the individual products' service levels that are realized with the expected product service levels in the aggregate model (i.e., at the upper level).

![Diagram](image)

Figure 1. *Inventory pattern of a single product on a multi product installation.*

The performance of this single stage system is determined by two kinds of system parameters, namely the *target inventory level* and the *cycle time*. This is clarified in Figure 1. This figure shows the idealized inventory pattern of a specific product on a multi-product installation. A product is manufactured until the target inventory level has been reached. Then, the other products are manufactured. The cycle time of the product is defined as the time period between two consecutive starts of production runs of the same product. It is clear that the service level is increased if the target inventory level is increased and vice versa. Also, the inventory costs will rise if the target inventory level is increased. This enables a trade-off between extra contribution* and extra inventory costs. On the other hand, the cycle time will have to be increased, since extra productive capacity is required (and, consequently, less time is available for set-up). The two control parameters are used at the short-term production order scheduling level to perform the lot-sizing and sequencing decisions.

In summary, the two levels of our hierarchical planning system are (see figure 2):

- Long-term capacity coordination

---

* The *contribution* is the net contribution for each product, i.e. the selling price of the product minus the variable costs (mainly material cost and energy cost)
The long-term capacity coordination function takes a decision on the level of the control parameters. This decision is made by optimizing an aggregate model of the production system (objective: maximization of profit). This function will be discussed in Section 3 of this paper. The short-term production order scheduling function determines when to produce which products for how long, using the system parameter settings determined at the long-term level. This function will be discussed in Section 4 of this paper.

3. Long-term Capacity Coordination

In this section, the problem will be formulated as a nonlinear programming problem. This mathematical description of the problem is a formulation of the aggregate model. The long-term capacity coordination function should balance the control parameters. Since these parameters determine the performance of the production system, their settings can be evaluated and optimized in an aggregate model.

The control parameters are, for each product \( i \), the target inventory level \( (I_i^*) \) and the integer multiples \( k_i \). The resulting (target) cycle time is the result of these two parameters. This relation is expressed in equation 1.
The target cycle time $T_i^* \overset{\text{target cycle time}}{=} k_i \sum_{j=1}^{n} \left\{ \frac{1}{k_i} I_j^* - \alpha_j \sqrt{T_j^*} \left( \frac{d_j T_j^* - I_j^*}{\sigma_j \sqrt{T_j^*}} \right) \right\} + \frac{c_j}{k_j} \right\}$

where $d_i = \text{average demand for product } i \text{ per period}$

$\sigma_i = \text{standard deviation of demand for product } i \text{ per period}$

$c_i = \text{set-up time for product } i \text{ (periods per set-up)}$

$p_i = \text{production rate of product } i \text{ (units per period)}$

$E(.) = \text{partial expectation function according to Brown (1963)}$

The target cycle time $T_i^*$ is an integer multiple ($k_i$) of a basic cycle time $T^*$. The target cycle time is comprised of the production runs of each product (rationed according to the respective $k_i$'s) and the set-up time of each product (rationed in the same way). The length of the production run is determined by the difference between target inventory level and actual inventory level at the start of a new production run. The expected inventory level at the start of a new production run is a function of the target inventory level, the target cycle time and the demand rate of a product. This function is based on the partial expectation function (Brown 1963, p. 371). Note that equation 1 controls the allocation of capacity. Observed from the aggregate level, the selection of the target cycle times ensures that the required service level will be met. Of course, on disaggregation some infeasibilities may occur. These are not accounted for in the aggregate model, because we do not want this model to be too detailed. Its level of detail should be related to the organizational level taking this decision. This issue will be discussed later in this paper.

Given the target inventory levels, the expected service levels can be determined. The service level is defined as the fill rate, i.e. the portion of demand that is filled out of stock. Note that any demand that is not filled out of stock gets lost (no backordering). We define the expected fill rate of a product $i$ as the fill rate that is expected for product $i$, given the characteristics of demand (mean and variance), the target cycle times, and the target inventory levels. An expression for the expected fill rate ($EFR_i$) is presented in equation 2.

Equation 2 shows that the expected fill rate is computed by deducting the portion of
demand that will not be delivered (on average) from 1. If we know the expected fill rate, the target inventory level, and the target cycle time, we can determine the expected profit per period. The expected profit consists of the expected contribution, the expected inventory holding cost, and the expected set-up cost. These are represented in equation 3.

\[
\text{EFR}_i = \frac{\sigma_i \sqrt{T_i^*} E \left[ \frac{I_i^* - d_i T_i^*}{\sigma_i \sqrt{T_i^*}} \right]}{d_i T_i^*} = 1 - \frac{\sigma_i}{d_i \sqrt{T_i^*} E} \left[ \frac{I_i^* - d_i T_i^*}{\sigma_i \sqrt{T_i^*}} \right]
\]

\[\sum_{i=1}^{n} \left( b_i d_i \text{EFR}_i - \frac{1}{2} I_i^* \left( 1 - \frac{d_i}{p_i} \right) + \sigma_i \sqrt{T_i^*} E \left[ \frac{d_i T_i^* - I_i^*}{\sigma_i \sqrt{T_i^*}} \right] h_i - \frac{u_i}{T_i^*} \right) \]

where

- \( b_i \) = contribution margin per unit of \( i \)
- \( u_i \) = set-up cost

It is obvious that the portion of demand that is expected to be filled determines the contribution (first term of the equation). We estimate the average inventory as the mean of the highest and the lowest expected inventory positions. The highest inventory position is the target inventory level, reduced by the demand during production. The lowest inventory position is the inventory level at the end of the cycle. Finally, the set-up costs are directly related to the target cycle time (which is proportional to the reciprocal of the set-up frequency).

The set of equations mentioned above describes the operational behavior of the system in aggregate terms. The parameters to be influenced at this level are the target inventory level \( I_i^* \) and the integer multiples \( k_i \). If these are set, then the target cycle time is determined according to equation 1. Maximization of equation 3 as a function of \( I_i^* \) and \( k_i \) (replace \( T_i^* \) in equation 3 by equation 1, and \( \text{EFR}_i \) in equation 3 by equation 2) should result in the approximate optimal setting of the parameters. Obviously, the objective function is very complex. The function is non-linear in its decision variables \( I_i^* \)

8
and $k_p$, and also the interaction between each of the product cycle times and inventory levels is very complex.

The objective function is subject to a service level constraint. It is important that some management-specified minimum service level is guaranteed for each product. This decision should be taken at the strategic decision level and should be a constraint for the tactical decision described in this chapter. A minimum service level is realistic from a business point of view. If it did not exist, it might be possible that for some product(s) – especially those with a small contribution margin – no demand will be accepted during the year. In that case, it is unlikely that the product would be part of the product mix of this business. In order to obtain a feasible production schedule, it is necessary that the minimum service level can be met within the capacity constraints. On the other hand, a maximum service level is necessary to account for realistic planning. These service level constraints are represented in equation 4.

$$a_{ji} \leq EFR_i \leq a_{2i} \tag{4}$$

where $a_{ji} =$ predefined minimum service level for product $i$.
$a_{2i} =$ predefined maximum service level for product $i$.

The problem of maximizing (3) subject to (4) will be called the capacity coordination problem. The capacity coordination problem is aimed at the determination of the cycle times and the distribution of the available capacity over the product range. The model considers the cost structure of the different products, the capacity they consume, and the demand distribution of each of the products. The model also takes into account the differences in service levels due to the fact of some $I_i^*$'s being larger than some $d_i T_i^*$'s.

A heuristic approach has been developed to find a solution to the capacity coordination problem. In this paper, we will focus on the hierarchical structure of the model. The details of the heuristic and results of simulation experiments may be found in Fransoo et al. (1993).
4. Short-term Production Order Scheduling

The short-term production order scheduling function consists of two decisions: the sequencing decision and the lot-sizing decision. Both decisions are administered on-line, at the latest moment possible, i.e., the moment when a new run has to be started. Taking the decision earlier would be less advantageous, since less information would be available. Also, the decision space would not be expanded by taking the decision earlier, i.e., scheduling freedom is not limited by taking decisions as late as possible.

With regard to the sequencing decision, however, the new information does not lead to a better decision. In Fransoo (1993), it is proven that in case of a single machine, a limited number of products, extremely high and stationary demand and lost sales, it is better to use a fixed sequence than to use a sequencing rule which takes into account more recent information, such as the runout time sequencing rule. Therefore, a fixed sequence is used in this concept.

The lotsizing decision, however, is based upon the most recent information. The idea is that the average lotsize decided upon leads to a stable cycle time. Each time a specific lotsize is set, it is chosen such that the target inventory level be reached, and — consecutively —the target service level be realized during the cycle to come. The lotsize is determined at the start of the production run and equals the target inventory level $I^*_i$ minus the actual inventory level $I_i$.

It is important to note that the operational scheduling function is rather myopic in nature. This can only be allowed if the parameter settings as determined by the upper level can be maintained, since this will guarantee the proper behavior of the system as compared to the long-term objectives.

5. Use of the hierarchy concept in a single-stage system

The model presented in this text is a two-tiered hierarchical model. The hierarchical approach used in the development of this model has some specific characteristics. We refer especially to the setting of the control parameters at the upper level and the more
myopic view at the operational decision level.

In this section we will analyze the hierarchical structure of the model presented in this text. In this analysis, some of the insights provided by Schneeweiß (1993) and Bertrand et al. (1990) will be used.

First, let us consider the top level of the model. The top level decides on the control parameters of the system (target inventory level and target cycle time). There is however a difference in the nature of each of these control parameters. The target inventory level decision is a *final* decision, i.e., it has a direct consequence for the physical production system. The production system always produces up to the target set by the capacity coordination function. The decision on the target cycle time is however not final, since this cycle time will not be realized exactly by the production system. In line with Schneeweiß (1993), we call this decision *factual*, i.e., once the decision is taken by the top level, it is a fact for the bottom level and has to be taken into account in its own decision process.

The top level takes its decision based upon an aggregate model of the bottom level. This aggregate model consists of equations (1) to (4). An aggregate representation may be used, if the steady state performance of the detailed model is the same as the steady state performance of the aggregate representation (Chandy et al. 1975). In the design of the hierarchical procedure presented in this text, this consistency has been tested by a series of simulation experiments. In these experiments, performance measures were evaluated for the aggregate and the detailed models (such as profit, individual product service levels and cycle times).

In the aggregate model, some detailed information is not included. First, there is an aggregation concerning the modeling of time. In the aggregate model, time buckets are not distinguished and only rates (for demand and production) are considered, while in the detailed model at the bottom level time periods are distinguished and individual events (order arrival, setups, etc.) are considered. Secondly — and an extension of the time aggregation —, there is an aggregation of the demand process. The demand is modelled as normally distributed demand over the cycle time. This is more or less independent of the actual demand process. To illustrate this, let us suppose the actual
The demand process has a Poisson distribution. Since only the demand during the cycle time influences the performance of the system, we may use the characteristic that the Poisson distribution— and a number of other distributions — approximates the normal distribution if it is accumulated over a certain time period. Obviously, this aggregation is only valid if the cycle time is sufficiently long to justify this approximation. Thirdly, the interactions between the products are not modeled at the aggregate level. It is assumed that it will be possible to maintain the individual product cycle times at the detailed level.

Since these cycle times are intended to serve as a target for the lower level, this assumption consists of two premises of the bottom level decision procedure. The first premise is that the objective of the bottom level is to minimize the difference between the actual cycle times and the target cycle times (the "objective function" of the bottom level). As mentioned in the previous section, the performance of a number of possible decision procedures has been evaluated in order to determine the decision procedure at the bottom level. This evaluation was based upon the difference between the target (cycle times and service levels) determined by the top level and the actual realization. The second premise of the bottom level is that it accepts lost sales, i.e., demand gets lost if it cannot be delivered from the actual inventory. Backordering of demand is not allowed.

Thusly, in the aggregate model, the lower level decision and the primary process are represented. It is anticipated what will be the expected reaction of the bottom level. All decisions of the bottom level (sequence and lotsize) take effect in the object system and are — by definition — final.

The consistency exposed here is important in the design of the decision procedures. The aggregate model is based upon a number of assumptions about the bottom level behavior. These assumptions are however in the design process used as premises in the design of the detailed model, leading to consistency in the planning and control hierarchy. The use of these consistency requirements in the design process is in line with the methodology presented by Bertrand et al. (1990).

The conceptual aggregation used here is clearly different from the mathematically more transparent aggregation procedures in the well-known approach by Hax and Bitran (e.g., Bitran and Hax 1977). Their approach is characterized by a dual aggregation criterion:
data are aggregated by products and by time. At the item level, individual items are considered for short time periods. At the family level, groups of items are considered for longer periods of time. At the type level, groups of families are considered for even longer periods of time. In our approach, only two levels of aggregation are considered. Basically, the product level of aggregation is the same at both levels. In this paper, the products are considered at the item level. However, if very homogeneous families can be formed (i.e. families which share a common set-up, and no set-up is required between the products of the same family), these families can be considered as well. This requires the addition of an extra level to decide on the scheduling of individual products.

Another important difference between the two models is the impact of the top level decision. In the model by Hax and Bitran, all top level decisions are final. This leads to a very restricted decision space at the bottom level. The advantage is a clear consistency between the two levels. On the other hand, however, a lot of flexibility is lost and the ability of the bottom level to react to new information is decreased. In our model, we maintain the necessary consistency, although the disaggregation is an approximate one. Consequently, some differences exist between the two levels (as outlined above), but the major (steady-state) behavior of both models is consistent.

The choice of a single level of product aggregation is in line with Bitran and Hax's remarks in their paper (1977) and our observation (Fransoo 1993) that in flow process industries set-ups have a primary impact on total production costs. Even more dominant is the impact of set-ups on total production capacity. Bitran and Hax (1977) state that the inclusion of set-up cost forces the model to be formulated at a lower level of aggregation. In Bitran et al. (1981) the problem is captured by introducing an extra feedback loop to correct the type level decision. The feedback loop explicitly considers set-up cost. The approach presented in this text decides directly, at the top level, on cycle times and inventory positions, taking not only the set-up costs into account, but also the set-up time. This results in increased model complexity at the top level. More decision variables are introduced, and the resulting model is not linear. On the other hand, only one time slot – in terms of long-term rates – is considered at this level, which reduces complexity again.

Finally, we would like to discuss the organizational setting. The difference between
mathematical aggregation and conceptual aggregation reflects a different view on production control and operations management. The approach by Hax and Bitran carefully minimizes cost at each of the decision levels, using the decision variables that can be influenced at each of these levels. This reflects a decentralized organizational structure, where each department has to strive for cost minimization within aggregate volume constraints. In this paper we have chosen for an approach which is more explicitly based on an integral model of the system dynamics. We have studied the behavior of the system to be controlled and we have noted the central parameters which determine the behavior of the system. These parameters have appeared to be the cycle time and the target inventory level. These control parameters have to be set for a longer period of time, in order to reach stability in the system and consequently improve operational control. Our design also leads to a closer cooperation between the production and sales departments, since it makes the relationship between these two departments more explicit. This is possible in this environment, because of the well defined capacity constraints in production. It remains to be seen whether a joint decision procedure for production and sales at the tactical decision level makes sense if these constraints are less clear than in the problem considered in this study.

Translated into an organizational setting, this means that management determines the required long-term results of the business (in the annual production and sales plans), which are transformed into settings of the logistics parameters. These parameters are then used for operational production control. This is also reflected in the two-tiered theoretical model presented in this text. If the departments at the operational levels maintain these parameter settings, the long term results can be achieved. Clearly, no optimization takes place any more at the operational level. This should account for much more stability in the operational schedules. The purpose of the operational decision level is not cost minimization, but maintaining control over the system.

Acknowledgment
Sincere appreciation went to Ch. Schneeweiss and the ESI IX participants, since the discussions in Mannheim considerably improved the quality and scope of this paper.

References


