On the capacity-aspect of inventories

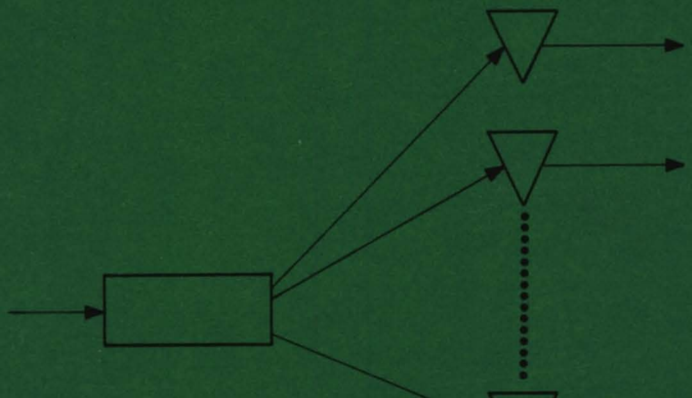
DOI:
10.6100/IR178916

Document status and date:
Published: 01/01/1985

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

ON THE

CAPACITY-ASPECT OF INVENTORIES



R.P.H.G. BEMELMANS

ON THE CAPACITY-ASPECT OF INVENTORIES

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR IN DE TECHNISCHE WETENSCHAPPEN AAN DE TECHNISCHE HOGESCHOOL EINDHOVEN, OP GEZAG VAN DE RECTOR MAGNIFICUS, PROF. DR. S. T. M. ACKERMANS, VOOR EEN COMMISSIE AANGEWEZEN DOOR HET COLLEGE VAN DEKANEN IN HET OPENBAAR TE VERDEDIGEN OP DINSDAG 28 MEI 1985 TE 16.00 UUR

DOOR

ROLAND PIETER HUBERTUS GERARDUS BEMELMANS

GEBOREN TE MHEER

Druk: Dissertatiedrukkerij Wibro, Helmond.

Dit proefschrift is goedgekeurd
door de promotoren

Prof. Dr. J. Wijngaard

en

Prof. Dr. J. Wessels

aan Mieke
aan mijn ouders

Acknowledgements.

The research for preparing this text has been carried out under the supervision of Prof. dr. J. Wijngaard. The author is much indebted to him for his inspiration and research guidance. The author also expresses his gratitude to Dr. Attwood for his many suggestions with respect to the use of English in this text. Finally the author wishes to thank W. Tjin for producing the drawings in this text.

Contents.

Chapter 1. Scope of the text.

1.1 Introduction.

Controlling the production in an industrial organisation is very complex. There are two different reasons for this complexity. On the one hand, complexity is due to the variety in range and in level of detail of the activities that play a role in such a control (think of manufacturing process development, capacity planning, coordinating the flow of material through the production process, releasing of workorders, and scheduling). On the other hand, the production process itself may be complex (many products, many stages, complex interrelationships between resources, and uncertainty in the availability of resources).
To deal with the first cause for complexity, one creates different, but coordinated levels of control. At each of these levels a specific part of the control of the production process is accounted for (see Anthony [3]). To deal with the second cause for complexity, one groups manufacturing steps into so-called Production Units (see Bertrand [8]). Each Production Unit is responsible for a specific part of the production process. Of course, these Production Units have to be coordinated to ensure that the products are manufactured timely and efficiently. This activity will be referred to as Material Coordination (see Bertrand [8]).

In Chapter 2, we will discuss this decomposition approach in more detail. Material Coordination will be part of such a decomposition approach to Production Control. On the level of Material Coordination, different Production Units are discerned in the process and there is a flow of material over these Production Units. It is the task of

Material Coordination to coordinate the activities of the different Production Units in order to realise a given delivery performance target, like minimizing the number of stock-outs. Material Coordination, thus, does not influence the demand or influence the resource availability, but has to reach certain performance targets for a given demand and with a given resource-availability. In the next Chapter, we will go further into Material Coordination, and discuss its relationship to other parts of Production Control.
To have an idea of the place that Material Coordination takes within a framework for Production Control, one can think of existing Material Coordination Systems, like Material Requirements Planning, the Reorder Point System or the Base Stock System.

1.2 Topic of the text.

The way Material Coordination can deal with uncertainty, is important. The following types of uncertainty can be distinguished:

- uncertainty in the availability of raw materials
- uncertainty in the behaviour of the resources
- uncertainty in the actual delivery pattern that will be required
- uncertainty in the registration of inventories and work-in-process.

Since, for Material Coordination, it is not possible to influence the required delivery pattern or the availability of resources, other methods must be used to protect against these uncertainties.
Some of these uncertainties may be due to inadequate information processing capabilities. If the information processing system can be improved without much effort, it will be sensible to do so. However, in general, it will be impossible (or much too expensive) to remove all the uncertainty. In order to be able to cope with the remaining uncertainty, it is necessary to create stocks (note that also in case there is no uncertainty, stocks may be created, for example due to

For the design of a good Material Coordination System, it is important to have insight into how to use these stocks efficiently. This insight will also enable us to make a trade-off between investments to reduce uncertainties (e.g. information processing systems) and investments to cope with uncertainties (e.g. inventories, work-in-process, or flexible resources).

Whybark and Williams [54] have shown that the control of buffer stocks should be adjusted to the sources of uncertainty. Therefore, let us first consider these sources of uncertainty. There are fundamentally two different types of uncertainty: there is uncertainty due to the behaviour of individual products (e.g. uncertain demand, inventory registration or yield of the production process) and there is uncertainty due to the behaviour of resources (e.g. worker-availability or machine break-down).

Consequently, two fundamentally different approaches to using buffer stocks can be distinguished, namely a product-oriented approach and a capacity-oriented approach.

In the product-oriented approach, the buffer stocks are based on the behaviour of individual products. The delivery pattern is translated to a production pattern by Material Coordination, using standard throughput-times for orders. The production patterns for the products are coordinated over a short horizon. A well-known example of the product-oriented approach is Material Requirements Planning. In this approach, a certain inventory is created for each product to protect it against uncertainties. This can be done by hedging the demand (i.e. systematically over-estimating the demand), by using a safety lead-time or by using a fixed safety stock for each product. Note that the stocks created in the product-oriented approach, also have to protect Material Coordination against uncertainty due to the behaviour of resources.

In the capacity-oriented approach, the accumulation of buffer stocks is based on a comparison of demand and availability of capacity, which means that the inventory of different products is no longer viewed in the first place as a buffer against uncertainties in the behaviour of that product. Instead, in this approach, the inventories are viewed as a form of stored capacity. In case the demand for capacity during periods in the future is larger than the available capacity, this stored capacity will be used to solve the problem. In the capacity-oriented approach we aggregate over the individual inventories of the products to find a measure for the amount of stored capacity in the inventories. Thus, an aggregate production pattern is generated. This production pattern is disaggregated over a shorter horizon. Note that the stocks created in the capacity-oriented approach also have to protect Material Coordination against uncertainties that are due to the behaviour of individual products.

These two approaches differ fundamentally, but both yield a feasible Material Coordination System. The question is which approach should be used in what type of situation. The objective of this text is to provide the reader with insight into which characteristics of the situation are important for making this decision and thus to provide a tool for deciding in a given situation, which of the two approaches is best.

1.3 The approach.

As we have seen in the previous Sections, the aim of this text is to suggest when to use a product-oriented approach and when to use a capacity-oriented approach to designing a Material Coordination System.

The problem is studied in this text via the systematic analysis of simple, but relevant models. In this research, we have restricted ourselves to models of the single-phase type, which means that there is a single capacity bottle-neck in the production process and there are many products. The reason for using these models is that they are the best starting point to compare the capacity-oriented and the product-oriented approaches. We will not go into multi-phase situations, since more research is needed for these situations. However, in situations with only one bottle-neck, these results will help the reader to choose an adequate approach to design a Material Coordination System.
We start by discussing a fairly simple single-phase multi-product planning problem. Then, we will introduce more and more aspects that can play a role. For each model, we will formulate both the capacity-oriented and the product-oriented approach and compare their

performance. This performance evaluation is mainly done by simulation experiments.

1.4 Review of the text.

In Chapter 2, we will describe the place that Material Coordination has within a general framework for Production Control. We will also show why we have chosen to investigate the single-phase multi-product problem. Related literature to this Chapter is Anthony [3] and Galbraith [23].

In Chapter 3, we will consider a simple single-phase multi-product planning problem with identical products and stochastic demand. A review of single-phase models has been presented by Elmaghraby [19]. However, mostly deterministic models have been considered in the literature. An exception to this is the work of Graves [24], Williams [55] and Zipkin [61].
For the single-phase model in Chapter 3, we will compare the performance of capacity-oriented and product-oriented strategies, when confronted with uncertainty with respect to availability of the resource and with respect to demand.

In Chapter 4, we consider a model in which demand is partly known beforehand. Thus, a forecast for demand of each product is available. Difficulties arise since different forecasts for future demand make the products not-identical in the short-term. The question whether to use a capacity-oriented approach or a product-oriented approach is intertwined with how the forecast is used.

In Chapter 5, we will describe a model with non-identical products. In this model, there are obvious slow-movers and obvious fast-movers, and the difference between them is no longer only caused by short-term forecasts, but there are big differences between them in the long-run as well. This introduces new problems, since the capacity-oriented approach has to be restricted to fastmovers.

To show more clearly how the results obtained, can be used, we will include a simple example of a plastic products factory in Chapter 6. For this factory, we will describe a framework for Production Control and we will show how a Material Coordination System for this situation can be designed, based on the results of this text.

Chapter 2. Material Coordination.

2.1 Production Control.

When controlling an industrial organisation, all kinds of activities have to be considered. For example the following activities should be part of control:
budgeting decisions, scheduling decisions, release of actual workorders, selection of suppliers, marketing, financial planning, decisions on workforce levels.
In order to create some order in this range of activities, one has to distinguish several separate control processes. Each of these control processes is directed at a specific part of the control of the organisation, whereas it must be possible to coordinate the separate processes in order to gain control over the whole organisation. Common processes that can be distinguished are (see Burbidge [15]):

. Sales Control
. Production Control
. Purchase Control
. Financial Control
. Quality Control

In this text, we will consider Production Control. One way to define Production Control is (see e.g. Greene [25] and Bertrand and Wortmann [9]):

"The Production Control function is defined as the set of activities in a production organisation that are directed at the control of volumes and types of products produced at specific places as a function of time"

According to Bertrand and Wortmann [9] this means that Production Control includes long-range planning, product-development, manufacturing process development, customer service control, factory lay-out planning, transportation and physical distribution, manpower planning, material supplies control and materials handling, capacity planning, scheduling, loading, dispatching and expediting, and inventory control.

At a high level of the organisation, Production Control is integrated with the other control processes. For example long-range planning for Production Control has to be combined with

- -long-range sales planning in order to ensure that the production activities comply with the marketing activities.
- -long-range purchase planning in order to ensure that the timely supply of raw materials is possible.
- -long-range quality planning in order to ensure that the quality remains within certain limits.
- -long-range financial planning in order to ensure that the capital necessary for realizing the plan, is acquired at the right time.

The reason to distinguish these different control processes is that they are relatively independent. It is possible to reduce the interference between these control processes to simple relationships (e.g. by a budgetary system). Slack is required to reduce this interference (compare Galbraith [23]). The main benefit of investing in this slack is that each separate control process becomes easily understandable, which in general leads to a better control of the organisation.

We will restrict ourselves to Production Control. We will not discuss the question of how to create slack efficiently in order to make Production Control independent from the other control processes, but we will just assume that the interference has been reduced in some way. The reason why we will not go into this any further, is not that we believe that the problem of creating slack between the control processes is relatively simple or unimportant. On the contrary, there is still a lot of work to be done in this field and the importance is obvious. However, to keep the research that was needed for preparing this text, manageable, we have restricted ourselves to Production Control (even to a specific part of Production Control, but we will return to that in the next Sections). We believe that, before discussing efficient ways to invest in slack between different control processes, it is necessary to have a good insight into the performance of each individual control process.

2.2 Reduction of control complexity.

Production Control, as described in the previous Section, is still very complex.
The first cause for this complexity becomes clear when we consider the list of activities that are part of it (mentioned in the previous Section). There is a big difference in the range and the level of detail between the activities. Yet, there are clear relationships between different activities, that make coordination necessary. The usual way to attack this problem is to create different "levels of control", each with its own details and range of decisions. Each level is then considered to be relatively independent, as the interference between different levels is reduced to a simple one, e.g. by generating goals and restrictions. This requires investment in slack at each level in order to be able to separate it from other levels. We will discuss the idea of levels of control more deeply in Subsection 2.2.2.
A completely different cause for the complexity of Production Control may be that there are many products in various stages of progress, complex interrelationships between resource restrictions and much uncertainty with respect to the availability of these resources. In order to reduce the complexity of the production process, "Production Units" are created. These Production Units are comparatively independent and only simple methods for coordinating them will be permitted. Of course, again, this requires an investment in slack

within the Production Units. We will return to this subject in the next Subsection.

Before discussing both of the methods to reduce the complexity of Production Control (namely the creation of levels of control and Production Units), we must mention that they are interrelated. When detailed production plans for the near future (at a lower level of control) are being considered, it is necessary to have some insight into the way that a given Production Unit functions, whereas it is sufficient to have a rough concept of the Production Unit if a long-term plan for Production Control has to be determined.

2.2.1 Production Units.

To simplify Production Control, several manufacturing steps and resources are grouped into so-called Production Units.
The aim of creating these Production Units is to reduce the complexity of Production Control. Therefore, the following conditions have to be taken into account:

On the one hand, the control of each of the Production Units has to be relatively simple. This requires a stable environment and stable operational norms for the Production Units. If this stability is not implied in the process the Production Units are imbedded in, it will be required to invest in slack between the Production Units in order to guarantuee this stability.

On the other hand, the coordination over Production Units has to be simple too. For this coordination the Production Units are considered as black boxes with simple production characteristics. The model of the production process, in which the Production Units are treated as black boxes is referred to as the aggregate process. This aggregate process then must be easy to control.
Notice that the analysis of the aggregate process only aims at setting objectives for the Production Units in order to ensure coordination

(like setting due dates for work-orders), but it does not solve all the problems for the Production Units in detail. It is left to the Production Units to solve these detailed problems (scheduling, loading, etc.). In order to be able to leave the solution of these detailed problems to the Production Units, when analyzing the aggregate process, it is necessary to invest in some slack and flexibility within the Production Units.

As an example of the creation of Production Units, we will describe the model that was considered by Bitran and von Ellenrieder [12]. Note that this model will only be used as a point of reference for discussing different aspects of Production Control. Therefore the reader does not need a thorough understanding of the model in order to read the rest of this text.
In Bitran and von Ellenrieder [12], a firm was considered that manufactured castings and nipples for use in the construction industry. The number of different products that were produced and sold, was about 1200.

In Figure 2.1, the production process has been shown as a diagram.
In the first stage of the production process, the cores are prepared in two parallel stages. These cores are stored and used for assembling the casts, which are prepared by "moulds preparation" and are sent to the third stage, the melting of one of three ferrous alloys. The molten material is prepared in three batteries of electric furnaces. In this way, for each battery of furnaces a reserve supply is provided.
From these supplies, the items are passed through a furnace for annealing and grain allignment (heat treatment). In the gauging stage of the process the finishing operations take place that create the last significant intermediate stock of products. Sometimes, items are dispatched to the customers directly from this stock and sometimes they are submitted to some additional process.
As will be clear, this is a complex process and it would be difficult to control it without structuring the production process first.
Therefore Bitran and von Ellenrieder aggregated over some manufacturing steps and thus constructed the "aggregate process" as in Figure 2.2.

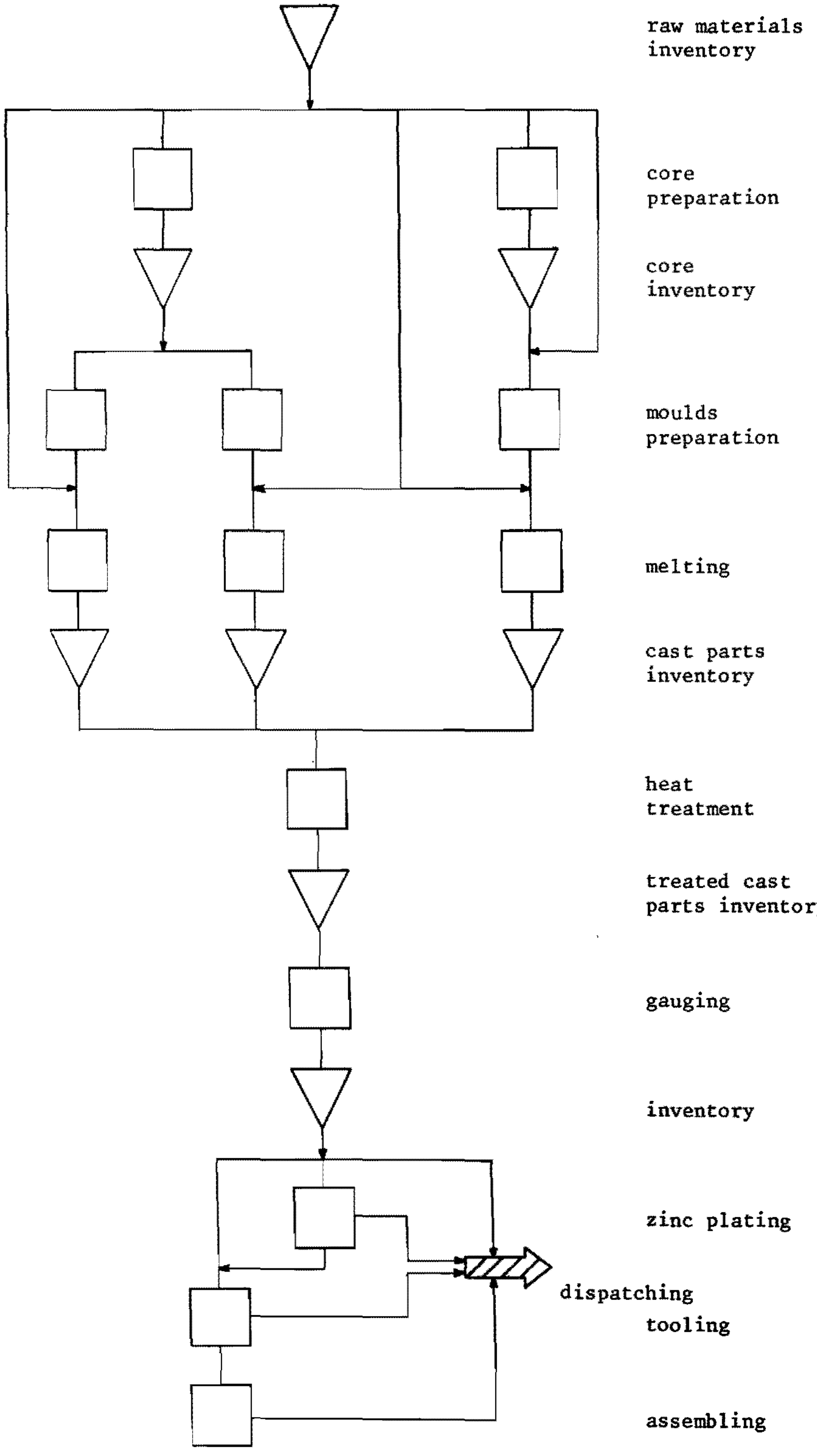


Figure 2.1. Flowchart of the production process for the castings and nipples manufacturing.

The reason for making this particular division into Production Units, was that heat treatment was one of the most complicated stages in the production process (from a planning point of view), because of the large variety of types and sizes that have to be dealt with.

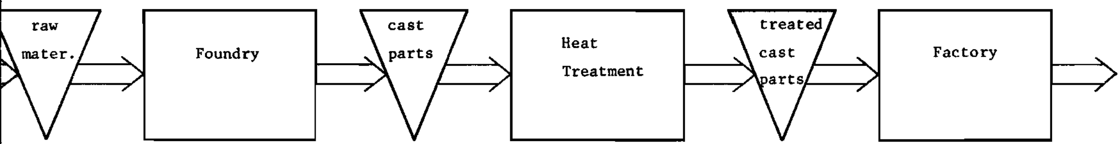


Figure 2.2. Aggregate process for the castings and nipples manufacturing (see Figure 2.1).

Notice that the introduction of Production Units decreases the decision freedom. This effect has to be compensated by the fact that, due to a reduction of the complextity, the control can be improved (see Bertrand [8]).

2.2.2 Levels of Control.

We group the decisions into decision levels. The most important reason for doing so, is that consequences of decisions are so different that a monolithic approach is impossible. Of course, if this were not the case, a hierarchical approach might still be preferable because of its relative simplicity: we want a simple structure for taking planning decisions in order to make an easy coordination possible with the other control processes in the organisation (think of budgets, objectives and production levels).

This grouping of the decisions leads to a so-called hierarchy of planning decisions. Roughly, one can distinguish three levels in such a hierarchy (see Figure 2.3). This distinction presents us with a natural framework for planning and control in practical situations (see e.g. Anthony [3], Bitran and Hax [13], Jönsson [31], Manz [38]), although a too rigid classification into exactly three levels will certainly not always be right. We will discuss each of the levels in some detail. The discussion of each level starts with the definition given by Anthony [3], who (to our knowledge) was the first to formulate such a framework in a systematic way.

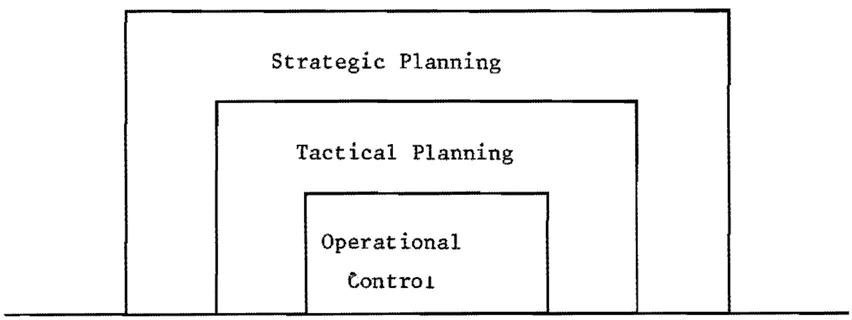


Figure 2.3. A planning hierarchy.

Strategic planning is "the process of deciding on objectives of the organisation, on changes in the objectives, on the resources used to attain these objectives, and on the policies that are to govern the acquisition, use and disposition of these resources".
For example, a typical decision that should be taken on this level is whether to enter the market with a completely new type of product. This requires large investments in the design of new production facilities or even building new plants. Such decisions obviously interfere with other control processes in the organisation, like Sales Control, which has to estimate the possibilities of the new market, and Financial Control, in order to acquire the capital that is needed.

The different control processes are balanced in outline on this level. This requires a high degree of aggregation. Another reason for using a high degree of aggregation on this level, is the following: since the decisions on this level have long-lasting effects on the organisation, it is necessary to have a long planning horizon (about two to five years). The information that is available on this term is often only

qualitative or characterized by a great deal of uncertainty. To be able to take realistic decisions on this level, it is necessary to consider aggregate quantities.
The outcomes of the decisions on this level often have a large and long-lasting effect on the behaviour of the organisation and therefore require the attention of top management.

Tactical planning (or management control) is "the process by which managers assure that resources are obtained and used effectively and efficiently in the accomplishment of the organisation's objectives". Before discussing this level, we should first mention that this level is known under two different names in the planning and control literature, namely management control and tactical planning. Originally, Anthony [3] used the term management control, but later, others preferred the term tactical planning (see e.g. Ackoff [1] and Hax and Meal [28]). We believe that the latter term is more common in recent literature and, therefore, we will use it in this text too.

On the tactical level, one must use certain prescribed facilities to attain the objectives that have already been set by the strategic level. Looking at the example in Section 2.2.1, typical activities that fall under this heading include the replacement of electric furnaces, the decision to start using a fourth alloy that only differs slightly from the existing ones, the make or buy decisions for cast parts, the trade-off between customer service rate and inventory levels and setting work force levels in each Production Unit.
Often the planning period for this level is about one year and this reduces much of the uncertainty of the strategic level, where the planning period is much longer. Consequently, the tactical level can react more efficiently to later developments. Therefore the strategic level must retain some slack for the tactical level in order for this level to be able to react to uncertainty of the environment. The tactical level must, in turn, formulate guidelines for operational control.

An important point is the interference with Sales Control on this level. As we have seen, both control processes have already been coordinated on the strategic level. On the tactical level, coordination for a shorter period is considered. We have already mentioned the trade-off between customer service rate and inventory levels, but this is not a matter for Production Control only. Sales Control and Financial Control should take part in this particular decision, because (usually) there are conflicting interests between different control processes at this point. Production Control requires a stable production situation and does not like to be disturbed by an unpredictable, fluctuating demand. Sales Control, however, wants to provide a good customer service rate and therefore requires more flexibility of Production Control in the short term, no matter what investments in slack (inventories, excess of capacities) are needed therefore. Financial Control wants to keep the required capital (that has been tied up in e.g. inventories) within certain limits. These conflicting interests have led, in many organisations, to the formulation of lateral relations (see Galbraith [23]). Bertrand and Wijngaard [10] distinguish structural and operational coordination within this context. Structural coordination implies aggregate agreements with respect to delivery performance and sales patterns. Operational coordination takes the actual status of production, sales and finance into account.

Notice that the structural coordination falls under the heading of tactical planning, since it ensures that the resources are used effectively and efficiently without actually being concerned with specific tasks. The operational coordination falls rather under the heading of operational control, which we discuss below.

The degree to which coordination has to be structural or operational, depends on the particular production situation. The outcome of this coordination, will be referred to as the Master Production Schedule. This Master Production Schedule should be a (normative) statement of production, sales and finance.

<u>Operational control</u> is "the process of assuring that specific tasks are carried out effectively and efficiently".

On the operational control level the daily actions have to be coordinated. The aim is no longer to set budgets for inventory but to actually control the inventories, no longer do we set the workforce levels but actual hiring and firing takes place, as the situation requires.

In the example of Section 2.2.1, this level is responsible for the cast parts stock being sufficiently high to ensure that heat treatment can do its work, also scheduling of jobs in the foundry is a task of operational control, as is daily allocation of workmen to the machines. On this level it also has to be ensured that the flow of material over the Production Units is coordinated to guarantee a certain performance rate to the customers.
As we see this level of planning has a short planning period (say a few weeks) and it must come up with detailed proposals for action.
The interference on this level with other control processes is rather limited, the coordination has taken place on a higher level and now the commitments on the higher levels have to be realised.

2.3 Introduction of Material Coordination.

In this text, we focus on the part of Production Control that consists of coordinating the flow of material over the Production Units. This task will be referred to by the term "Material Coordination", a term that is proposed by Bertrand [8].

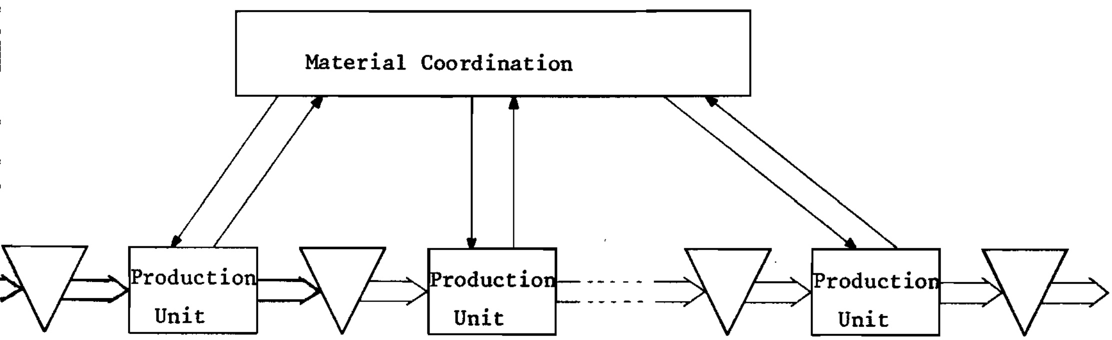


Figure 2.4. Material Coordination.

On this level of control the availability of resources as well as the "demand" can (generally) no longer be influenced. This demand may be the outcome of some balancing between control processes as formulated in a Master Production Schedule and therefore it need not be the "customer demand". However, since Material Coordination has no control over the Master Production Schedule, we will refer to the Master Production Schedule as the demand for Material Coordination.

The task of Material Coordination is to coordinate the activities of the Production Units in order to realise the commitments that have been made on the tactical level with respect to customer service rate, inventory budgets and workforce levels. So, Material Coordination is not involved in the trade-off between different performance criteria, but has to take necessary actions to reach the given performance targets. Typical tasks that belong to Material Coordination are setting due dates for orders, ensuring material being available and controlling the inventories.
Because of the nature of Material Coordination, it is seen as a part of the operational control level.
In the next Section, we will illustrate the concept of Material Coordination by describing some well-known examples of Material Coordination.

2.4 Some well-known examples of Material Coordination.

In this Section, we want to elucidate the concept of Material Coordination by describing some well-known examples of Material Coordination. The examples that we restrict ourselves to in this Section, are:

- the Reorder Point System
- the Base Stock System
- Material Requirements Planning

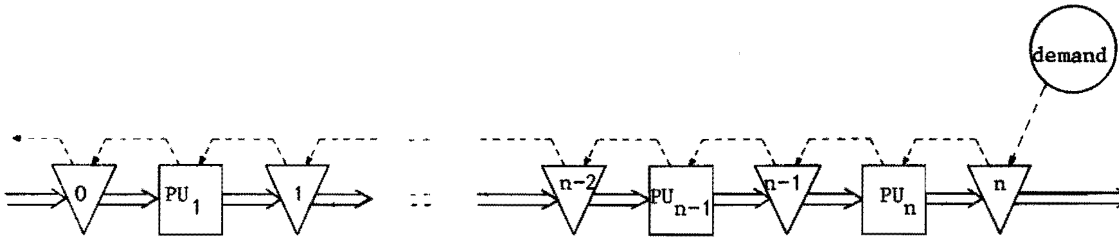
2.4.1 The Reorder Point System.

For an extensive study of this approach the reader is referred to Hadley and Whitin [27]. We will use their notations in this Section.

Let us first consider a single Production Unit.
In the Reorder Point approach a replenishment order for a product is released if the inventory position of that product is below a predetermined, critical level (with the inventory position we mean the inventory on hand minus back-orders plus outstanding replenishment orders). This critical level is determined on the basis of the distribution of the demand over the production leadtime and on the performance criterium that is used (e.g. minimizing inventory holding costs and stock-out costs over time). Depending on whether the inventory is reviewed periodically or continuously, this level (reorder point) is denoted by "T", respectively "r".
Just as important as the question when to produce, is the question how much to produce. Therefore, together with a critical level a production quantity is determined on the basis of the mean and the lumpiness of the demand so as to optimize some performance criterium, like the expected number of stock-outs. In the Reorder Point approach one usually produces a fixed batch "Q", or one replenishes the inventory to a fixed level "R".
Combination of both leads to the familiar Reorder Point strategies: <R,r>, <Q,r>, <R,T> and <Q,T>.

Now consider a production process with several Production Units and see how to use the Reorder Point approach then.

The philosophy of the Reorder Point approach is as follows: The Master Production Schedule (which in case of a Reorder Point System usually conforms to customer demand) is satisfied from stock-point n (see Figure 2.5). For each product critical levels are set as above. If the inventory for a product drops below this level, then an order is released to Production Unit n for replenishment. After such an order is



PU = Production Unit

⟹ = Goodsflow

--→ = Information flow

Figure 2.5. Reorder Point System.

released, Production Unit n receives its "raw materials" from stock-point n-1 and manufactures these to end-items for stock-point n. This leads to a reduction of the inventories at stock-point n-1. In the Reorder Point approach, this reduction is observed as independent demand for stock-point n-1. Based on the characteristics of this demand again critical levels are set for the inventories at stock-point n-1. For the control of the inventory at stock-point n-1, Production Unit n-1 receives raw materials from stock-point n-2, which leads to an "independent" demand at stock-point n-2, etc.

The Reorder Point System has some disadvantages, namely:

1. The production leadtime of a Production Unit is assumed to be independent of the release of replenishment orders. However, the actual production leadtime will depend on the Work-In-Process in the Production Units. Thus different products interfere with each other. Since this Work-In-Process fluctuates widely due to the lumpiness of the demand (and the effects this has on the release of production orders) it will usually be difficult to give a good estimate for the leadtime.

2. Each Production Unit buffers demand until the inventory position drops below the reorder point before passing the demand to the preceding Production Unit. This leads to a delay of information about demand (see e.g. Forrester [21] and van Aken [2]). Even if there are only gradual changes in the demand process, the delay

of information may have large consequences. For example, if demand increases, then Production Unit n will, after some time, adapt its reorder levels. Production Unit n-1 notices this change in the demand process only after a longer time period. Thus it reacts much later on a change in the demand process. This, however, also means that the Production Unit does not only have to keep pace with the new demand process, but it will be necessary to over-react in the short term. This over-reacting is necessary, since for some time the production has been systematically less than the demand.

3. Increasing variability of demand. Another aspect of buffering demand, is that demand appears to be more lumpy if one goes further back in the production process.

The disadvantages 2 and 3 are a consequence of the fact that the decrease of the inventories in each stock-point is seen as independent demand, while there are obvious dependencies, between demand in different stock-points. Van Dierdonck and Bruggeman [17] describe this as a lack of vertical integration, i.e. integration between the control of subsequent manufacturing stages.
The first disadvantage is due to lack of "horizontal" integration (see van Dierdonck and Bruggeman [17]), i.e. integration between different products at the same manufacturing stage.

An obvious advantage of the Reorder Point System is its simplicity, which makes it easy to implement. Only a straightforward flow of information is necessary, which means that there is no need for a large investment in information processing systems. Therefore, this approach is often used, especially for "cheap" products.

2.4.2 The Base Stock System.

In the Base Stock System, the idea of dependent demand is used to make a better coordination between Production Units possible, which leads to vertical integration (see Figure 2.6).

For an extensive study of the Base Stock System, the reader is referred to Kimball [34], Magee [37], and Timmer et al. [51].

The information about demand is not only used to control the inventories in stock-point n, but it is exploded to all stages in the production process so that each Production Unit can react on it. At each stock-point certain inventory levels (base stocks) are determined for each product. As in the case of the Reorder Point System, a replenishment order is released to the preceding Production Unit if the inventory position of a given product drops below its level. The big difference with a Reorder Point System is the way demand is experienced in the stock-points. In the Base Stock System, one keeps track of the demand for end-items and explodes this into demand for components. Consequently, there is no delay in information about demand, which leads to smaller investments in safety-stocks.

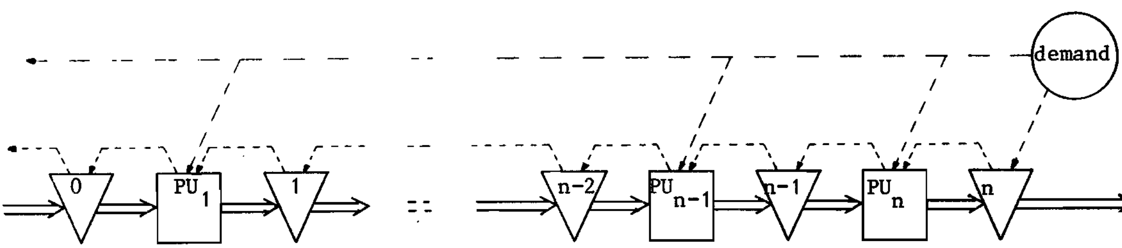


Figure 2.6. Base Stock System.

When translating the demand for stock-point n to demand for stock-point n-k, all inventories in between these stock-points have to be taken into account too. Therefore, the so-called "echelon inventory" is introduced to base the production decision on (compare e.g. Clark and

Scarf [16]). This echelon inventory is the inventory for a given product at the stage where it is produced, and downwards in the production process as it is "assembled" into other products.
Notice that if the production process is divergent, there is a possibility that a component is manufactured into a wrong product (that means that it is manufactured into a product for which no demand has occured, whereas it should have been manufactured into another product for which a stock-out occurs). If one wants to implement a Base Stock System in such a situation, the definition of base stocks has to be adapted. However, in case the production process is convergent, the echelon inventory can be used straight forwardly.

For the Base Stock System, vertical integration is provided for. Consequently, the disadvantages 2 and 3, mentioned for the Reorder Point System, are circumvented. In the Base Stock System, there is still lack of horizontal integration (see disadvantage 1 of the Reorder Point System).
Note that the Base Stock System requires more information processing than the Reorder Point System.

2.4.3 Material Requirements Planning.

In a Material Requirements Planning system the Manufacturing Bill Of Material plays a central role. The Manufacturing Bill of Material describes the product structures from the Material Coordination point of view. Starting from the final products in the Master Production Schedule it is possible to determine what components have to be manufactured in which quantities to assemble the final products. Of course, it is not only required to know how much to produce, but also when to produce. Therefore standard leadtimes are introduced that indicate how long it takes to manufacture the components into the next subassembly.
This leads to the following first step in a Material Requirements Planning system: The Master Production Schedule for the final products

is exploded via the Manufacturing Bill of Material to find a time-phased gross requirement for the components. Any possible independent demand for a component is forecasted and added to this gross requirement. The next step in a Material Requirements Planning system is the so-called "netting procedure". In this netting procedure the gross requirements are converted to net requirements at each production phase on the basis of inventory on hand, the orders that already have been issued and (sometimes) the safety stock. Orlicky [44] gives the following example to illustrate this netting procedure:

Gross requirements		120
On hand	25	
On order	50	
	==	
	75	
Safety stock	-20	
		55
============	======	======
Net requirements		65

After determining the net requirements for each component it is possible to determine "planned orders". Usually some lot-sizing technique is used for this final step in the Material Requirements Planning system.

Of course, we have only given a very rough description of Material Requirements Planning. For a study of Material Requirements Planning, the reader is referred to Orlicky [44]. What we have aimed at, in this Subsection, is to sketch the general idea behind Material Requirements Planning, which is very straightforward. In the APICS News of february 1973, L.J. Burlinger stated that the logic of Material Requirements Planning is inescapable. This seems true, but in order to be able to use this kind of system some important conditions should be met. The most important ones, in our view, are:

- The Master Production Schedule consists of a deterministic requirement for final products and may not be seen as a stochastic variable.
- The resource restrictions may not be tight: It is not clear in the Material Requirements Planning approach how to react if it proves that the released orders cannot be realised.
- It must be possible to keep the production leadtimes constant. Usually these production leadtimes will depend on the Work-In-Process in the Production Units.
- The situation has to be so that safety stocks are only necessary for the Master Production Schedule products. The netting procedure that we have mentioned treats a deplenishment of the component safety stock in the same way as a stock-out for the component.

Consequently, if demand can be forecasted perfectly over the whole production leadtime and if there are no (severe) capacity-restrictions, Material Requirements Planning can be used best. In situations with a stable demand, there will be no advantage of using the Material Requirements approach instead of the Base Stock System. Therefore, Material Requirements Planning is often used in situations with a highly variable demand. Notice that in situations, where the conditions for applying Material Requirements Planning are met, it in fact corresponds to a very powerful information processing system.
In other situations Material Requirements Planning is often used too. Its performance then relies on the ability to make a realistic Master Production Schedule, and on the possibility to react on exception messages (rescheduling).

Notice that for all Material Coordination Systems described in this Section, production runs are started on the basis of information about individual products. In situations with a tight capacity restriction (or more generally in situations where horizontal integration plays an important role), this approach may give poor results (we will return to this in the next Section).

2.5 The role of stocks.

In the previous Section, we have described some well-known examples of Material Coordination. How well a given Material Coordination System works, depends not only on the characteristics of the Material Coordination System, but also on the characteristics of the environment (like how stochastic is demand, how uncertain is the availability of the resources, etc.).

Galbraith [23] has put forward that "the ability of an organisation to successfully coordinate the activities by goal setting, hierarchy and rules depends on the combination of the frequency of exceptions and the capacity of the hierarchy to handle them". Consequently, for Material Coordination, a trade-off has to be made between investments that are necessary to reduce the uncertainty and investments to be able to cope with existing uncertainty. In order to reduce uncertainty, investments are required in information processing systems or in lateral relations. In order to be able to cope with existing uncertainty, Material Coordination is provided with flexible resources or Material Coordination creates safety stocks.
In this text, we want to gain insight into efficient ways to create safety stocks on the level of Material Coordination in order to be able to cope with uncertainty. The results of this text may then be used in making this more general trade-off.

When we want to investigate efficient ways to create safety stocks, it is interesting to consider the way that such stocks are created in the Material Requirements Planning approach. Material Requirements Planning supports three fundamentally different ways to create safety stocks (compare Whybark and Williams [54]):

1. safety stock per product: a production run for a product is started as soon as the inventory drops below the safety stock.

2. safety leadtime per product: a larger leadtime than necessary is used in the planning.

3. hedging the demand: demand is systematically over-estimated.

Whybark and Williams [54] have compared the first two methods. Their main conclusion is that the way to buffer against uncertainty should depend on the nature of uncertainty. If each period the demand fluctuates around the forecast, then it is best to use a safety stock per product, but if the main source of uncertainty is that customers often put their large orders (lumpy demand) in another period than expected, then a safety leadtime performs better.

The lesson that is to be learnt from their research, is that one must first know what type of uncertainty one is confronted with before starting to create buffers against it.
Looking at the possible sources of uncertainty at the level of Material Coordination, one can distinguish two types of uncertainty:
On the one hand, there are uncertainties that are a consequence of the behaviour of individual products, e.g. uncertain demand, inventory registration or yield factor. A common way for Material Coordination to buffer against this type of uncertainty, is to create a safety stock for each individual product, which has to absorb the stochastic behaviour of that product.
On the other hand, there are uncertainties due to differences between demand and availability of resources (for example due to worker availability or machine breakdowns). The safety stock that is created to absorb these uncertainties is largely exchangeable between products: If, for some reason, it proves that a Production Unit cannot produce more than c in a specific period and it is necessary to produce y_1 for product 1 and y_2 for product 2 with $y_1+y_2 > c$, then an inventory of $(y_1+y_2)-c$ solves the capacity problem, no matter how distributed over the products (as long as the inventories do not exceed y_j).
Notice that the discrepancy between the availability of capacity (c) and the demand for capacity (y_1+y_2) may be due to the behaviour of the capacity or to the behaviour of the aggregate demand of the products. This shows that the two types of uncertainty, that we have mentioned, are interrelated. Consequently, the stocks to buffer against these uncertainties should not be determined independently of each other. However, both aspects of uncertainty require a different approach to creating safety stocks: The product-aspect of uncertainty requires decomposition over the individual products so as to isolate the

behaviour of each product, whereas the capacity-aspect requires aggregation over the products in order to be able to consider the behaviour of the total demand put on the resources. As a consequence, the stock that is meant to buffer against the capacity-aspect of uncertainty is largely exchangeable between products, whereas for the product-aspect this exchangeability is limited.

Connected with these two aspects of uncertainty, there are two extreme approaches to the design of a Material Coordination System, namely a product-oriented approach and a capacity-oriented approach. Roughly these approaches can be described as follows:

- product-oriented approach. The required delivery patterns have to be translated by Material Coordination to production patterns. In a product-oriented approach the first step is to determine the requested production patterns by straightforward offsetting, not taking capacity restrictions into account but using standard throughput-times. The second step in the product-oriented approach is to coordinate the different production patterns. In this step the capacity restrictions are taken into account. Typically the horizon in the second step is smaller than in the first step. Uncertainties in required delivery patterns and capacity availibility and the interference between products because of restricted capacities can be attacked by safety stocks and safety leadtimes in the first step, so per product.
  Material Requirements Planning is an example of the product-oriented approach.

- capacity-oriented approach. Material Coordination first makes a production level plan, possibly combined with a capacity adjustment plan. This requires aggregation of delivery patterns and inventories to capacities. Then, in a second planning step, the production level plan for the first period is distributed over the different products, only using short-term detailed information. This disaggregation can be based, for instance, on the run-out times of the individual products. With the run-out time of a product, we mean the expected time until a stock-out occurs for that product.

Uncertainties in the capacity availability and the total required deliveries can be taken into account in the first planning step. Imbalances between the individual products, resulting from this procedure, may also be estimated in an aggregate way. It is possible to determine how much extra (aggregate) inventory is necessary because of these imbalances.
Such capacity-oriented approaches have been proposed by van Beek [6], Magee [37] and Meal [40]. They stress the capacity adjustment in the first step and assume in the second step that the capacity usage and the capacity availability are equal.

Both approaches are feasible. It is not clear however when to use what approach. It may be so that both approaches work well in certain situations, while in other situations only a mixture of both approaches is satisfying. An interesting mixture of both approaches for the single capacity case has been proposed by Graves [24].

2.6 The single-phase multi-product planning problem.

In the previous Section, we have mentioned two extreme approaches to Material Coordination. We want to investigate their weak and their strong points in this text. Thus we hope to provide the reader with a tool to decide which approach to use when designing a Material Coordination System in a practical situation.

For this investigation we have used the simplest model in which there is a distinction between both approaches, namely the single-phase multi-product planning model (with one clear capacity bottle-neck). The reason to consider this model is not because it is such a good model for many realistic situations (although it may be so for certain situations), but because it is the most straightforward starting point for the analysis of the weak and strong points of both approaches.

The single-phase multi-product planning problem has a long history in the theoretic research. However this research has been dominated by models with a deterministic demand.
Elmaghraby [19] presents a good overview of the work in this field. However, when one is interested in the question how to buffer effectively against uncertainties, the results from that research are not very helpful since it proves to be difficult to extend them to stochastic situations (see e.g. Graves [24]).
More recently, the research in this field has also incorporated stochastic elements in the model. Consider for example the work of Federgruen and Zipkin [20], Graves [24] and Williams [55]. There is a lot of similarity between the research described in this text and their work. Our interpretation however differs from theirs because we have a different notion of the single phase. We view this phase as a controlled Production Unit, consisting of more machines and workers, whereas in the mentioned research this phase is meant to represent a single machine. This leads to somewhat different characteristics for the behaviour of the capacity in the models. This means that their results cannot simply be applied if analyzing the control of a Production Unit. Yet, some results can be used and we will refer to these in the next Chapters.

In the single-phase models, that we will consider in this text, the batch-sizes have been fixed. The reason for this is that the possibility to use the capacity efficiently, usually, interferes heavily with the choice of the run sizes. Therefore, at the level of Tactical Planning, at least restrictions have to be imposed on the batch-sizes, in order to be able to decide whether the availability of the resources has to be adjusted. On the level of Material Coordination, that falls under the heading of Operational Control, the availability of the resources is given. Material Coordination has to provide for the timing of production orders. It would have been possible to work with restrictions for the batch-sizes on the level of Material Coordination, without actually fixing the run-sizes. However, in order to simplify the analysis, we will assume that the batch-sizes are fixed on the Tactical Level (we will return to this in Chapter 5).

Notice that in situations with incidental, large demands (think of project-situations) it will not be sensible to introduce such a decomposition between the level where the batch-sizes are determined and the level where the timing of production orders is provided for. Therefore, we will restrict ourselves to situations with a relatively smooth demand. In the situations that we will consider, the demand follows a stochastic process, that may be partly known beforehand. For such situations, de Bodt and van Wassenhove [14] have shown that forecast errors have a large impact on the cost effectiveness of lot-sizing techniques when used in a rolling schedule approach. Therefore, there will be little sense in leaving the decision on the batch-sizes to Material Coordination in such situations.
Once the batch-sizes have been fixed, the total set-up times and set-up costs can no longer be influenced. Therefore, these set-up costs and times may be ignored at the level of Material Coordination (the set-up times are then viewed as part of the processing time for a batch).

In this text, we will compare product-oriented and capacity-oriented approaches to Material Coordination. Before starting to investigate the product-oriented and the capacity-oriented approaches, there is one advantage of the capacity-oriented approach, that we want to mention already, since it is connected to the levels of control that are described in this Chapter:
The capacity-oriented approach makes the relationship to higher levels of control easier. It is possible to combine capacity adjustment decisions with production level decisions. In case of a product-oriented approach one needs a separate (aggregate) model to make the capacity adjustment decision and it is not always easy to couple this level of decision making properly to the (detailed) product-oriented approach for Material Coordination.

Chapter 3. Identical products; purely stochastic demand.

3.1 Introduction.

In the previous Chapter, we have described the level of Material Coordination within a general framework for Production Control. Also, we discussed the uncertainties of the environment to which Material Coordination is exposed, and we described the need for effective ways to buffer against these uncertainties. This led us to choose the single-phase multi-product model for this research.
In the single-phase multi-product model, we can distinguish two basically different approaches to the design of a Material Coordination System, namely the product-oriented approach and the capacity-oriented approach. Since these two approaches, of which we want to investigate the weak and strong points, differ fundamentally in the way they buffer against uncertainties on the level of Material Coordination, we will study a stochastic single-phase model. Consequently, a situation with a stochastic arrival process for demand and a stochastic availability process for the resource will be considered. To facilitate the analysis in this Chapter, we will assume that for all products the demand processes are the same. Also the production characteristics for all products are the same (we speak of "identical" products).
In the purely stochastic case, demand follows a given stochastic process. In subsequent Chapters, we will extend the analysis to situations where demand is partly known beforehand, and also to situations with "non-identical" products.

Although the reason to introduce the single-phase model is to compare the product-oriented approach with the capacity-oriented approach, we will first formulate the problem of finding the overall optimal strategy as a Markov decision problem. Solving this Markov decision problem for some special cases, we have a point of reference to measure how the performance of strategies deteriorates when we restrict ourselves to simple approximating strategies. At hand of this Markov decision problem, we can illustrate the difficulties that are inherent in using the overall optimal strategy. Thus, we can discuss the reasons for searching simpler strategies. We will then describe approaches that are more generally used in optimization theory to overcome these difficulties, and show the relationship with the capacity-oriented and product-oriented strategies that we want to compare in this text.

If one is going to restrict oneselves to approximating strategies, it is necessary to investigate what opportunities there are for doing so in general. Wismer [56] says on this subject in the preface of "Optimization methods for large scale systems ... with applications":

> "We have confined our attention to those developments which are applicable to large-scale systems. As a rule such developments are characterized by (1) decomposition of the large system into smaller subsystems which are later composed or coordinated to reconstruct the original system, or by (2) aggregation of the variables in the large system thereby reducing its dimensionality. The ultimate gain from these methods is not only to obtain a computer solution in a reasonable time (or at all) but also to aid in the conceptualization and understanding of large-scale interactions."

We will show, in Section 3.4, that a sensible form of aggregation in the model actually conforms to restricting the class of strategies in the Markov decision problem to capacity-oriented strategies. Analogously a sensible form of decomposition in the model conforms to restricting the class of possible strategies to product-oriented strategies (see Section 3.5). This brings us back to our main theme, namely the comparison of the capacity-oriented and the product-oriented approaches.
For further information on aggregation in production planning, the reader is referred to Axsäter and Jönsson [4], Ritzman and Krajewski [45], Wijngaard [58] and Zipkin [62]. A more general discussion of the concept of aggregation is found in Liesegang [35].

As will be shown, in Sections 3.4 and 3.5, within each class of strategies (capacity-oriented and product-oriented strategies) an optimal strategy can be found. However, this strategy may not be attractive for practical use. Therefore, we will also develop simple heuristics within each class. With the determination of these heuristics, we will find an approximation of the actual costs too.

In Section 3.6, we will concentrate on a more specific single-phase multi-product planning model. For this specific model, we will compare the strategies and heuristics by means of simulation (Sections 3.7 and 3.8).
In Section 3.9, we will discuss the sensitivity of the results for the choice of the specific model (of Section 3.6).

3.2 General formulation of the model.

In this Chapter, we will consider a production system in which N products are made for stock by a single Production Unit (see Figure 3.1).

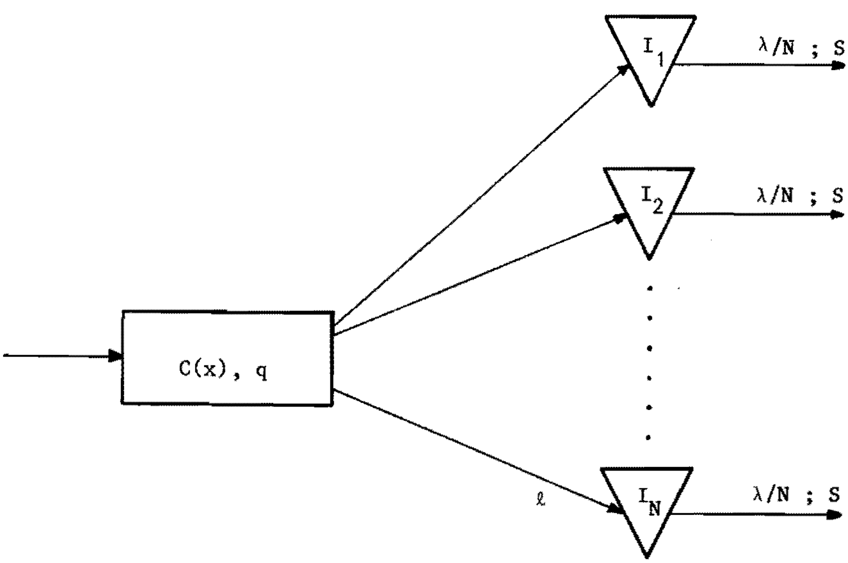


Figure 3.1. A single Production Unit.

The demand for different products is not correlated. For each product the demand follows a stochastic process. The interarrival time between successive "demand instants", for a given product, is negative-exponentially distributed with mean N/λ. If a demand occurs, the demand size follows a discrete stochastic distribution. The stochastic demand size will be denoted by S ($S \geq 0$). The demand for a given product, say j, is satisfied from the stock-point of that product. If this stock is too small, the remaining part of the demand is backordered.
The inventory level of product j at time t is denoted by $I_j(t)$, which may be negative to indicate a backlog. To be more precise, we will define $I_j(t)$ to be left-continuous. This means that if at time t a demand or a production opportunity arrives, this is accounted for in the inventory of product j in the open interval (t,∞), as is shown in Figure 3.2. All variables, that we will introduce in this text, for which a choice between left-continuous or right-continuous has to be made, will be assumed to be left-continuous.

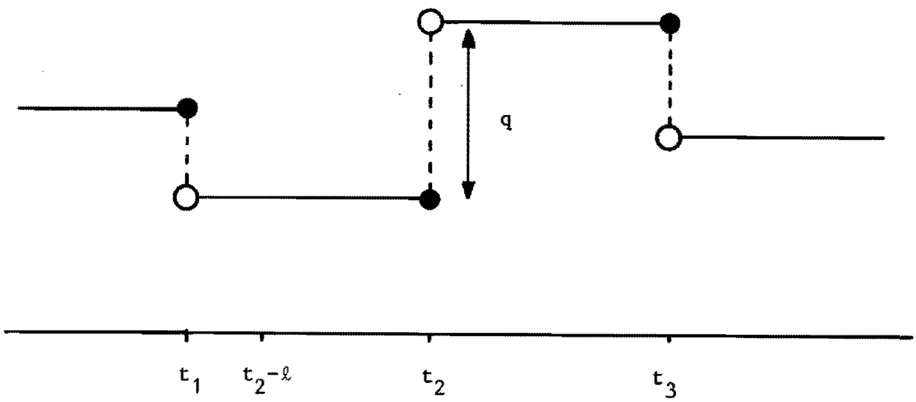


Figure 3.2. Example of the inventory pattern of a product. At times t_1 and t_3, a demand arrives, whereas at time t_2 a batch of size q (which has been started at time $t_2-\ell$) arrives.

Note that the inventory level corresponds to the physical inventory only when $I_j(t) \geq 0$.

To the inventory of a product j, a certain cost is assigned. Therefore, we will define a cost rate $p(i)$. If the inventory level at time t for product j equals i_j (i.e. $I_j(t)=i_j$), then the total cost increases at time t with rate:

$$\sum_{j=1}^{N} p(i_j) \tag{3.1}$$

In our model, we will only consider cost rates p(i), that are non-negative and convex. We will also require that p(i) goes to infinity as i goes to plus or to minus infinity.

In order to control the inventory levels, production runs can be started at specific instants of time, which will be referred to as "production opportunities". Production opportunities arrive at independent interarrival times with a non-lattice distribution function C(t). For convenience with notations, we assume that C(0)=0.
At any production opportunity, for at most one product, a production run of size q (which is chosen integer and supposed to be fixed beforehand) can be started. It may be decided for which product this run is started. The production system is controllable, if we impose the following restriction on q, which ensures that the utilization rate of the Production Unit is less than one:

$$\lambda \cdot ES < \frac{q}{\int_0^\infty t \, dC(t)} \tag{3.2}$$

where ES denotes the expected value of the stochastic demand S.

If a production run is started for product j at a given production opportunity, then the batch q will arrive at the stock-point for that product after a production throughput-time ℓ. This implies that there may be more production orders in process at the same time.
Consequently, the results of this Chapter can apply to situations with more production facilities in the Production Unit, or to situations with more Production Units in series, as long as there is only one bottle-neck (see Figure 3.3).

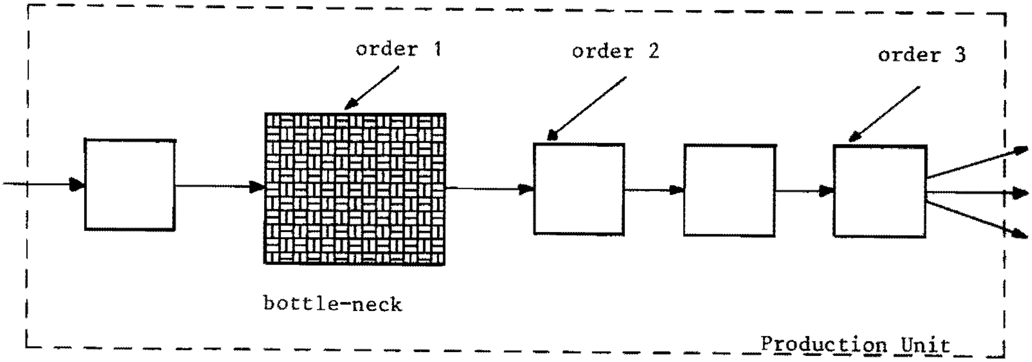


Figure 3.3. A Production Unit with a single bottle-neck.

The performance objective of this model is to minimize the average cost that is incurred, over time. However, from a mathematical point of view, this average cost needs not be well-defined if we apply an arbitrary strategy. Therefore, we define the cost of a strategy π, as in (3.3):

$$\limsup_{T\to\infty} E_\pi \left[\frac{1}{T} \int_0^T \sum p(I_j(t))dt \right] \tag{3.3}$$

With $E_\pi(X)$, we mean the expectation of X when applying strategy π.

3.3 Overall optimal strategy.

Suppose a production opportunity arrives at time s. Due to the throughput-time of an order, only the inventory levels over the interval $(s+\ell,\infty)$ can be influenced by the production decision. Therefore, we will introduce the inventory position of product j at any time t, denoted by $Ipos_j(t)$, as the inventory $I_j(t)$ plus the amount for which a production run has been started before t which will arrive at the stock-point before $t+\ell$.

We introduce the "shifted cost rate" as follows: The shifted cost rate for product j at time t is the expected cost rate at time $t+\ell$, if the inventory position at time t, for that product, is given. If we denote this shifted cost rate by $L(ip_j)$, if the inventory position of product j equals ip_j, then we can write this as in (3.4):

$$L(ip_j) := E[\ p(I_j(t+\ell)) \mid Ipos_j(t)=ip_j\ ] \tag{3.4}$$

where the formula behind the vertical bar denotes a condition on the expectation.
Note that $L(ip_j)$ does not depend on t or on j or on π.

If we define $G^{(s)}(x)$ as the probability that the demand for a given product in the interval $[t,t+s)$ equals x, then we can write (3.4) as (just sum over all possible realisations of demand in the interval $[t,t+\ell)$):

$$L(ip) = \sum_{x=0}^{\infty} p(ip-x)\cdot G^{(\ell)}(x) \tag{3.5}$$

Notice that since p(i) is convex in i, L(ip) is convex in ip.

We will now show that we may replace the cost rate p(i) in (3.3) by the shifted cost rate, if the expected costs over the first ℓ units of time are finite. Notice that we cannot influence these costs by the production decisions, so that this assumption seems reasonable. Using that the expectation of the conditional expectation is the unconditional expectation $(E(E(X|Y))=EX)$, we find

$$E_\pi[\tfrac{1}{T}\ {}_0\!\int^T \sum_{j=1}^N p(I_j(t+\ell))dt\,] = \tfrac{1}{T}\ {}_0\!\int^T \sum_{j=1}^N E_\pi p(I_j(t+\ell))dt =$$

$$\tfrac{1}{T}\ {}_0\!\int^T \sum_{j=1}^N E_\pi[Ep(I_j(t+\ell))|Ipos_j(t)]dt = E_\pi[\tfrac{1}{T}\ {}_0\!\int^T \sum_{j=1}^N L(Ipos_j(t))dt\,] \tag{3.6}$$

Therefore, we can formulate the production planning problem entirely in terms of the inventory positions. For notational convenience, we will define $\underline{Ipos}(t)$ as the vector of inventory positions (and the realisations $\underline{ip}(t)$ analogously), via (3.7).

$$\underline{Ipos}(t) := (Ipos_1(t),\ Ipos_2(t),\ \cdots,\ Ipos_N(t)) \tag{3.7}$$

Notice that at a given production opportunity, the time until the next demand instant is negative-exponentially distributed, and the demand-size has a fixed distribution. Consequently, the demand process is regenerated at any production opportunity. Since the interarrival times between successive production opportunities have a fixed distribution as well, also this process is regenerated at a production opportunity. Consequently, the distribution of demand for any product j between successive production opportunities remains constant over time. For any vector $\underline{x} = (x_1, x_2, \cdots, x_N)$, with non-negative integer entries, let $r(\underline{x})$ denote the probability that the realisation of demand for product j between successive production opportunities equals x_j ($j = 1, 2, \ldots, N$). It is easy to see that $r(\underline{x})$ is given by (3.8).

$$r(\underline{x}) = \Pi_{j=1}^{N} \left\{ \int_{t=0}^{\infty} G^{(t)}(x_j)dC(t) \right\} \tag{3.8}$$

Since the demand process and the capacity-availability process are both regenerated at the production opportunities, we will restrict ourselves to strategies in which the production decision depends solely on the realisation of the vector of inventory positions upon arrival of a production opportunity. Such strategies can be written as a function $\pi(\underline{ip})$, where $\pi(\underline{ip})=r$ means that a production run for product r is started if upon arrival of a production opportunity the realisation of the vector of inventory positions is as given (r=0 indicates the decision not to start a production run).

If we define $\underline{Ipos}_m$ as the vector of inventory positions upon arrival of the m-th production opportunity, then $\{ \underline{Ipos}_m,\ m = 1, 2, \ldots\}$ is a Markov Chain with countable state-space for any given strategy π.

The transition probabilities in this Markov Chain depend on the strategy π and on the parameters of the demand process between successive production opportunities (determined by (3.8)). It is therefore straightforward to determine the probability $r_\pi(\underline{i},\underline{k})$ of a transition from $\underline{i}$ to $\underline{k}$.

Suppose that the process starts with all inventory positions equal to zero and that immediately a production opportunity is available. Then we can define $\tilde{H}_\pi^{(m)}(\underline{k})$ as the probability that we are in state $\underline{k}$ upon arrival of the m-th production opportunity. We assume that $\tilde{H}_\pi(\underline{k}) := \lim_{m\to\infty} \tilde{H}_\pi^{(m)}(\underline{k})$ exists and that $\sum_{\underline{k}} \tilde{H}_\pi(\underline{k}) = 1$. This means that the Markov Chain will tend to a steady-state.
These "steady-state" probabilities then satisfy the following equilibrium relations:

$$\tilde{H}_\pi(\underline{k}) = \sum_{\underline{i}} r_\pi(\underline{i},\underline{k}) \cdot \tilde{H}_\pi(\underline{i}) \qquad \text{for all } \underline{k} \tag{3.9}$$

To solve this set of equilibrium relations, one may use successive approximation (after truncating the state-space to a finite one). However, the size of the state-space will be large, which will present numerical problems, especially in case of many products. This is mainly due to the fact that the state-space is N-dimensional.

Now suppose that the system has reached its equilibrium. Consider an arbitrary point in time. The time that has elapsed since the previous production opportunity has distribution function R(t) as in (3.10) (see Ross [46], notice that it is required that C(x) is not lattice).

$$R(t) = \frac{\int_0^t (1-C(x))dx}{\int_0^\infty (1-C(x))dx} \tag{3.10}$$

This enables us to determine the probability that at an arbitrary point in time (if the system is in equilibrium) the vector of inventory positions is given by $\underline{k}$. Denote this probability by $H_\pi(\underline{k})$ and let

$r_\pi^t(\underline{i},\underline{k})$ denote the probability that we are in state $\underline{k}$, given that the elapsed time since the previous production opportunity equals t and that the state upon arrival of that production opportunity was $\underline{i}$ (notice that $r_\pi^t(\underline{i},\underline{k})$ can be determined analogously to $r_\pi(\underline{i},\underline{k})$). Then $H_\pi(\underline{k})$ satisfies (3.11).

$$H_\pi(\underline{k}) = \int_{t=0}^{\infty} \sum_{\underline{i}} r_\pi^t(\underline{i},\underline{k}) \cdot \tilde{H}_\pi(\underline{i}) \; dR(t) \quad \text{for all } \underline{k} \tag{3.11}$$

Via (3.11), we can determine $H_\pi(\underline{k})$ if we know $\tilde{H}_\pi(\underline{k})$ (again after truncating the state-space).

Under weak regularity conditions, the average expected costs in the long run are equal to the expected costs in the steady-state situation, i.e. (3.12) holds.

$$\lim_{T\to\infty} E_\pi[\frac{1}{T} \int_0^T \sum_{j=1}^N L(Ipos_j(t))dt] = \sum_{\underline{ip}} [\sum_{j=1}^N L(ip_j)] \cdot H_\pi(\underline{ip}) \tag{3.12}$$

In theory, now, it is possible to determine the overall optimal strategy by minimizing expression (3.12) over π. Usually, however, this approach is not attractive for practical use. The main reason for this inattractivity is the complexity of the encountered numerical problems. As we have seen, the state space is N-dimensional. Consequently, only for limited values of N, the optimal strategy can be found. Besides these numerical problems, it may be so that the optimal strategy is complicated, which makes it difficult to react on unexpected events. In order to avoid these problems, we will restrict ourselves beforehand to strategies that are easy to find and easy to understand.

When searching for simple, approximating strategies, two techniques will be used, namely aggregation and decomposition. As we will see in the next Sections, this brings us back again to the comparison of capacity-oriented and product-oriented strategies for this single-phase model.

3.4 Aggregation; Capacity-oriented strategies.

The main source of the numerical difficulties, encountered when using the strategy that is discussed in the previous Section, is the dimensionality of the state-space. Therefore, we will consider approaches to reduce this dimensionality. The two most important techniques for doing so are aggregation and decomposition. In this Section, we will describe the aggregation approach. The decomposition approach is discussed in Section 3.5.

Since all products are identical, a natural way of defining an aggregate inventory position at time t, is via

$$\mathrm{Ipos}(t) := \sum_{j=1}^{N} \mathrm{Ipos}_j(t) \tag{3.13}$$

The decision whether or not to produce, will now be based only on the aggregate inventory position upon arrival of a production opportunity. Since L(ip) is convex, it seems reasonable to restrict oneselves to strategies of the following type:

> Start a production run if and only if the aggregate inventory position at the production opportunity is less than or equal to a prescribed level β (β is chosen integer).

Remark. Under some weak regularity conditions, it can be proven that for N=1 the optimal strategy is of the above form (see Wijngaard [59]).

If, at a given production opportunity, it is decided to start a production run, then we have to assign this run to a product. Obviously, for such a disaggregation step we need detailed information on the distribution of the aggregate inventory position over the individual products. However, due to the convexness of L(ip), this disaggregation step is simple, since it will be optimal to assign the production run to the product with the lowest inventory position (in this situation with identical products).

Notice that the strategies that are found using aggregation, correspond to the ones which we have introduced (in Chapter 2) as capacity-oriented strategies: The production decision ignores the status of individual products, but it is based on the aggregate inventory position that reflects how much capacity is stored in the system as a whole.

In order to be able to rewrite (3.12), we define $a(\underline{ip}) := \sum_{j=1}^{N} ip_j$. If we use index β instead of π (this is possible since capacity-oriented strategies are entirely determined by the choice of β), we can write the average, expected costs for any capacity-oriented strategy β as:

$$\sum_{\underline{ip}}[\sum_{j=1}^{N} L(ip_j)]\cdot H_\beta(\underline{ip}) =$$

$$\sum_k \sum_{\underline{ip} \text{ with } a(\underline{ip})=k}[\sum_{j=1}^{N} L(ip_j)]\cdot H_\beta(\underline{ip}|a(\underline{ip})=k)\cdot H_\beta^{(a)}(k) =$$

$$\sum_k g_\beta(k)\cdot H_\beta^{(a)}(k) \qquad (3.14)$$

where $H_\beta(\underline{ip}|a(\underline{ip})=k)$ is the conditional probability that the vector of inventory positions equals $\underline{ip}$ given that the aggregate inventory position is k, and $H_\beta^{(a)}(k)$ is the probability that the aggregate inventory position equals k (when applying the capacity-oriented strategy β). Notice that $g_\beta(k)$ is the conditional expectation of the cost rate, given that the aggregate inventory position equals k.

To evaluate (3.14), it is required to determine the steady-state probabilities $H_\beta^{(a)}(k)$. This can be done by introducing the steady-state probabilities $\tilde{H}_\beta^{(a)}(k)$, that denote the probability that the realisation of the aggregate inventory position, upon arrival of a production opportunity, equals k.

Define, analogous to Section 3.3, $ra_\beta(i,k)$ as the probability of a transition of the aggregate inventory position to k upon the arrival of a production opportunity, given that upon arrival of the previous production opportunity the aggregate inventory position was i. Notice that this probability only depends on the choice of β and on the parameters of the aggregate demand process between successive production opportunities. Instead of (3.9), we now find:

$$\tilde{H}_{\beta}^{(a)}(k) = \sum_i ra_{\beta}(i,k)\cdot\tilde{H}_{\beta}^{(a)}(i) \qquad \text{for all } k \tag{3.15}$$

Notice that these equilibrium equations are one-dimensional. Therefore it is much simpler to solve these equations (again after truncating the state-space) than to solve (3.9).

Now consider an arbitrary point in time given that the system is in equilibrium. Define $ra_{\beta}^{t}(i,k)$ as the probability that the aggregate inventory position equals k at this point in time, given that the elapsed time since the previous production opportunity is t and that the aggregate inventory position upon arrival of that production opportunity equalled i. Then, analogous to (3.11), we find the following expression for $H_{\beta}^{(a)}(k)$:

$$H_{\beta}^{(a)}(k) = \int_{t=0}^{\infty} \sum_i ra_{\beta}^{t}(i,k)\cdot\tilde{H}_{\beta}^{(a)}(i)dR(t) \quad \text{for all } k \tag{3.16}$$

The above shows how $H_{\beta}^{(a)}(k)$ can be calculated.

To find the optimal capacity-oriented strategy, that minimizes (3.14), we also have to determine $g_{\beta}(k)$. The determination of $g_{\beta}(k)$ requires the calculation of the conditional probabilities $H_{\beta}(\underline{ip}|a(\underline{ip})=k)$. Unfortunately, however, this makes the problem again as complex as the determination of the overall optimal strategy, since it will lead to an N-dimensional state space. The optimal capacity-oriented strategy is only interesting therefore as a point of reference, but is not useful in practice. For practical use, it is necessary to consider strategies that are based on simple approximations of $g_{\beta}(k)$. The simplest approximation can be found by assuming that it is possible to keep the inventory positions of all products equal. Of course this will not always be possible, but the objective of the production allocation is to keep the inventory positions equal. This assumption, which is independent of the choice of β, leads to the following approximation of $g_\beta(k)$, that we denote as $\hat{g}(k)$:

$$\hat{g}(k) := N \cdot L(k/N) \tag{3.17}$$

The strategy that is found when minimizing (3.18) over β will be referred to as the simple capacity-oriented heuristic.

$$\textstyle\sum_k \hat{g}(k) \cdot H_\beta^{(a)}(k) \tag{3.18}$$

The actual cost of using this heuristic is not given by (3.18), but is determined by (3.14). The actual cost, which will be denoted by SCH, will usually be difficult to determine.

The convexity of $L(ip)$ implies that $\hat{g}(k) \leq g_\beta(k)$ applies for all values of k. Therefore the minimum of (3.18) over β is a minimum for the average expected costs, not only for the capacity-oriented strategies, but also for the overall optimal strategy. This "capacity lower bound" will be denoted by CL.
If we define OC as the cost that corresponds to the overall optimal strategy, and we let OCC be the cost of the optimal capacity-oriented strategy, then the following inequalities will hold:

$$CL \leq OC \leq OCC \leq SCH \tag{3.19}$$

This means that the simple capacity-oriented heuristic is (almost) optimal if for a given situation CL≈SCH. Therefore, if CL≈SCH, there is no reason to search for more advanced capacity-oriented heuristics that take into account that keeping the inventory positions equal will not be possible.

In Section 3.7, we will return to the capacity-oriented approach for a more specific model. Using the results presented there, we will also discuss the situations for which it may be useful to derive more advanced heuristics, which are based on an approximation of $g_\beta(k)$ taking into account the fact that keeping all inventory positions equal will not be possible.

3.5 Decomposition; product-oriented strategies.

In this Section, we will use the technique of decomposition to reduce the complexity of the production planning problem. Consequently, instead of considering one model with an N-dimensional state-space, we will consider N models with one-dimensional state spaces (one for each product). We have introduced strategies that are based on this approach in Chapter 2 as product-oriented strategies.
In each one-dimensional model it will be optimal, since L(ip) is convex, to start a production run if the inventory position upon arrival of the production opportunity, is less than or equal to a predetermined, critical level, say β (see Wijngaard [59]).
Once it is decided for each product whether a production run is required for that product, a run will be started at the production opportunity if and only if there is at least one product for which such a run is required. Consequently, since the demand process for all products is the same (and therefore also the critical level β, which is chosen integer), the decision to use a production opportunity to start a production run depends only on the product with minimal inventory position. Therefore, we define product-oriented strategies as strategies in which the decision to start a production run can be characterized by:

> Start a production run if and only if the minimal inventory position at the production opportunity is less than or equal to a prescribed level β.

Since it is possible that there are several products for which a production run is required, we have to choose a rule to assign the run to one of the products. For reasons, that have been explained in

Section 3.4, it is optimal to assign the run always to the product with minimal inventory position (in case there are several products with minimal inventory position, we will choose one of them randomly).

Define $H_\beta^{(d)}(k)$ as the marginal probability (if the system is in equilibrium) that the realisation of the inventory position of a given product equals k, under application of the product-oriented strategy characterized by β. Then we can write the average cost of (3.12) as in (3.20)

$$N\cdot[\textstyle\sum_k L(k)\cdot H_\beta^{(d)}(k)] \tag{3.20}$$

The optimal product-oriented strategy is the strategy that minimizes (3.20). The corresponding cost will be referred to as OPC. To find OPC, we have to determine the steady-state probabilities $H_\beta^{(d)}(k)$, which unfortunately again is an N-dimensional problem: The possible delay due to other, more urgent products, also demanding a production run at a given production opportunity, has to be taken into account.
Therefore it is necessary to search for simple product-oriented strategies that are based on an approximation of $H_\beta^{(d)}(k)$ which can be determined easier.
The most straightforward approximation we find if assuming that the delay due to other products is always equal to zero. This conforms to the situation where at a given production opportunity, we may start more than one production run. Consequently, there is no interference between the products on the capacity. Therefore, it is optimal to decompose over the products. If we denote the steady-state probability upon arrival of a production opportunity, under this assumption, by $\hat{H}_\beta^{(d)}(k)$, then we can derive analogous relations for $\hat{H}_\beta^{(d)}(k)$ as for $H_\beta^{(a)}(k)$ in the previous Section.
The optimal strategy that is found under this assumption, that means the strategy minimizing (3.21), will be referred to as the simple product-oriented heuristic.

$$N\cdot[\sum_k L(k)\cdot\hat{H}_\beta^{(d)}(k)] \qquad (3.21)$$

It is easy to see that the theoretic cost (3.21) that corresponds to the product-oriented heuristic, gives a lower bound for the cost of the overall optimal strategy.
This "product lower bound" will be denoted by PL. The real cost of this heuristic is denoted by SPH (and can be determined by (3.20)).
Thus for the product-oriented approach, we find:

$$PL \leq OC \leq OPC \leq SPH \qquad (3.22)$$

Again, as in the case of the capacity-oriented approach, we see that if PL≈SPH there is no reason to search for more advanced heuristics, since the simple product-oriented heuristic is (almost) optimal. In situations where there is a large gap between SPH and OPC it may be useful to construct more advanced heuristics by taking the interference on the capacity into account. This can be done by introducing a delay in the start of orders of a certain product that is not equal to zero. Williams [55] has suggested a queueing type of analysis to estimate the delay that is due to other products. An analogous approach has been proposed by Graves [24].
Though one may expect this queueing type approach to give better results in case SPH-OPC is large, it will usually not yield the optimal product-oriented strategy. The reason for this is that subsequent delays are not independent of each other and not independent of the inventory position of the product.
In Section 3.8, we will return to the product-oriented strategies and we will also discuss the question of when it may be useful to construct more advanced heuristics.

3.6 Choice of a specific model.

In the previous Section, we have described the optimal strategy, the capacity-oriented strategies and the product-oriented strategies for the general model. In this Section, we want to compare these different strategies with each other. Therefore we will derive numerical results for these strategies for a specific model within this general class of models. These numerical results will be used to draw certain conclusions. As to how far it is possible to extend the conclusions to other situations is investigated in Section 3.9.

The model that we choose to analyse numerically has the following characteristics:

- the demand size equals one at each demand instant ($S=1$)

- inventory holding costs and stock-out costs are linear: $p(i)=ai^{+}+bi^{-}$, where $i^{+}=\max(0,i)$ and $i^{-}=\max(0,-i)$

- the interarrival time between successive production opportunities is negative-exponentially distributed with mean $1/\mu$. Consequently, the utilization rate of the capacity equals $\rho := \lambda/(\mu\cdot q)$.

In Section 3.7, we will first derive numerical results for the overall optimal strategy when there are only two products ($N=2$). These results are meant to provide us with a yardstick to measure the performance of the capacity-oriented strategies and the product-oriented strategies. These strategies are considered in Section 3.8, also for situations with $N > 2$.

3.7 Overall optimal strategy; numerical results.

In this Section, we will present some numerical results for the overall optimal strategy in the model that has been chosen in the previous Section. To obtain these results, we have made use of the analysis in Section 3.3, that is given for a general class of models. It should be noticed that the assumption that the interarrival time between successive production opportunities is negative-exponentially distributed, makes the analysis a lot simpler. Due to this assumption, namely, it holds that the distribution function, at an arbitrary steady-state moment in time, for the amount of time that has elapsed since the previous production opportunity, has the same negative-exponential distribution:

$$C(t) = R(t) \tag{3.23}$$

This means in the first place that $H_\pi(\underline{ip}) = \tilde{H}_\pi(\underline{ip})$, so that we only have to determine the steady-state probabilities upon arrival of a production opportunity. It also means that we can simplify the equilibrium equations (3.9), by no longer considering the process upon arrival of production opportunities, but on so-called "change instants". By a change instant we mean a point in time in which either a production opportunity or a demand arrives. Note that, since we may decide not to produce at a given production opportunity, it is not necessary for the state of the system to change at a change instant, but it is only possible then.
Since both demand instants and production opportunities are generated by a Poisson process (with parameter λ, respectively μ), the process that generates the change instants is also a Poisson process (with parameter $\lambda+\mu$).
A given change instant corresponds to a production opportunity with probability $\frac{\mu}{\lambda+\mu}$ and to a demand instant with probability $\frac{\lambda}{\lambda+\mu}$.

Introduce the following definitions:

$$\begin{cases} \underline{e}_r = (0,\dots,0,\underset{\uparrow \text{ r-th entry}}{1},0,\dots 0) & \text{for } r=1,2,\dots,N \\ \underline{e}_0 = \underline{0} & \end{cases} \tag{3.24}$$

and

$$\chi_{\{\pi(\underline{k})\}}(r) = \begin{cases} 1 & \text{if } \pi(\underline{k}) = r \\ 0 & \text{otherwise} \end{cases} \tag{3.25}$$

Define $H_\pi(\underline{k})$ as the probability that the vector of inventory positions is given by $\underline{k}$. Then it is easy to see that the following equilibrium equations (that are analogous to (3.15)) hold:

$$H_\pi(\underline{k}) = \frac{\lambda}{\lambda+\mu}\cdot\frac{1}{N}\sum_{r=1}^{N} H_\pi(\underline{k}+\underline{e}_r) + \frac{\mu}{\lambda+\mu}\cdot\sum_{r=0}^{N} H_\pi(\underline{k}-q\cdot\underline{e}_r)\cdot\chi_{\{\pi(\underline{k}-q\cdot\underline{e}_r)\}}(r) \tag{3.26}$$

To solve these equilibrium equations numerically, by means of successive approximation, we must restrict the range of inventory positions that are allowed. Therefore we introduce the "bounded model", in which we have values Ilower and Iupper with:

- if a demand instant arrives in the bounded model for a product with inventory position equal to Ilower, then this instant is ignored.

- if, at a production opportunity, it is decided to start a production run that could result in an inventory position higher than Iupper, then the inventory position of the corresponding product is set equal to Iupper.

The introduction of Iupper has no influence on the optimal strategy or the corresponding cost, as long as the chosen value of Iupper is sufficiently large. To check the sensitivity of the results for the choice of Ilower, we give the results for two different choices for the value of Ilower.

Note that we only give results for the case of two products (N=2). The reason for this is that the size of the state space grows exponentially in N. This analysis for two products will be used to have a point of reference when we consider capacity-oriented and product-oriented strategies. Therefore, we will explain in detail why we have chosen these values for the parameters in the next Section, when we compare capacity-oriented and product-oriented approaches.

It is interesting to remark that the optimal strategy for the second situation (λ=0.5) is a product-oriented strategy (with β=-1), while the optimal strategy for the third situation (λ=1.8) is a capacity-oriented strategy (with β=18). These results are not surprising, because the utilization rate ($\rho = \frac{\lambda}{q \cdot \mu}$) is very low in the second situation and very high in the third situation. We will return to this in the next Section.

Table 3.1. Results for the overall optimal strategy with N=2 and ε=0.01, for negative-exponentially distributed interarrival times between production opportunities.

a	b	λ	μ	q	ℓ	ρ	Iupper	Ilower	up	Ilower	up
1	3	1.67	1	2	0	0.84	15	-80	11.67	-100	11.67
1	3	0.5	1	2	0	0.25	7	-20	1.87	-40	1.87
1	3	1.8	1	2	0	0.9	25	-80	19.77	-100	19.77
1	1	1.67	1	2	0	0.84	7	-40	5.92	-65	5.92
1	10	1.67	1	2	0	0.84	25	-30	19.98	-80	20.00
1	3	8.35	1	10	0	0.84	50	-80	41.47	-100	41.91
1	3	1.67	1	2	5	0.84	25	-60	12.73	-80	12.73
1	3	1.67	1	2	10	0.84	25	-60	13.76	-80	13.76

Explanation of Table 3.1: In the model, that is described in Section 3.6, the inventory positions are bounded from above by Iupper and from below by Ilower (we have given the results for two different choices of Ilower). We have solved (3.26) by means of successive approximation (see van der Wal [53]). Successive approximation gives an upper bound and a lower bound for the cost of the overall optimal strategy. As soon as the difference between these two bounds was less than ε=0.01, we stopped the iterations. In the Table we have given the value of the upper bound (up).

3.8 Product-oriented strategies and capacity-oriented strategies; numerical results.

For the model, described in Section 3.6, we will examine the numerical results that are obtained for the capacity-oriented strategies and the product-oriented strategies.
Note that for the capacity-oriented strategies it is necessary to calculate the steady-state probabilities $H_\beta^{(a)}(k)$, in order to evaluate the cost for a given capacity-oriented strategy (via (3.14)). As was shown in Section 3.4, the calculation of these steady-state probabilities is relatively easy. Only one-dimensional equilibrium relations have to be solved. In this, more specific, case, where the interarrival time between production opportunities is negative-exponentially distributed, it proves to be possible to find a closed form expression for $H_\beta^{(a)}(k)$. To see this, consider the possible transitions for the aggregate inventory position on the change instants, as depicted in Figure 3.4. Notice that (3.23) holds in this case, so that we can normalize on the change instants.

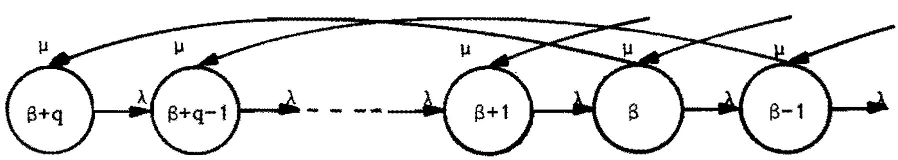


Figure 3.4. Possible transitions for the aggregate inventory position.

The transition equations in this (queueing type) model are given by

$$\begin{cases} k>\beta+q: & H_\beta^{(a)}(k)=0 \\ \beta+q\geq k>\beta \ : & \lambda\cdot H_\beta^{(a)}(k)=\mu\cdot H_\beta^{(a)}(k-q)+\lambda\cdot H_\beta^{(a)}(k+1) \\ \beta \ \geq k \ : & (\lambda+\mu)\cdot H_\beta^{(a)}(k)=\mu\cdot H_\beta^{(a)}(k-q)+\lambda\cdot H_\beta^{(a)}(k+1) \end{cases} \tag{3.27}$$

We will first postulate the solution of this system of equations as in (3.28) and then we will show that this solution is correct.

$$\begin{cases} k>\beta+q \ : & H_\beta^{(a)}(k)=0 \\ \beta+q\geq k>\beta \ : & H_\beta^{(a)}(k)=c\cdot\frac{\mu}{\lambda}\sum_{s=0}^{\beta+q-k} z^{-s} \\ \beta \ \geq k \ : & H_\beta^{(a)}(k)=c\cdot z^{k-\beta} \end{cases} \tag{3.28}$$

where $c>0$ and $z>1$, while z satisfies $f(z)=0$, with $f(y)$ defined by:

$$f(y) := \lambda y^{q+1} - (\lambda+\mu)y^q + \mu \tag{3.29}$$

It will be clear that if such a z exists, then the choice of (3.28) will satisfy the system of equations in (3.27) for $k \leq \beta-1$ (on the left hand side). For $k \geq \beta+1$ it is also easy to check that (3.27) is satisfied (start with checking $k = \beta+q$). The only equation which we have to check is the one with $k = \beta$. Consequently, we have to verify ($\overset{?}{=}$ denotes that we have to check whether the equality holds):

$$c\cdot(\lambda+\mu) \overset{?}{=} c\cdot\mu z^{-q} + c\cdot\mu\sum_{s=0}^{q-1} z^{-s} \tag{3.30}$$

which, since $c>0$ and $z>1$ conforms to

$$(\lambda+\mu)z^q \overset{?}{=} \mu + \mu z\frac{1-z^q}{1-z}$$

or equivalently

$$(\lambda+\mu)z^q - (\lambda+\mu)z^{q+1} \stackrel{?}{=} \mu - \mu z + \mu z - \mu z^{q+1} \tag{3.31}$$

which can be written as

$$(\lambda+\mu)z^q - \lambda z^{q+1} \stackrel{?}{=} \mu \tag{3.32}$$

That this equation holds, follows from f(z) = 0 (see (3.29)).

Notice that, since c > 0, all $H_\beta^{(a)}(k)$'s are nonnegative. Thus we only have to choose the normalizing constant such that $\sum_k H_\beta^{(a)}(k)=1$. Therefore, we choose c as in (3.33).

$$c = [\sum_{k=\beta+q}^{\beta} \frac{\mu}{\lambda}\sum_{s=0}^{\beta+q-k} z^{-s} + \sum_{k=\beta}^{-\infty} z^{k-\beta}]^{-1} \tag{3.33}$$

Finally, we have to check whether there exists a z > 1 such that f(z) = 0. This follows from the continuity of f(y) combined with:

* if y is very large then f(y)>0
* f(1)=0
* $(\frac{\delta}{\delta y}f(y))_{y=1} = \lambda-\mu q < 0$ (since the utilization rate is less than one).

Since each z>1 with f(z)=0 leads to a solution of the system of equations (3.27), and we know that the steady-state probabilities are unique, it follows that z is unique. This implies that e.g. the method of Newton-Raphson to determine the root z>1 of f(z)=0 converges very fast (see e.g. van der Griend [26]).

Notice that $H_{\beta-1}^{(a)}(k) = H_\beta^{(a)}(k+1)$ for all k, so that we have to determine the steady-state probabilities for only one β to find them all. This leads to a simple method for finding the simple capacity-oriented heuristic.

Since analogous results can be obtained for $\hat{H}_{\beta}^{(d)}(k)$, except that λ is replaced by λ/N, we can also find the simple product-oriented heuristic easily.
This leads us not only to the two simple heuristics, but also to their corresponding theoretic cost CL and PL.
As far as the actual costs of these strategies are concerned, the solution of an N-dimensional problem is required. Since this leads to

Table 3.2. Results for the simple heuristics for negative-exponentially distributed interarrival times between production opportunities (μ=1).

a	b	λ	N	q	ℓ	ρ	CL	PL	capacity-oriented heuristic β	capacity-oriented heuristic SCH	product-oriented heuristic β	product-oriented heuristic SPH
1	3	1.67	2	2	0	0.84	11.42	4.58	9	11.75/11.82	0	17.58/17.28
1	3	0.5	2	2	0	0.25	1.50	1.74	0	2.37/2.37	-1	1.86/1.87
1	3	1.8	2	2	0	0.9	19.63	5.08	18	19.95/20.09	1	29.82/30.81
1	3	1.67	5	2	0	0.84	11.42	5.24	9	12.84/12.92	-1	19.90/19.70
1	3	1.67	20	2	0	0.84	11.42	11.94	9	21.78/21.89	-1	21.43/21.31
1	1	1.67	2	2	0	0.84	5.71	2.28	4	5.83/5.85	-1	7.24/7.16
1	10	1.67	2	2	0	0.84	19.75	7.82	18	20.20/19.39	2	35.32/36.88
1	3	8.35	2	10	0	0.84	41.35	16.56	35	41.86/42.29	2	60.28/59.88
1	3	1.67	2	2	5	0.84	11.95	7.08	18	12.77/12.61	5	17.45/16.98
1	3	1.67	2	2	10	0.84	12.47	8.79	27	13.88/13.93	10	16.81/16.80

Explanation of Table 3.2: Using (3.28) the steady-state probabibilities $H_{\beta}^{(a)}(k)$ can be determined for any given capacity-oriented strategy β. Next, we have determined an approximation of the cost of using the capacity-oriented strategy β by applying (3.18). By minimizing this over β, we find an optimal choice for β, with corresponding cost. This approximate cost is denoted by CL, and the optimal value of β is given in the Table. In order to determine the actual cost when applying this capacity-oriented heuristic β, we have simulated the process. The results of two different simulation runs are given in the Table (under SCH).
For the simple product-oriented heuristic an analogous approach has been used.
For an explanation of the meaning of the parameters, consider the introduction of the model in Section 3.6.
The results for what we will call the reference case are given in the first line (utilization rate is about 0.84). In the subsequent lines, we give results for cases that differ from the reference case with respect to only one parameter. The only exception is the case where we change q. Since we want to investigate the sensitivity for a more lumpy production process without changing the utilization rate, we have changed λ in this situation as well. We have fixed μ=1 in Table 3.2, since the results only depend on the ratio of λ and μ (just rescale the time axis).

numerical difficulties in situations with more than two products (and we want to consider such situations too), we have used simulation instead to find these actual costs SCH and SPH.

We have determined CL,PL,SCH and SPH for a limited number of situations. The results are given in Table 3.2.
If we look at the results of Table 3.2, then we see that the simple capacity-oriented heuristic is better than the simple product-oriented heuristic, except in the case with a very low utilization rate (λ=0.5) and the case with many products (N=20). This may be expected, since in case of a limited, stochastically available capacity, the queueing phenomenon will lead to large delay times. In such a case, the capacity-oriented heuristic, that explicitly takes into account the effects of a finite capacity, performs best. Consequently, it may be

expected that if the utilization rate is low, or if the capacity-availability is less stochastic, then the simple product-oriented heuristic performs better. We will return to this hypothesis in Section 3.9, where we consider several other capacity-availability processes.

That the simple product-oriented heuristic gives better results for N=20, seems reasonable since, while the aggregate inventory position tends to be a worse measure for the state of the system, it becomes more harmless to model the products individually in case of many products. This second effect can be explained as follows: the delay due to other products depends (mainly) on the utilization rate, which stays constant if we increase the number of products. The average demand per product decreases. Therefore the demand per product during the period it queues for allocation of the capacity decreases.

When ℓ increases the simple product-oriented heuristic gets better. To understand this, it has to be realised that a change in ℓ will have similar effects as a change in L(ip) (see equation (3.5)). In Figure

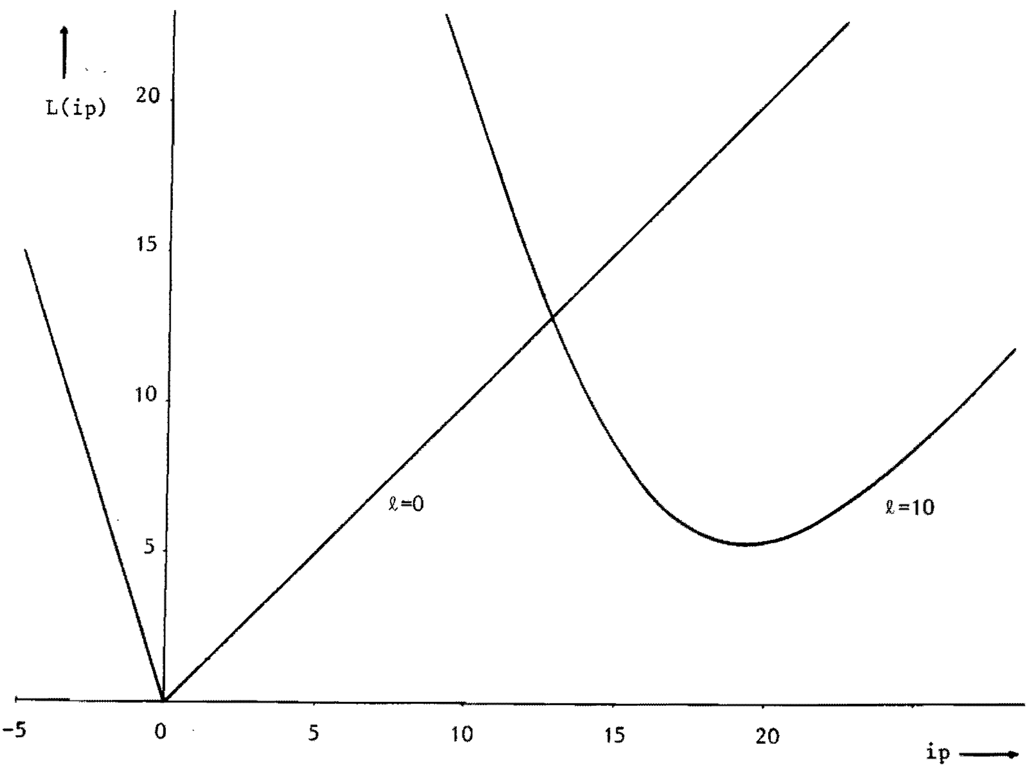


Figure 3.5. Comparison of L(ip) for ℓ=0 and ℓ=10; a=1, b=3.

3.5, we compare L(ip) for ℓ=10 with ℓ=0. In the case of ℓ=10, the minimum of L(ip) is higher and the slopes are less steep. Consequently, the influence of a delayed production due to other products is smaller in the case of ℓ=10 than in the case of ℓ=0.

From Table 3.2, we also see that max(CL,PL) = CL, except in the cases where λ = 0.5 and N = 20. These are also the only cases where SPH < SCH. This suggests an operational criterion for choosing between the simple capacity-oriented heuristic and the simple product-oriented heuristic. This criterion may be explained as follows (see Figure 3.6): since both lower bounds, PL and CL, are less than the cost of the overall optimal strategy (OC), it is obvious that the maximum of both is nearest to OC. If one lower bound is nearest to the actual cost, it may be argued that the assumptions that have led to this lower bound are most realistic and therefore the corresponding heuristic will perform best.

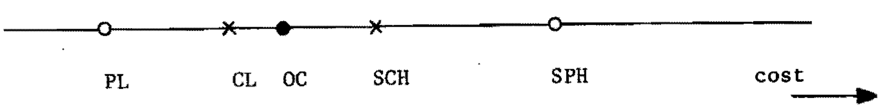


Figure 3.6. Relation between the costs.

At this point, the criterion may seem somewhat speculative, but we will verify it in Section 3.9. For the present, we will use this criterion in order to choose a heuristic.

From Table 3.2 it also follows that only in the cases with N=5, N=20, ℓ=5 or ℓ=10, the difference between the highest lower bound and its corresponding heuristic is significant. Note that the cost of the overall optimal strategy always lies between max(PL,CL) and min(SPH,SCH) (see Figure 3.6). Consequently, in all other cases, the heuristic, corresponding to the highest lower bound, is almost optimal. For the cases ℓ=5 and ℓ=10, we have determined the cost of the overall optimal strategy (see Table 3.1). It proves that in these cases, the optimal cost is almost equal to the cost of the simple capacity-oriented heuristic (OC ≈ SCH). Therefore, the best of the two simple heuristics can only be improved substantially in case of many products. In these cases numerical problems make it impossible to find the optimal strategy. We would like to know whether it is easy to find more advanced heuristics that are better than these simple heuristics in case of many products.

In order to find more advanced heuristics, it is possible to approximate the size of the delay in the availability of the capacity for a single product in the product-oriented approach, or it is possible to estimate the imbalance between the products in the capacity-oriented approach. The first approach leads to a more advanced product-oriented heuristic and the second to a more advanced capacity-oriented heuristic. These more advanced heuristics, however, will not perform better than the optimal product-oriented strategy, respectively the optimal capacity-oriented strategy. Therefore, we have determined the optimal capacity-oriented and product-oriented strategies by means of simulation so that we have an idea of how much we can improve on the simple heuristics in case of many products. The results for these strategies are given in Table 3.3.

Table 3.3. Results for the optimal capacity-oriented and product-oriented strategies for negative-exponentially distributed interarrival times between production opportunities.

a	b	λ	N	q	ℓ	ρ	optimal capacity-oriented strategy β	optimal capacity-oriented strategy OCC	optimal product-oriented strategy β	optimal product-oriented strategy OPC
1	3	1.67	5	2	0	0.84	10	12.83/12.67	1	12.77/12.54
1	3	1.67	20	2	0	0.84	12	20.66/20.69	-1	21.43/21.31

Explanation Table 3.3: By means of simulation, we have determined the optimal capacity-oriented and the optimal product-oriented strategies in case of many products. The costs that are found for two different simulation runs are given in this Table.

As follows from this Table, there is little improvement when using the optimal capacity-oriented or the optimal product-oriented strategy instead of the heuristic that has been chosen via the criterion. Therefore, in the situations considered here, there is no need to use more advanced heuristics than the simple ones. In situations where N is large and at the same time ρ is large, it may be sensible to consider more advanced heuristics (we will return to these more advanced heuristics in Chapter 4).

Note that the optimal capacity-oriented strategy and the optimal product-oriented strategy give about the same performance (OCC≈OPC). We will return to the implications of this in Chapter 4.

3.9 Sensitivity Analysis.

In Sections 3.2 to 3.5, we have discussed a general single-phase multi-product problem and we have put forward some strategies and heuristics. These strategies and heuristics have been evaluated in the previous Section. This evaluation, however, was based on the analysis of a specific model (see Section 3.6). We obtained the following results:

1. The best of the two simple heuristics is close to optimal.

2. It is best to choose the simple capacity-oriented heuristic if the corresponding lower bound is highest (CL>PL) and the simple product-oriented heuristic otherwise.

In this Section, we want to investigate whether these results can be extended to the general model (of Section 3.2) or whether they result from the choice of the specific model (of Section 3.6). This concludes our analysis of the single-phase model with identical products under a purely stochastic demand. In subsequent Chapters, we will investigate models where demand is partly known beforehand and models with non-identical products.

Since choosing the specific model consisted of three parts (see Section 3.6), the sensitivity analysis, in this Section, will consist of three parts as well:

-sensitivity for the cost rate p(i)

-sensitivity for the process that generates production opportunities.

-sensitivity for the distribution of the demand-size (S)

3.9.1 Sensitivity for the cost rate p(i).

For any given, convex cost rate p(i), with $p(i)\to\infty$ for $i\to\pm\infty$, the same approach can be used in order to find the heuristics and their corresponding lower bounds. As has been illustrated at hand of Figure 3.5, the performance of the strategy depends on the height of the minimum for p(i) (or equivalently L(ip) in case $\ell>0$) and on the steepness of the slopes near the minimum.
Therefore the sensitivity for p(i) has been checked sufficiently by comparing the reference case in Section 3.8 with the cases when b=1, b=10, $\ell=5$ or $\ell=10$. Consequently, one may expect to find a strategy that is almost optimal in situations with other cost rates as well, if one chooses the simple heuristic that is indicated by the criterion.

3.9.2 Sensitivity for the process that generates production opportunities.

To get an insight into the sensitivity of the results for the process by which capacity becomes available, we have derived heuristics for other distribution functions of the interarrival time between successive production opportunities. As we have mentioned in Section 3.8, we can rescale the first moment of the interarrival time to one.

Table 3.4. Results for the simple heuristics with distribution function $C_2(x)$ for the interarrival time between production opportunities. For an explanation see Table 3.3.

a	b	λ	N	q	ℓ	ρ	CL	PL	capacity-oriented heuristic β	SCH	product-oriented heuristic β	SPH
1	3	1.67	2	2	0	0.84	7.72	3.43	6	7.97/8.17	0	10.94/11.10
1	3	0.5	2	2	0	0.25	1.15	1.50	-1	2.23/2.23	-1	1.59/1.59
1	3	1.8	2	2	0	0.9	12.73	3.71	11	13.09/14.11	0	21.28/22.02
1	3	1.67	5	2	0	0.84	7.72	4.34	6	9.30/9.49	-1	13.40/13.29
1	3	1.67	20	2	0	0.84	7.72	11.39	6	19.84/20.01	-1	17.24/17.26
1	1	1.67	2	2	0	0.84	3.88	1.74	2	4.16/4.17	-1	4.83/4.88
1	10	1.67	2	2	0	0.84	13.34	5.81	12	13.70/13.68	1	25.14/26.12
1	3	8.35	2	10	0	0.84	23.13	12.10	18	24.14/24.94	0	33.68/34.56
1	3	1.67	2	2	5	0.84	8.52	6.36	15	9.77/9.80	5	11.54/11.43
1	3	1.67	2	2	10	0.84	9.27	8.32	24	11.04/11.15	10	12.05/12.04

In Table 3.4, respectively 3.5, we present the results for heuristics for other distribution functions of the interarrival time between production instants, namely

$$C_2(x) = 1-e^{-2x}(1+2x) \tag{3.34}$$

and

$$C_3(x) = 1 - 0.7887e^{-1.5774x} - 0.2113e^{-0.4226x} \tag{3.35}$$

It is easy to check that both distributions have the same mean as the distribution that is used in Section 3.8 (namely mean one), whereas the variance for $C_2(x)$ is 0.5 and the variance for $C_3(x)$ is 2.

Table 3.5. Results for the simple heuristics with distribution function $C_3(x)$ for the interarrival time between production opportunities. For an explanation see Table 3.3.

a	b	λ	N	q	ℓ	ρ	CL	PL	capacity-oriented heuristic β	SCH	product-oriented heuristic β	SPH
1	3	1.67	2	2	0	0.84	19.34	7.35	17	20.16/20.44	1	28.25/30.21
1	3	0.5	2	2	0	0.25	2.04	2.33	0	2.93/2.93	-1	2.56/2.59
1	3	1.8	2	2	0	0.9	29.46	8.24	29	35.40/35.23	2	52.11/52.97
1	3	1.67	5	2	0	0.84	19.34	7.52	17	20.91/21.19	-1	34.13/35.47
1	3	1.67	20	2	0	0.84	19.34	13.25	17	27.77/28.13	-1	31.71/33.26
1	1	1.67	2	2	0	0.84	9.62	3.57	7	9.83/10.27	-1	12.02/12.44
1	10	1.67	2	2	0	0.84	33.38	13.33	31	34.42/36.52	4	64.05/66.84
1	3	8.35	2	10	0	0.84	82.06	30.53	71	83.95/85.94	6	116.87/120.83
1	3	1.67	2	2	5	0.84	19.67	9.09	25	20.47/21.77	6	27.63/29.16
1	3	1.67	2	2	10	0.84	19.97	10.52	34	20.72/23.00	11	25.92/28.11

As can be seen, the criterion to use the simple heuristic for which the corresponding lower bound is maximal, again chooses the best heuristic in each situation. This suggests that the criterion is insensitive for the underlying capacity process.
In situations where the highest lower bound is close to the performance of the heuristic that is indicated by the criterion, this heuristic is almost optimal (compare Section 3.8). Consequently, in Table 3.4, only for situations with large values of ℓ and N, the simple heuristics may be improved. These are the same situations as in the original model (see Table 3.2). In Table 3.5, only in situations with large values of N and in the situation with λ=1.8, the simple heuristic may be improved.

We also want to mention the influence that changing the second moment of the distribution of the interarrival times has on the choice between the capacity-oriented and the product-oriented approach. If we compare the results of Table 3.2, 3.4 and 3.5, we see that a higher variance makes it more attractive to use a capacity-oriented approach. This seems reasonable, since the capacity-oriented approach takes the capacity-restriction explicitly into account.

3.9.3 Sensitivity for the distribution of the demand-size (S).

For the specific model in Section 3.6, we have assumed that S=1. To check whether the assertions for the specific model depend on this assumption, in this Section, we will consider a compound Poisson process. The only difference from the model of Section 3.6, is that S will be stochastic now. More explicitly, we have chosen the following distribution for S (notice that ES=1):

$$S = \begin{cases} 0 & \text{with probability } 2/3 \\ 3 & \text{with probability } 1/3 \end{cases}$$

The results for this model are given in Table 3.6.

As we see, the simple capacity-oriented heuristic performs best when

CL > PL. Otherwise, the product-oriented heuristic performs best. Therefore the criterion seems insensitive to the distribution of S. Also the performance of the best heuristic is, in most situations, close to its corresponding lower bound, which again indicates insensitivity.

Table 3.6. Results for the simple heuristics if the demand follows a compound Poisson process. For an explanation see Table 3.3.

a	b	λ	N	q	ℓ	ρ	CL	PL	capacity-oriented heuristic β	SCH	product-oriented heuristic β	SPH
1	3	1.67	2	2	0	0.84	21.72	8.03	10	23.00/21.76	1	29.17/30.28
1	3	0.5	2	2	0	0.25	2.50	2.62	-1	3.56/3.54	-1	2.88/2.86
1	3	1.8	2	2	0	0.9	36.45	8.88	19	36.46/37.32	1	60.29/58.25
1	3	1.67	5	2	0	0.84	21.72	8.34	10	25.48/24.23	-1	35.25/35.97
1	3	1.67	20	2	0	0.84	21.72	14.68	10	34.85/36.07	-1	35.63/36.65
1	1	1.67	2	2	0	0.84	10.34	3.71	4	10.57/10.87	-1	12.66/12.74
1	10	1.67	2	2	0	0.84	40.99	13.99	20	40.66/43.47	4	70.93/68.49
1	3	8.35	2	10	0	0.84	49.87	20.07	36	51.06/50.75	4	66.97/67.68
1	3	1.67	2	2	5	0.84	22.29	12.29	20	23.73/25.00	7	27.54/29.39
1	3	1.67	2	2	10	0.84	23.57	15.55	29	26.39/27.06	11	32.39/32.68

3.10 Discussion of Chapter 3 and preview of the next Chapters.

In this Chapter, we have compared the capacity-oriented and the product-oriented approaches for a single-phase model with identical products in which demand is purely stochastic.
It has proven that simple heuristics, that are based on the capacity-oriented approach or on the product-oriented approach perform well in most situations. These simple heuristics can be determined by analyzing one-dimensional models. This analysis leads also to approximations of the costs of these heuristics. Using these approximations, a criterion has been found to choose between the capacity-oriented and the product-oriented approaches. According to this criterion, the heuristic is chosen for which the approximation of the cost is maximal.
The heuristic that is indicated by this criterion performed well in the situations that have been investigated in this Chapter. As might be expected, the capacity-oriented approach performed best in case of a tight capacity-restriction, whereas the product-oriented approach became better as the number of products increased. It has also proven that the variance of the availability of the capacity played a role for the choice between the capacity-oriented and the product-oriented approaches. In case of a high variance, the capacity-oriented approach is more attractive.
We have finished this Chapter with a sensitivity analysis for the parameters of the general model with identical products and a purely stochastic demand.

In Chapter 4, we will consider a model where demand is partly known beforehand, whereas the stochastic processes that underlie demand are still the same for all products. In such a model, the products are no longer identical in the short-run. This leads to some difficulties for the capacity-oriented and the product-oriented approaches. Intertwined with the choice of the approach is the question of how the information that is available about demand is used in the planning. In Chapter 4, we will study some different ways to deal with this information.

In Chapter 5, we will consider a situation where the stochastic processes that underlie demand are different for the products. In such a situation, it will prove that the capacity-oriented approach has to be restricted to so-called fast-movers, whereas the slow-movers are controlled via a product-oriented approach.

We will finish this text by designing a Material Coordination System for a simple example of a plastic products factory, using the results of this text. In the plastic products factory, there will be only one capacity bottle-neck, which makes the relation with the research, presented in this text, more easy.

Chapter 4. Identical products; partly known demand.

4.1 Introduction of the model.

In this Chapter, we will consider a model in which demand is partly known beforehand. The model, that we consider, is similar to the model of the previous Chapter (see Figure 3.1). Again there are N products competing for the allocation of the same resource. Each product has the same demand-characteristics. In the model of this Chapter, however, the realisation of a part of the demand of each product is known some time in advance. Since these realisations may differ, it is necessary to treat the products in the short-run as though they were not identical. Since it is more natural to introduce forecasts for the demand in a periodic review model, we have chosen to analyze such a model in this Chapter.
For the purpose of investigating the effect of different forecasts for the demand of different products on the results of the previous Chapter, it is not so important how these forecasts are generated. It is only important that (due to different forecasts) the products cannot be treated as identical in the short-run. Therefore, we have chosen the demand of product j in period t, say $D_j(t)$, to consist of two parts as denoted in (4.1).

$$D_j(t) = U_j(t) + K_j(t) \tag{4.1}$$

The realisation of $U_j(t)$ is supposed to be known at the end of period t, whereas the realisation of $K_j(t)$ is known already at the end of period t-T. We assume that $U_j(t)$ has a discrete distribution, which is independent of j and t. The probability that $U_j(t)$ equals x will be denoted by $\eta(x)$, for x = 0, 1, 2, ... Analogously, we have a discrete distribution $\{\kappa(x)\}_{x=0}^{\infty}$ for the $K_j(t)$'s. We assume that there are no correlations between demand realisations for different products, between different periods or between $U_j(t)$'s and $K_j(t)$'s.

Analogous to Chapter 3, we define the N-vectors $\underline{U}(t)$ and $\underline{K}(t)$ as follows (compare (3.7)):

$$\underline{U}(t) := (U_1(t),\ U_2(t),\ \ldots,\ U_N(t))$$
$$\underline{K}(t) := (K_1(t),\ K_2(t),\ \ldots,\ K_N(t))$$

At the start of each period, we have the opportunity to start a production run for at most one product. Production is in batches of size q. A batch, that is started at the start of period t, occupies the capacity only during period t, and will arrive at the stock-point (of the corresponding product) in period t+ℓ (for an interpretation of this flowtime ℓ see Figure 3.3)

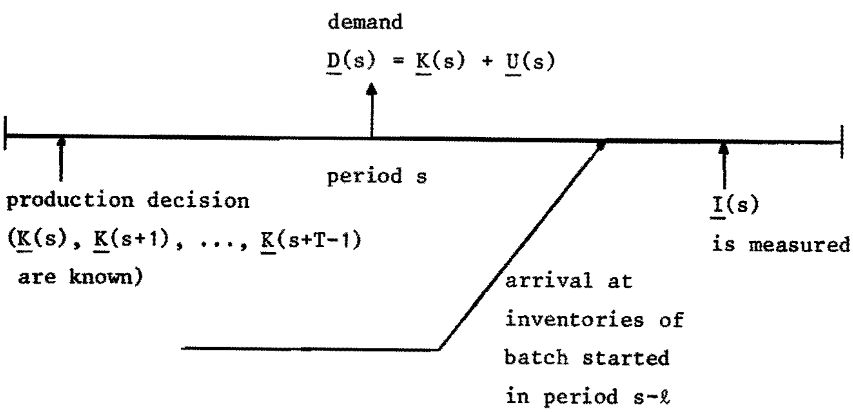


Figure 4.1. Schematic notation of the sequence of events in period s.

In Figure 4.1, we have given a scheme of events, occurring in an arbitrary period s, which may help to understand the model.

In this Chapter, the costs are reviewed periodically instead of continuously as in the previous Chapter:
Let $I_j(t)$ denote the inventory level of product j at the end of period t. With a given realisation of $I_j(t)$, say i, a cost p(i) is incurred. This p(i) has again the same characteristics as in Chapter 3.
To simplify the analysis, we will assume that the cost that corresponds to $I_j(t)$ is incurred in period t+1.

The reason to consider this model is that we want to compare the capacity-oriented and the product-oriented approaches in a situation where the products are not identical in the short-run. However, the choice of the approach may depend on the way the information about future demand is used. Therefore, we will discuss several methods to deal with this information. For each of these methods, we will describe the capacity-oriented and the product-oriented approaches. To avoid that the topic of comparing the capacity-oriented and the product-oriented approaches gets lost, due to the separate discussion of different ways to deal with the information that is available about future demand, we will overview the results in Section 4.6.4.

In Sections 4.2 to 4.5, we will describe approaches that lead to feasible strategies for this periodic review model. To avoid complications with notations, we will describe the approaches at hand of the situation with $\ell=0$, which means that if a production run is started at the start of period t, the batch will arrive at the stock-point during the same period. As we have seen in Chapter 3, it is straightforward to extend the description to situations with $\ell>0$, by minimizing the "shifted cost", which is the expected cost ℓ periods later. In this situation, where demand is partly known beforehand, this will lead to expectations that are conditioned on the realisations of the k_j's, but the same approach still works. We will return to this in Section 4.4.
In Section 4.6, we will give numerical results for a specific periodic review model and we will draw some conclusions.

4.2 Purely stochastic approach.

At the start of period t, the realisations $\underline{k}(t)$, $\underline{k}(t+1)$, ..., $\underline{k}(t+T-1)$ are known. It is possible, to use this information about the future behaviour of demand when making the production decision in period t. In the purely stochastic approach, however, this information is ignored. This will lead to a relatively simple approach, especially if we restrict ourselves to so-called capacity-oriented and product-oriented strategies (as we did for an analogous continuous review model - see Chapter 3). In separate Sections, we will describe both types of strategies for this periodic review model.

4.2.1 Capacity-oriented strategies.

In order to be able to define the capacity-oriented strategies, in this periodic review model, we introduce the aggregate inventory at the end of period t, $I(t) := \sum_{j=1}^{N} I_j(t)$, as a measure for the amount of capacity that is stored in the system.
For capacity-oriented strategies, the decision whether or not to start a production run at the start of period t, depends only on the realisation of the aggregate inventory at the end of the preceding period ($i(t-1)$). A production run will be started if and only if the aggregate inventory is less than or equal to a certain critical level, say β. The production run will then be assigned to the product with minimal inventory (if there are several products with minimal inventory we will choose one of them randomly).

As we have seen in Chapter 3, it is difficult to find an optimal capacity-oriented strategy, since it requires insight in the distribution of the aggregate inventory over the individual products. Consequently, the problem of determining this optimum becomes N-dimensional. To avoid this, we will use an approximation of the distribution of the aggregate inventory over the individual products. In Chapter 3, we have used the approximation that is based on the assumption that it is always possible to keep the inventories of individual products equal. Since this approximation performed well, we will use it here again. However, other approximations would be possible (as a matter of fact, we will suggest another approximation in Section 4.6.3).

With the approximation that the inventories can be kept equal, we can replace the cost $\sum_{j=1}^{N} p(i_j)$ by $N \cdot p(\sum_{j=1}^{N} i_j/N)$.

Using this approximation, we will determine an optimal strategy (the simple capacity-oriented heuristic). Therefore, notice that $\{I(t),\ t=1,\ 2,\ \ldots\}$ is a Markov Chain with countable state space, under application of any capacity-oriented strategy β. The transition probabilities in this Markov Chain depend on the choice of β and on the distribution of the aggregate demand over one period. Consequently, it is possible to calculate the transition probabilities $ra_\beta(k,h)$ that the aggregate inventory goes from k at the start of one period to h at the start of the next.

We will introduce the functions $t_\beta(k)$ and $c_\beta(k)$ as the expected time, respectively the expected (approximate) cost until the aggregate inventory equals β, under application of the capacity-oriented strategy with critical level β, if the aggregate inventory equals k at the end of a given period. We assume that these variables exist and that they are finite for all k. Then it is easy to see that the following relations hold (by definition $t_\beta(\beta)$ and $c_\beta(\beta)$ are set equal to zero):

$$\begin{cases} t_\beta(k) = 1 + \sum_h ra_\beta(k,h) \cdot t_\beta(h) & \text{for } k \neq \beta \\ t_\beta(\beta) = 0 & \end{cases} \tag{4.2}$$

and

$$\begin{cases} c_\beta(k) = N \cdot p(k/N) + \sum_h ra_\beta(k,h) \cdot c_\beta(h) & \text{for } k \neq \beta \\ c_\beta(\beta) = 0 & \end{cases} \tag{4.3}$$

After truncating the state-space, we can determine the solution of these one-dimensional relations, for example by means of successive approximation. It is easy to see that $t_\beta(k) = t_{\beta+1}(k+1)$, so that we have to solve (4.2) only for one choice of β, to find the solution for all choices of β.

If the approximate, expected cost per period is finite, under application of the capacity-oriented strategy denoted by β, this average cost can be denoted as in (4.4). Notice that (4.4) represents the quotient of the total expected cost and the expected time between successive visits to β.

$$\frac{N \cdot p(\beta/N) + \sum_h ra_\beta(\beta,h) \cdot c_\beta(h)}{1 + \sum_h ra_\beta(\beta,h) \cdot t_\beta(h)} \tag{4.4}$$

Since only the numerator of (4.4) depends on β, we only have to solve (4.3) for several choices of β to find an optimal choice under the approximate cost. The strategy that we find if minimizing this approximate, average cost per period, will be referred to as the simple capacity-oriented heuristic. We will denote this heuristic by the choice of β that it corresponds to, say βc. As we have seen in Chapter 3, the approximate, average cost per period for the simple capacity-oriented heuristic, is a lower bound for the actual cost of any strategy. Therefore, we will denote the approximate cost that corresponds to this heuristic by CL (Capacity Lower bound). The actual cost of using this heuristic will be higher. We will denote this cost by SCH (Simple Capacity-oriented Heuristic).

We will return to numerical results for both the simple capacity-oriented heuristic and the optimal capacity-oriented strategy in Section 4.6.

4.2.2 Product-oriented strategies.

Product-oriented strategies in the purely stochastic approach are found by decomposing over the products. Since in the purely stochastic approach, the products are identical, a production run will be started if and only if the minimal inventory has dropped below a certain critical level, say β. As for the capacity-oriented strategies, a production run will be assigned to the product with minimal inventory.

To find the optimal product-oriented strategy, we have to take into account the possible delay in the availability of the capacity due to other products. For reasons, that have been explained in Chapter 3, we search for simple approximations of this delay. The simplest approximation, we find by assuming that the products never interfere with each other on the capacity, so that there is no delay in the availability of the capacity. In Section 4.6.3, we will also discuss another approximation, but seeing the results of this simplest approximation in the continuous review model (see Chapter 3), it seems reasonable to use it here as well. The resulting strategy will be referred to as the simple product-oriented heuristic. The critical level that this simple product-oriented heuristic conforms to, will be denoted by βp. As we have discussed in Chapter 3, the cost that we find in the model under the assumption that the products never interfere on the capacity, will be a lower bound for the actual cost of any strategy. Therefore, we will denote this approximate cost by PL (Product Lower bound). The actual cost of using the simple product-oriented heuristic will be higher. We will denote this actual cost by SPH (Simple Product-oriented Heuristic).

Notice that in order to find the simple product-oriented heuristic, the same approach as for the simple capacity-oriented heuristic can be used, except that we now consider the Markov Chain
$\{I_j(t),\ t=1,\ 2,\ \ldots\}$.
In Section 4.6, we will return to numerical results for both the optimal product-oriented strategy and the simple product-oriented heuristic.

4.3 Introduction of rolling schedules.

Now we will consider the strategies that we find if we use the information that is available about future demand. Consequently, the

production decision at the start of period t, will be based on the following realisations: $\underline{i}(t-1)$, $\underline{k}(t)$, $\underline{k}(t+1)$, ..., $\underline{k}(t+T-1)$.
The strategies that we consider, can be written as a function $\pi(\underline{i}(t-1), \underline{k}(t), \underline{k}(t+1), \ldots, \underline{k}(t+T-1))$, where $\pi(\underline{i}(t-1), \underline{k}(t), \underline{k}(t+1), \ldots, \underline{k}(t+T-1))=r$ corresponds to the decision to start a production run for product r, if $1 \leq r \leq N$, or to the decision not to start a production run if r=0.

Notice that $\{[\underline{I}(t-1), \underline{K}(t), \underline{K}(t+1), \ldots, \underline{K}(t+T-1)], t=1, 2, \ldots\}$ is a Markov Chain with countable state space under application of any strategy π. The transition probabilities in this Markov Chain can be determined, straightforwardly, since the distribution of $\underline{U}(t)$ and $\underline{K}(t+T)$ is given. This makes it, theoretically, possible to determine an optimal strategy. However, even for small values of T and N, the numerical problems, inherent to determining an optimal strategy, make this approach impossible. Even after aggregation or decomposition over the products, the state of the system is described by a (T+1)-dimensional vector. For larger values of T this will present us with severe numerical problems. Indeed, one usually applies a completely different approach in such a situation. This is the so-called rolling schedule approach. In such an approach, a production plan is formed by solving a multi-period problem and implementing only the production decision for the first period (see Baker [5]). The next period, the variables are updated and the procedure is repeated.
The standard way to determine the production plan (each period anew) in a rolling schedule approach, is by minimizing the total costs over a finite planning period (see Baker [5], McClain and Thomas [39], Morton [41], Nuttle and Wijngaard [43] and Zabel [60]).
Notice that in the rolling schedule approach, the tuple $(\underline{i}(t-1), \underline{k}(t), \underline{k}(t+1), \ldots, \underline{k}(t+T-1))$ is given and the schedule is constituted as though no new realisations of the $\underline{K}$'s will become available. This makes it more simple than analysing the Markov Chain that we have mentioned above.

Let Thor denote the horizon over which we want to minimize the costs, when determining a production schedule. Remark that the planning horizon need not be equal to the horizon over which demand is partly known (i.e. Thor $\neq$ T). In case T > Thor, more realisations of the K-part of demand are available than will be used in determining the production plan (if ℓ=0). In case T<Thor, no realisations are available for K-part of demand in the last periods of the planning period. In our examples, we will restrict ourselves to the situation where Thor = T, so that the length of the planning period and the horizon over which demand is partly known, are in balance. The reason for restricting ourselves to this situation, is that it is our aim to consider the influence a partly known demand has on the choice between capacity-oriented and product-oriented strategies. Therefore, we are not so much interested in varying T and Thor independently. However, when describing different strategies, we will not require that Thor = T, so that it is clear how to use this approach in situations with Thor ≠ T.

Notice what happens if we apply such a rolling schedule approach, in case we would not account for any costs after the planning period. The planning will then be based on the idea that there is no sense in keeping inventory after Thor, because p(0) = 0. Consequently, the solution will force us to start selling our safety-stock as we approach the horizon of the schedule. In case the capacity is infinite and Thor is large, this effect may be of minor importance, but as the capacity becomes more tight, the depletion of (capacity) safety-stock may have long-lasting effects. To avoid this problem, one might increase the planning period, so that there is hardly any influence on the first period decisions. However, extending the planning period will increase the complexity of the problem that has to be solved each period anew. Therefore, instead of extending the planning period, we prefer to use a short planning period but we add an extra component to the cost in the final period (see Baker [5]). This "final cost function" has to denote the preference for ending with enough safety-stock for the future (after the horizon).
If no final cost function is used, even in case the capacity restriction is not tight, it may be decided not to produce a batch because the inventory holding costs are too high in the short-term, while in fact such a batch could have been required. However, in case of an excess of capacity, a deficit of stock will only have short-term effects. Even with a short planning horizon, the decision not to start such a production run at the end of the planning period, has little effect on the first period decisions (which will be implemented). The effects of postponing the start of a production run are more severe in case of a tight capacity restriction. Therefore, we will only use a

final cost function in combination with a capacity-oriented approach. This final cost function should denote the preference for ending with enough capacity buffered in the inventories at the end of the horizon. Therefore, the final cost function has to depend only on the aggregate inventory at the end of the planning period. To keep the analysis simple, we will base the final cost function on a purely stochastic approach. We define the final cost function as (for the definition of the variables, see the previous Section):

$$p_f(k) := c_{\beta c}(k) - CL \cdot t_{\beta c}(k) \tag{4.5}$$

where k is the aggregate inventory at the end of the planning period. Notice that via definition (4.5), the final cost function denotes the total deviation from the approximate, average cost per period until the aggregate inventory reaches βc, under application of the simple capacity-oriented heuristic. Thus the final cost function is a relative measure for the capacity problems that may be expected if the aggregate inventory at the end of the planning period is given.
We will return to numerical results, using this final cost function, in Section 4.6.

Now, let us return to the problem of finding a production plan over the horizon Thor. There are two, fundamentally different, approaches to determine a production plan, namely:

-the Stochastic Dynamic Programming approach
-the Deterministic Dynamic Programming approach.

The difference between the two approaches is that in the Stochastic Dynamic Programming approach, one accounts for the stochastic behaviour of the demand, whereas in the Deterministic Dynamic Programming approach, it is assumed that the demand can be forecasted accurately. We will describe both approaches separately in Sections 4.4 and 4.5.

4.4 The Stochastic Dynamic Programming approach.

As we have mentioned in Section 4.1, the cost incurred in period t depends on the realisation of the inventories at the end of period t-1. Consequently, we can picture the costs that we must minimize if we construct a production plan for the planning period as in Figure 4.2.

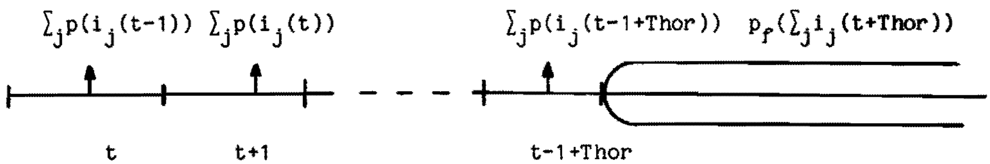


Figure 4.2. The total costs over the planning period.

For the Stochastic Dynamic Programming approach, the stochastic behaviour of demand is taken into account. If we want to take the stochastic behaviour of demand into account, the determination of a production plan over the planning period still requires that we consider a large state space. Therefore, we will restrict ourselves to product-oriented and capacity-oriented strategies for the Stochastic Dynamic Programming approach.

4.4.1 Product-oriented strategies.

As we have seen, product-oriented strategies require decomposition over the products. Consequently, the decision whether or not to start a production run for a given product j, is determined on the status of that individual product.
Therefore, in a product-oriented approach, first for each individual product a production plan is determined over the whole horizon. As usual in a rolling schedule approach, only the first period decisions are implemented. However, it is not allowed to start more than one production run per period. Consequently, we need a rule that determines

which product will be produced if there are several products that require such a run.
First, we will describe how a production plan per product is determined in such a product-oriented approach and next we will discuss the form of the "composition rule", on which we decide how to combine the production plans in the first period. Note that we have mentioned in Section 4.3, that no final cost function will be used in combination with a product-oriented approach.

We will restrict ourselves now to a specific product j and given realisations $k_j(t)$, $k_j(t+1)$, ..., $k_j(t+T-1)$.
Suppose that the inventory of product j (as projected in executing the Deterministic Dynamic Programming plan) equals i at the start of period t+s (i.e. $I_j(t+s-1) = i$). Define the total expected costs from period t+s until the end of the planning period (see Figure 4.2), under application of a strategy that minimizes the total cost over this period, as $v_p^s(i)$.
Introduce the transition probabilities $rd_j^s(i,h,0)$ and $rd_j^s(i,h,1)$ as the probability of a transition of the inventory for product j from i at the end of period t+s-1 to h at the end of period t+s, if we do not produce product j, respectively if we do produce product j (s = 0, 1, ..., Thor-1). These transition probabilities can be determined straightforwardly, since the realisations $k_j(t)$, $k_j(t+1)$, ..., $k_j(t+T-1)$ are given.

Using these transition probabilities it is easy to see that the $v_p^s(i)$'s satisfy the following recurrence relations:

$$\begin{cases} v_p^s(i) = p(i) + \\ \qquad \min\left[\sum_h rd_j^s(i,h,0)\cdot v_p^{s+1}(h),\ \sum_h rd_j^s(i,h,1)\cdot v_p^{s+1}(h)\right] \\ \qquad\qquad\qquad \text{for } s=0,1,\ldots,\text{Thor}-1 \\ v_p^{\text{Thor}}(i) = 0 \end{cases} \tag{4.6}$$

Since the above recurrence relations are one-dimensional, they can be solved (after truncating the state-space) without much computational effort e.g. by backward recursion.

In Section 4.1, we have announced that we would discuss the strategies in this Chapter for the situation where $\ell=0$, because the extension to situations with $\ell>0$ is simple. At hand of the product-oriented strategies in the Stochastic Dynamic Programming approach, we want to demonstrate this extension. Therefore, we introduce the "conditional shifted costs" $L_T^s(ip_j \mid k_j(s), k_j(s+1), \ldots, k_j(T))$ as the expected cost in period $s+\ell$ if the inventory position of product j equals ip_j at the start of period s, conditioned on the realisations of future demand that are available. Note that it is straightforward to calculate these shifted costs since the distributions of the U_j's and K_j's are known. By using these conditional shifted costs in (4.6) instead of p(i), we find a production plan for $\ell>0$.

Above, we have shown how we can determine a production plan for each product. Now, we will discuss how to compose the production plans for different products in the first period. To find an optimal allocation of the production run in the first period, we should consider the coordination of the products over the whole horizon. However, this leads to a problem that is again as complex as the original problem without decomposing over the products. To avoid the numerical problems that are inherent to such an approach, we will only consider simple coordination rules. Such a simple rule is the "Myopic rule". This Myopic rule has the following form:

Myopic rule: Consider only the first period in which the allocation of the production run influences the cost. Assign the run to the product for which the expected inventory is minimal then. In case $\ell=0$, this means that the production run is assigned to the product for which $i_j(t-1)-k_j(t)$ is minimal (if $\ell>0$, more realisations of the k_j's and information about production runs that already have been started, will be required). Notice that if T=0, so $k_j(t)$ has not realised yet, we find the same allocation rule as in the purely stochastic approach.

To check how well this rule performs, we have also considered a more sophisticated composition rule. This rule is the "Value Function rule" and can be described as follows:

Value Function rule: In constructing the production plan, we have determined the Value Function $v_p^s(i)$, that denotes the expected costs from period t+s on. Using this, we can express the cost that is incurred if we do not produce a given product j now, by:

$$\sum_h rd_j^0(i_j(t-1),h,0)\cdot v_p^1(h) - \sum_h rd_j^0(i_j(t-1),h,1)\cdot v_p^1(h) \qquad (4.7)$$

Therefore, the Value Function rule assigns a production run to the product for which (4.7) is maximal. An analogous approach has been proposed by van Beek [6] and Wijngaard [57].

We have compared the two composition rules, numerically, for a specific situation, namely:

1) The known part of the demand ($K_j(t)$) has a Poisson distribution with parameter μ.

2) The unknown part of the demand ($U_j(t)$) has a Poisson distribution with parameter λ.

3) The cost function is chosen as $p(i) := ai^+ + bi^-$.

Notice that the utilization rate of the capacity is given by $\rho := N\cdot(\lambda+\mu)/q$.

Simulation results for both allocation rules are given in Table 4.1. As the results indicate there are only small differences between the two allocation rules. The only exception is the last situation, but there the utilization rate is very high, so that it is unlikely that one will use a product-oriented approach at all (we will return to this in Section 4.6). Therefore, we will use the Myopic rule in combination with product-oriented strategies.

Table 4.1. Comparison of the Myopic rule and the Value Function rule to allocate a production run for the product-oriented strategy in the Stochastic Dynamic Programming approach (Thor=T).

q	ℓ	N	a	b	λ	μ	Thor	ρ	Myopic	Value Function
12	0	2	1	3	2.5	2.5	2	0.83	11.86/11.99	11.89/12.01
12	0	2	1	3	2.5	2.5	3	0.83	11.93/12.04	11.93/11.81
12	0	2	1	3	2.5	2.5	5	0.83	11.95/11.95	12.01.11.86
12	0	2	1	3	0	5	2	0.83	10.75/10.71	10.58/10.66
12	0	2	1	3	5	0	2	0.83	12.20/12.26	12.30/12.28
20	0	2	1	3	2.5	2.5	2	0.5	15.86/15.96	15.88/15.87
12	0	5	1	3	1	1	2	0.83	27.60/27.36	27.40/27.35
12	0	10	1	3	0.5	0.5	2	0.83	48.74/48.99	48.86/48.89
12	0	2	1	1	2.5	2.5	2	0.83	7.47/7.47	7.49/7.40
12	0	2	1	10	2.5	2.5	2	0.83	18.66/18.66	18.62/18.44
12	5	2	1	3	2.5	2.5	2	0.83	16.88/16.85	16.94/16.83
12	10	2	1	3	2.5	2.5	2	0.83	21.35/21.38	21.15/21.20
12	0	2	1	3	0.5	0.5	2	0.17	9.05/9.05	9.11/9.12
12	0	2	1	3	2.75	2.75	2	0.92	17.07/16.75	16.83/17.08
12	0	2	1	3	2.9	2.9	2	0.97	41.98/41.98	37.25/34.69

Explanation of Table 4.1: We have used (4.6) to find a production plan for each product and next we have used one of the two composition rules to allocate the run to the products. Notice that if the planning horizon T=Thor=1, then we cannot apply the product-oriented approach, since $v^1(k)$ equals zero then. Therefore, we have only considered situations with Thor≥2.

To investigate the sensitivity of the results for the specific simulation run, we have simulated two different production runs. The results of both runs are given in this Table, namely as: (results first run)/(results second run).

4.4.2 Capacity-oriented strategies.

As we have seen in the previous Chapter, a capacity-oriented strategy requires aggregation over the products. Consequently, the decision whether or not to start a production run depends on the realisation of the aggregate inventory. Before we discuss a method to constitute an aggregate production plan that indicates whether or not to start a production run, first we will discuss a rule that indicates to which product a production run in the first period should be assigned. As in the case of product-oriented strategies, it is difficult to find an optimal rule to allocate the production run in the first period, since this requires some insight in the distribution of the aggregate inventory over the individual products. For reasons of symmetry, we will use the same Myopic rule as we did for the product-oriented strategies. Note that using a Value Function rule in this situation would be more difficult, since the production plan has only been determined in aggregate terms.

The aggregate production plan will be based on the aggregate realisations: $i(t-1)$, $k(t)$, $k(t+1)$, ..., $k(t+T-1)$
With an aggregate inventory, we have to connect a certain cost. As we have discussed for the purely stochastic approach, a reasonable approximation, which we will use, seems to be to connect the cost $N \cdot p(i/N)$ to an aggregate inventory i.

Under this approximation, we can determine analogous recurrence relations to (4.6) for an aggregate cost function. Since this approach follows straightforwardly from the approach in 4.4.1, we will not discuss it here.

In Section 4.6, we will consider numerical results for both the capacity-oriented and the product-oriented strategies for a specific periodic review model.

4.5 The Deterministic Dynamic Programming approach.

In the Stochastic Dynamic Programming approach, we account for the stochasticity of the demand. The determination of a production plan can be simplified considerably by assuming that the demand can be forecasted accurately. Such an approach has for example been used by Billington, McClain and Thomas [11], Bitran and Hax [13], Gabbay [22], and Hax and Meal [28]. Because of its simplicity, this approach is often used in practical situations, compare e.g. the Material Requirements Planning approach (see Orlicky [44]).

Under some conditions, it can be proven that such an approximation yields an optimal strategy (certainty equivalence - see Holt et al. [30]). However, then it is required that the cost function is quadratic and that the transitions of the inventory are linear (this is disturbed in case the production level is restricted from below by zero or from above by a finite capacity). Consequently, in case the cost function is not quadratic or in case the capacity restriction plays a role, this approach may not be optimal. Still, the approach is often used, because it is relatively simple to determine a production plan using this approach.

For given realisations of $\underline{k}(t)$, $\underline{k}(t+1)$, ..., $\underline{k}(t+T-1)$, we must determine forecasts for the demand from period t to t+Thor-1, which will be assumed to be perfect if determining a production plan over the planning period.
For $t \leq s \leq t+T-1$, it seems reasonable to add the expectation of $U_j(s)$, say $EU_j(s)=u$, to the known part of demand. To make sure that the forecasts for future demand are integer, we have to round u off. Since always rounding off downwards or always rounding off upwards would give forecasts that systematically deviate, we must sometimes round off

upwards and sometimes downwards instead. Therefore, we introduce the following mechanism to determine forecasts $\delta_j(s)$, where [u] stands for the entier of u (i.e. the largest integer that is less than or equal to u):

$$\delta_j(s) := \begin{cases} [u] + 1 + k_j(s) & \text{with probability } u-[u] \\ [u] \quad\; + k_j(s) & \text{with probability } 1-(u-[u]) \end{cases} \quad \text{for } t \leq s \leq t+T-1 \tag{4.8}$$

Analogously, we determine forecasts after t+T-1, by rounding of $EU_j(s)+EK_j(s)$. Notice that this plays a role if Thor > T.

What makes the Deterministic Dynamic Programming approach relatively simple, is that the inventory of product j at the end of period t+s, can only be in a limited number of states (see van Beek [6]). Since we cannot start more than s+1 production runs, these possible states are:

$$i_j - \sum_{r=t}^{t+s} \delta_j(r),\ i_j - \sum_{r=t}^{t+s} \delta_j(r)+q,,\ \ldots,\ i_j - \sum_{r=t}^{t+s} \delta_j(r)+(s+1)\cdot q$$

The approach that has been described for the Stochastic Dynamic Programming approach can be used here, but now it is much simpler since the transition probabilities are either 0 or 1 and the state space is much smaller. This enables us not only to derive capacity-oriented and product-oriented strategies, but also to determine an optimal production plan (over the planning period), for situations where the number of products is not too large.

In Section 4.6, we will consider numerical results for the Deterministic Dynamic Programming approach. We will compare the influence of using the final cost function on the performance of the optimal strategy. We will also consider capacity-oriented and product-oriented strategies for this Deterministic Dynamic Programming approach.

4.6 Numerical results and conclusions.

In this Section, we will compare the different approaches that have been described in the previous Sections. To be able to derive numerical results for the proposed strategies, we have chosen more specific distributions for the two components of the demand. These are the same choices as have been used for the comparison of the composition rules in Section 4.4:

For the unknown part, $U_j(t)$, we have chosen a Poisson distribution with parameter λ. For the known part, $K_j(t)$, we have chosen also a Poisson distribution but this one has parameter μ (consequently, the utilization rate of the capacity is $\rho := N\cdot(\lambda+\mu)/q$).

The cost that is incurred for a product if the realisation of the inventory at the end of a period equals i, is chosen to be $p(i) := ai^{+}+bi^{-}$. Finally, we have restricted ourselves to the situation with T = Thor.

Most of the results in this Section are determined by means of simulation. For each parametersetting, we generated two different simulation runs. In the Tables, the costs for the strategies, that we consider, are given for both simulation runs.

4.6.1 Purely stochastic approach.

For the simple capacity-oriented heuristic and for the simple product-oriented heuristic, that are derived in case demand is purely stochastic, we give simulation results in Table 4.2. In the purely stochastic approach, the demand in each period is Poisson distributed with parameter $\nu := \lambda+\mu$.

Table 4.2. Simple capacity-oriented and product-oriented heuristics in the purely stochastic approach, with $\nu=\lambda+\mu$.

q	ℓ	N	a	b	ν	ρ	capacity-oriented heuristic CL	SCH	βc	product-oriented heuristic PL	SPH	βp
12	0	2	1	3	5	0.83	6.83	12.66/12.70	8	10.64	12.38/12.27	2
20	0	2	1	3	5	0.5	8.82	18.87/18.93	5	16.08	16.59/16.61	-1
12	0	5	1	3	2	0.83	6.83	30.60/30.64	8	24.22	28.20/28.17	-2
12	0	10	1	3	1	0.83	6.83	60.81/60.82	8	46.79	48.38/48.49	-2
12	0	2	1	1	2	0.83	4.21	7.54/7.52	5	7.17	7.59/7.55	-1
12	0	2	1	10	2	0.83	11.00	20.42/20.48	12	14.39	19.46/19.13	4
12	5	2	1	3	2	0.83	11.37	17.88/17.92	61	16.68	17.86/17.86	28
12	10	2	1	3	2	0.83	14.52	21.92/22.13	112	21.04	21.76/22.15	54
12	0	2	1	3	1	0.17	4.84	11.68/11.77	-2	9.36	9.32/9.28	-2
12	0	2	1	3	5.5	0.92	9.76	14.23/14.36	12	10.74	16.77/16.74	2
12	0	2	1	3	5.8	0.97	20.45	22.01/23.84	23	10.89	36.24/40.14	2

Explanation of Table 4.2: Using the relations that have been derived in Section 4.3, we can determine an approximation for the cost of using the capacity-oriented strategy (or the product-oriented strategy) that corresponds to any choice of β. By varying β, we have determined an optimal choice for β, which is given in the Table for both the capacity-oriented and the product-oriented approach. The corresponding approximate cost (CL for the capacity-oriented heuristic and PL for the product-oriented heuristic) is a lower bound for the cost of any strategy. The actual cost of using the proposed heuristics have been determined by means of simulation. This cost is denoted by SCH for the capacity-oriented heuristic and SPH for the product-oriented heuristic.

We see that the criterion to choose the heuristic which gives the highest lower bound (see Chapter 3), works well for this periodic review model too. An exception is the case with ν=5.5. The reason for this seems to be that the approximation of the cost, on which the lower bound is based, is not so good in this periodic review case as in the continuous review case, at least for the capacity-oriented heuristic. This can be understood if one realises that the assumption underlying the approximation in the capacity-oriented approach is that it is possible to keep all inventories equal. If a production run is started, we indeed allocate the run to the products so as to ensure this. However, the longer ago such a production run has been started, the more the inventories tend to diverge. In this periodic review model, the cost is incurred for the inventories at the end of each period, whereas the production decision is taken at the start of a period. Consequently, the imbalance between the inventories has a higher influence on the cost in the periodic review model than in the continuous review model.
That means that the difference between the highest lower bound and the cost of the best heuristic is larger in this periodic review model. This suggests that it may be sensible to introduce more advanced heuristics than the simple heuristics that are considered in Table 4.2, in this periodic review model. To check whether this is true, we have determined the optimal capacity-oriented and the optimal product-oriented strategies with corresponding costs by means of simulation. The results of these strategies are given in Table 4.3.
In order to make the comparison with Table 4.2 more easy, we have given the simulated costs for the best heuristic in Table 4.3 again (the P or C denotes whether this best heuristic is the Product-oriented heuristic or the Capacity-oriented heuristic).
Considering Table 4.3, there are two striking points:
Firstly, the performance of the optimal capacity-oriented and the optimal product-oriented strategy is about the same for almost all situations (an exception is the situation with ν=1, where the

Table 4.3. Optimal capacity-oriented and product-oriented strategies in the purely stochastic approach, with $\nu=\lambda+\mu$.

q	ℓ	N	a	b	ν	ρ	optimal capacity-orient. strategy simul. cost	β	optimal product-orient. strategy simul. cost	β	best simple heuristic simul. cost	
12	0	2	1	3	5	0.83	12.09/12.00	12	11.94/11.92	3	12.38/12.27	P
20	0	2	1	3	5	0.5	17.82/17.76	10	16.29/16.33	0	16.59/16.61	P
12	0	5	1	3	2	0.83	26.14/26.17	20	25.63/25.72	0	28.20/28.17	P
12	0	10	1	3	1	0.83	48.93/48.92	37	48.38/48.49	-2	48.38/48.49	P
12	0	2	1	1	2	0.83	7.54/7.52	5	7.42/7.38	0	7.54/7.52	P
12	0	2	1	10	2	0.83	16.79/16.50	18	16.86/16.81	6	19.46/19.13	P
12	5	2	1	3	2	0.83	17.34/17.43	63	17.32/17.53	29	17.86/17.86	P
12	10	2	1	3	2	0.83	22.04/21.43	116	21.32/21.69	56	21.76/22.15	P
12	0	2	1	3	1	0.17	10.59/10.59	2	9.32/9.28	-2	9.32/9.28	P
12	0	2	1	3	5.5	0.92	13.41/13.56	15	13.49/13.84	5	14.23/14.36	C
12	0	2	1	3	5.8	0.97	22.01/23.84	23	22.35/22.32	9	22.01/23.84	C

utilization rate equals 0.17). This suggests that in most situations it is not so important which of the two approaches is used, as long as one chooses a good strategy within the approach. How easy it is to find a simple strategy that performs good within a given class depends on the situation. It seems sensible to use the criterion that chooses the simple heuristic that gives the highest lower bound.
Secondly, in some cases there is a substantial difference between the best simple heuristic and the optimal strategy within the class of product-oriented and capacity-oriented strategies. Since it is possible to improve on the simple heuristics.

more advanced capacity-oriented heuristic.

When determining the simple capacity-oriented heuristic, we have assumed that it would be possible to keep all the inventories equal. This assumption leads to a simple approximation of the cost for a given aggregate inventory (see Section 4.3). The cost rate on a given aggregate inventory is then approximated by $N \cdot p(i/N)$ (which corresponds to the actual cost rate if all inventories equal i/N). In a more advanced capacity-oriented approach, we account for the fact that it will not be possible to keep the inventories equal. Instead, we now assume that the individual inventories, given a realisation i for the aggregate inventory, fluctuate uniformly around the mean i/N. Since the batch-size q plays a role in this imbalance between the inventories, we have chosen the following approximation of the cost rate on a given aggregate inventory i:

$$\tilde{g}(i) := \sum_{y=i/N - q/2}^{y=i/N + q/2} p(y) \cdot \frac{1}{q+1} \tag{4.9}$$

Of course, other approximations for the actual cost would be possible. For example, one could use an approximation that depends on the strategy that is used. However, the analysis may then become more complex. To keep the approach simple, we have used the approximation of the imbalance as in (4.9).

Since $\tilde{g}(i)$ depends only on the aggregate inventory, the relations of Section 4.3 can be used again to find the more advanced capacity-oriented heuristic, except that we have to replace the cost function $N \cdot p(i/N)$ by $\tilde{g}(i)$, but this will not make the analysis more difficult.

Table 4.4. More advanced capacity-oriented heuristic in the purely stochastic approach, with $\nu=\lambda+\mu$ (for an explanation see Table 4.2 except that (4.9) is now used as cost function).

q	ℓ	N	a	b	ν	ρ	more advanced capacity-oriented heuristic appr	simul	ß	simple cap.-or. heurist. SCH	optimal cap.-or. strategy simul
12	0	2	1	3	5	0.83	11.16	12.09/12.00	12	12.66/12.70	12.09/12.00
20	0	2	1	3	5	0.5	17.11	17.82/17.76	10	18.87/18.93	17.83/17.89
12	0	5	1	3	2	0.83	23.80	26.17/26.29	21	30.60/30.64	26.14/26.17
12	0	10	1	3	1	0.83	46.43	48.93/48.92	37	60.81/60.82	48.93/48.92
12	0	2	1	1	2	0.83	7.14	7.54/7.52	5	7.54/7.52	7.54/7.52
12	0	2	1	10	2	0.83	15.15	16.88/16.56	17	20.42/20.48	16.79/16.50
12	5	2	1	3	2	0.83	16.46	17.81/17.34	64	17.88/17.92	17.34/17.43
12	10	2	1	3	2	0.83	20.43	22.04/21.43	116	21.92/22.13	22.04/21.43
12	0	2	1	3	1	0.17	10.02	10.59/10.59	2	11.68/11.77	10.59/10.59
12	0	2	1	3	5.5	0.92	13.09	14.08/13.59	16	14.23/14.36	13.41/13.56
12	0	2	1	3	5.8	0.97	22.62	23.44/23.10	26	22.01/23.84	22.01/23.84

In Table 4.4, we give numerical results for this more advanced capacity-oriented heuristic. For the more advanced heuristic, we have not only given the simulated costs, but also the approximate cost that follows from the analysis of the theoretical model, using approximation (4.9) of the cost rate on the aggregate inventory. Note that (contrary to CL) the theoretic cost not necessarily has to be a lower bound.
To be able to set the results of the more advanced heuristic in a wider context, we have also given the results for the simulation runs of the

simple capacity-oriented and the optimal capacity-oriented strategy in Table 4.4

As we see, the more advanced heuristic is much better than the simple capacity-oriented heuristic in case of high stock-out costs (b=10). For the case of a strict capacity, the more advanced heuristic gets worse.

more advanced product-oriented heuristic.

The simple product-oriented heuristic has been derived under the assumption that the products never interfere on the capacity. In reality, the products compete for allocation of capacity. This may lead to a delay in the availability of the capacity for an individual product. To find a more advanced product-oriented heuristic, we estimated the size of this delay. This delay-time is then treated as a deterministic leadtime that will be added to the actual leadtime, ℓ, to find the so-called adjusted leadtime. Such an approach has been proposed by Graves [24] and Williams [55].
If we want to estimate the size of the delay, we have to realise that subsequent delays for the same product are not independent and that there is a correlation between the delays for the different products. However, the reason for introducing this more advanced product-oriented heuristic is only to see whether it is possible to improve the simple product-oriented heuristic in a simple way. Therefore, in order to keep the model, that we have to analyze, tractable, we neglect these correlations. Instead, we assume that there is a constant probability α that a product triggers a production run in an arbitrary period, whereas with probability $1-\alpha$ no production run is triggered for that product. To ensure that the utilization rate of the capacity is correct, we have to choose $\alpha = \nu/q$.

To estimate the size of the delay, under this approximation, we introduce a Markov Chain with state space $\{r : r \geq 0\}$, where state $r=0$ corresponds to the capacity being idle and state r $(r>0)$ corresponds to the situation with r unfinished orders (for all products together). That means that there is one order in process and $r-1$ orders are queueing for the capacity.

Transitions between the states in this Markov Chain take place at the end of each period. If $r>0$ then one order will be finished in a period. So, if no new orders are generated in the period, a transition occurs to state $r-1$. This happens with probability $(1-\alpha)^N$. With probability $p_k := \binom{N}{k}\alpha^k(1-\alpha)^{N-k}$ exactly k new orders are generated in a period (for $0 \le k \le N$), which conforms to a transition from state r to state $r+k-1$. Notice that the probability of a transition from state 0 to any state j is the same as the probability of a transition from state 1 to this state j.

Define π_r as the probability to be in state r (in the steady-state). Then the following transition equations can be derived for π_r.

$$\begin{cases} \pi_0 = p_0(\pi_0+\pi_1) \\ \pi_1 = p_1(\pi_0+\pi_1) + p_0\pi_2 \\ \pi_2 = p_2(\pi_0+\pi_1) + p_1\pi_2 + p_0\pi_3 \\ \quad\vdots \\ \pi_N = p_N(\pi_0+\pi_1) + p_{N-1}\pi_2 + \cdots + p_1\pi_N + p_0\pi_{N+1} \\ \\ \pi_r = \sum_{k=0}^{N} p_k\pi_{r+1-k} \qquad \text{for } r \ge N+1 \end{cases} \tag{4.10}$$

Starting from the upper equation and going down, we can expres all π_r's in terms of π_0. Since the utilization rate of the capacity is less than one, we can find a solution of (4.10), in this way, by choosing π_0 such that the probabilities sum up to one.

Using Little's formula (see e.g. Stidham [49]), we find the following expression for the average waiting time for an order, say w:

$$w = \frac{1}{N\alpha} \cdot \sum_{r=0}^{\infty} r\pi_r - 1 \tag{4.11}$$

By using the adjusted leadtime $\ell^* := \ell+w$, we have found the more advanced product-oriented heuristic.

Table 4.5. More advanced product-oriented heuristic in the purely stochastic approach, with $\nu=\lambda+\mu$ (for an explanation see Table 4.2, except that an adjusted leadtime has been used).

q	ℓ	N	a	b	ν	ρ	more advanced product-oriented heuristic appr	simul	β	simple prod.-or. heurist. SCH	optimal prod.-or. strategy simul
12	0	2	1	3	5	0.83	13.52	16.59/16.45	8	12.38/12.27	11.94/11.92
20	0	2	1	3	5	0.5	17.14	16.44/16.44	1	16.59/16.61	16.29/16.33
12	0	5	1	3	2	0.83	29.49	32.83/32.63	3	28.20/28.17	25.63/25.72
12	0	10	1	3	1	0.83	53.01	52.06/52.14	0	48.38/48.49	48.38/48.49
12	0	2	1	1	2	0.83	8.51	11.29/11.20	5	7.59/7.55	7.42/7.38
12	0	2	1	10	2	0.83	18.96	22.22/22.09	11	19.46/19.13	16.86/16.81
12	5	2	1	3	2	0.83	19.51	20.57/20.29	34	17.86/17.86	17.32/17.53
12	10	2	1	3	2	0.83	24.01	24.39/23.78	60	21.76/22.15	21.32/21.69
12	0	2	1	3	1	0.17	9.89	9.32/9.28	-2	9.32/9.28	9.32/9.28
12	0	2	1	3	5.5	0.92	16.02	31.81/31.89	18	16.77/16.74	13.49/13.84
12	0	2	1	3	5.8	0.97	21.71	79.75/79.49	47	36.24/40.14	22.35/22.32

Results for this more advanced heuristic are given in Table 4.5. If these results are compared to the results for the simple product-oriented heuristic and the optimal product-oriented strategy, it proves that this more advanced heuristic performs poor. It is even worse than the simple product-oriented heuristic. There may be two reasons for this bad performance:

- correlation between triggers for subsequent orders plays an important role
- other moments than only the first moment of the delay play an important role.

If we want to incorporate these correlations in the model or if we want to determine higher moments of the delay, the approach becomes more difficult. This supports the conjecture that one should use a capacity-oriented approach if one wants to account for the interference of the products on the capacity.

4.6.2 Stochastic Dynamic Programming approach.

In this subsection, we will compare the capacity-oriented and the product-oriented strategies that are described in Section 4.4. In Table 4.6, we give numerical results for these strategies.
These results lead to the following observations:

- if the part of demand that is known beforehand increases (that means the ratio $\mu/(\lambda+\mu)$ increases) both the capacity-oriented strategy and the product-oriented strategy improve, but there seems to be hardly any influence on their relative performance.

- if the utilization rate of the capacity ($\rho=N\cdot(\lambda+\mu)/q$) is high, the capacity-oriented strategy (in the Stochastic Dynamic Programming approach) performs best, whereas the product-oriented strategy performs best if this rate is low. This relation has also been found in Chapter 3, and seems intuitive.

- the performance of both strategies does hardly improve as the planning horizon increases. Therefore T=2 seems a reasonable choice.

- If the number of products increases, the performance of the capacity-oriented strategy gets worse rapidly. The reason for this is that the capacity-oriented strategy is based on the assumption that the inventories for all products can be kept equal. As the number of products increases, this assumption becomes more unrealistic. It is possible to account for the imbalance between the products in the cost of a given aggregate inventory and thereby improve the capacity-oriented strategy (see Section 4.6.1).

Table 4.6. Capacity-oriented and product-oriented strategies in the Stochastic Dynamic Programming approach.

q	ℓ	N	a	b	λ	μ	Thor	ρ	capacity-oriented	product-oriented
12	0	2	1	3	2.5	2.5	2	0.83	12.08/12.17	11.86/11.99
12	0	2	1	3	2.5	2.5	1	0.83	12.31/12.63	--------
12	0	2	1	3	2.5	2.5	3	0.83	11.98/12.00	11.93/12.04
12	0	2	1	3	2.5	2.5	5	0.83	11.93/11.90	11.95/11.95
12	0	2	1	3	0	5	2	0.83	11.28/11.41	10.75/10.71
12	0	2	1	3	5	0	2	0.83	12.51/12.86	12.20/12.26
20	0	2	1	3	2.5	2.5	2	0.5	18.36/18.47	15.86/15.96
12	0	5	1	3	1	1	2	0.83	30.38/30.21	27.60/27.36
12	0	10	1	3	0.5	0.5	2	0.83	60.70/60.54	48.74/48.99
12	0	2	1	1	2.5	2.5	2	0.83	7.11/7.12	7.47/7.47
12	0	2	1	10	2.5	2.5	2	0.83	22.00/21.41	18.66/18.66
12	5	2	1	3	2.5	2.5	2	0.83	17.00/17.23	16.88/16.85
12	10	2	1	3	2.5	2.5	2	0.83	21.53/21.34	21.35/21.38
12	0	2	1	3	0.5	0.5	2	0.17	11.42/11.46	9.05/9.05
12	0	2	1	3	2.75	2.75	2	0.92	13.85/14.09	17.07/16.75
12	0	2	1	3	2.9	2.9	2	0.97	23.20/20.93	41.98/41.98

Explanation of Table 4.6: Based on the known part of the demand and considering the schochasticity of demand, we can determine a production plan each period anew. In this Table we have given the simulation results that we find if we apply such an approach to find capacity-oriented and product-oriented strategies. For both types of strategies, we have used the Myopic rule to allocate a production run to the products (see Section 4.4).

4.6.3 Deterministic Dynamic Programming approach.

If we assume the forecast for future demand to be accurate, we find the Deterministic Dynamic Programming approach (see Section 4.5). Under the assumption that the forecast is accurate, we can determine an optimal production plan over the planning period, each period again. Doing so, we find the results of Table 4.7. Notice that we give results both if the final cost function is used and in case it is not used.

On the results of Table 4.7, we base the following conclusions:

-Only if the capacity restriction is tight or the stock-out costs are high, the use of a final cost function yields better results. Notice that for large N, the results become even worse. The reason for this may be that the cost function is based on the assumption that the simple capacity-oriented heuristic will be used after the planning horizon. However, especially as N increases, it proves that this capacity-oriented heuristic is not so good (using the more advanced capacity-oriented heuristic might yield a better cost function in this situation).

Table 4.7. Optimal strategy in the Deterministic Dynamic Programming approach (for an explanation see Table 4.6, except that the stochasticity of demand is no longer considered).

q	ℓ	N	a	b	λ	μ	Thor	ρ	without final cost function	with final cost function
12	0	2	1	3	2.5	2.5	2	0.83	11.58/11.79	11.37/11.68
12	0	2	1	3	2.5	2.5	3	0.83	10.90/11.31	10.90/11.11
12	0	2	1	3	2.5	2.5	5	0.83	11.03/10.95	11.00/10.96
12	0	2	1	3	0	5	2	0.83	10.58/10.66	10.37/10.51
12	0	2	1	3	5	0	2	0.83	12.33/12.32	12.13/12.35
20	0	2	1	3	2.5	2.5	2	0.5	15.85/15.84	15.94/15.84
12	0	5	1	3	1	1	2	0.83	25.32/25.23	26.34/25.99
12	0	2	1	1	2.5	2.5	2	0.83	7.27/7.26	7.07/7.20
12	0	2	1	10	2.5	2.5	2	0.83	20.31/20.36	19.37/19.11
12	5	2	1	3	2.5	2.5	2	0.83	17.79/17.60	17.74/17.43
12	10	2	1	3	2.5	2.5	2	0.83	23.56/23.21	23.85/23.13
12	0	2	1	3	0.5	0.5	2	0.17	9.29/9.30	9.66/9.59
12	0	2	1	3	2.75	2.75	2	0.92	15.37/15.73	13.61/14.51
12	0	2	1	3	2.9	2.9	2	0.97	32.50/33.55	23.46/21.21

- As the part of demand that is known beforehand increases (and thus the forecast is improved), the performance of the strategies gets better.

- in case we use Deterministic Dynamic Programming, the performance of the strategy seems more sensitive for the choice of the planning horizon than in the Stochastic Dynamic Programming approach. Therefore we have considered the same approach (without final cost function) again in Table 4.4, but now with planning horizon T=3.

Table 4.8. Optimal strategy in the Deterministic Dynamic Programming approach with planning horizon T=3 (for an explanation see Table 4.7).

q	ℓ	N	a	b	λ	μ	Thor	ρ	without final cost function
12	0	2	1	3	0	5	3	0.83	10.00/9.92
12	0	2	1	3	5	0	3	0.83	11.59/11.71
20	0	2	1	3	2.5	2.5	3	0.83	15.76/15.83
12	0	5	1	3	1	1	3	0.5	25.00/24.96
12	0	2	1	1	2.5	2.5	3	0.83	6.95/7.02
12	0	2	1	10	2.5	2.5	3	0.83	16.58/16.55
12	5	2	1	3	2.5	2.5	3	0.83	16.82/16.37
12	10	2	1	3	2.5	2.5	3	0.83	23.14/22.39
12	0	2	1	3	0.5	0.5	3	0.17	9.18/9.28
12	0	2	1	3	2.75	2.75	3	0.92	13.22/13.85
12	0	2	1	3	2.9	2.9	3	0.97	28.88/28.11

Comparing the results of Table 4.7 and 4.8 makes clear that increasing the planning horizon is especially useful if the stock-out costs are high or if the capacity restriction is tight. These are also the cases in which application of a final cost function is sensible.

If the number of products increases, the determination of an optimal production plan will become more difficult. Besides, the optimal production plan assumes that the forecast is accurate. It is not clear whether this "optimal" production plan will be better than plans that

Table 4.9. Capacity-oriented and product-oriented strategies in the Deterministic Dynamic Programming approach (for an explanation see Table 4.7).

q	ℓ	N	a	b	λ	μ	Thor	ρ	capacity-oriented	product-oriented
12	0	2	1	3	2.5	2.5	2	0.83	12.25/12.49	11.85/11.61
12	0	2	1	3	2.5	2.5	1	0.83	12.43/12.66	--------
12	0	2	1	3	2.5	2.5	3	0.83	12.24/12.44	11.63/11.61
12	0	2	1	3	2.5	2.5	5	0.83	12.46/12.28	11.55/11.52
12	0	2	1	3	0	5	2	0.83	11.28/11.41	10.75/10.71
12	0	2	1	3	5	0	2	0.83	11.96/13.27	12.39/12.36
20	0	2	1	3	2.5	2.5	2	0.5	18.33/18.52	15.81/15.93
12	0	5	1	3	1	1	2	0.83	30.60/30.57	25.32/25.33
12	0	10	1	3	0.5	0.5	2	0.83	61.33/61.79	54.51/54.44
12	0	2	1	1	2.5	2.5	2	0.83	7.13/7.20	7.26/7.31
12	0	2	1	10	2.5	2.5	2	0.83	25.71/25.92	20.35/20.60
12	5	2	1	3	2.5	2.5	2	0.83	19.21/18.80	17.74/17.86
12	10	2	1	3	2.5	2.5	2	0.83	25.53/24.68	23.19/23.36
12	0	2	1	3	0.5	0.5	2	0.17	11.18/11.11	9.32/9.21
12	0	2	1	3	2.75	2.75	2	0.92	14.76/15.11	15.97/15.52
12	0	2	1	3	2.9	2.9	2	0.97	23.35/21.15	36.73/39.16

are based on simpler approaches, if this assumption does not hold. Note that even if the assumption holds, the optimal production plan only optimizes the costs over a finite planning period. Therefore, we will also consider capacity-oriented and product-oriented strategies in this Deterministic Dynamic Programming approach.

In Table 4.9, we present numerical results for these strategies. In case of a capacity-oriented strategy the final cost function of Section 4.3 is used.
These results lead to the following conclusions:

- the capacity-oriented strategy is better than the product-oriented strategy if the capacity restriction is tight, whereas the product-oriented strategy is better otherwise.

- instead of searching an optimal strategy in the Deterministic Dynamic Programming approach, one can just as well use a capacity-oriented strategy in case of a tight capacity restriction and a product-oriented strategy otherwise (compare Tables 4.7 and 4.9).

4.6.4 Overview of the numerical results.

In Sections 4.2 to 4.5, we have mentioned several different approaches to the single-phase multi-product planning problem under periodic review, where demand is partly known in advance. So far, we have given numerical results for each approach separately. In this subsection, we want to compare the approaches. Therefore, we give the relative performance of the described strategies and heuristics in Table 4.10. These results are in percentages as compared to the best of them. For example if the best result is found for the purely stochastic approach, then there is a zero in the corresponding column. If the Deterministic Dynamic Programming approach gives a cost that is 1.2 times as high, then we have a 20 in the column that corresponds to the Deterministic Dynamic Programming approach.

Table 4.10. Comparison of the results for the different approaches (the numbers of the columns denote the approaches as is described in subsection 4.6.4).

q	ℓ	N	a	b	λ	μ	Thor	ρ	1	2	3	4	crit
12	0	2	1	3	2.5	2.5	2	0.83	5.5 P	2.1 P	0	0.4 P	P
12	0	2	1	3	2.5	2.5	3	0.83	11.0 P	7.9 P	0	4.6 P	P
12	0	2	1	3	2.5	2.5	5	0.83	12.1 P	8.5 C	0	5.0 P	P
12	0	2	1	3	0	5	2	0.83	16.1 P	1.0 P	0	1.0 P	P
12	0	2	1	3	5	0	2	0.83	0.8 P	0 P	0.8	1.2 P	P
20	0	2	1	3	2.5	2.5	2	0.5	4.8 P	0.4 P	0	0.2 P	P
12	0	5	1	3	1	1	2	0.83	11.5 P	8.7 P	0	0 P	P
12	0	10	1	3	0.5	0.5	2	0.83	0 P	0.9 P	---	12.5 P	P
12	0	2	1	1	2.5	2.5	2	0.83	5.8 C	0 C	2.1	0.7 C	P
12	0	2	1	10	2.5	2.5	2	0.83	3.4 P	0 P	9.0	9.7 P	P
12	5	2	1	3	2.5	2.5	2	0.83	5.8 C	0 P	4.9	5.5 P	P
12	10	2	1	3	2.5	2.5	2	0.83	2.8 P	0 P	9.5	8.9 P	P
12	0	2	1	3	0.5	0.5	2	0.17	3.6 P	0 P	2.7	2.4 P	P
12	0	2	1	3	2.75	2.75	2	0.92	2.3 C	0 C	11.3	6.9 C	P
12	0	2	1	3	2.9	2.9	2	0.97	3.9 C	0 C	49.7	0.8 C	C

The strategies and heuristics that we have compared in Table 4.10, are the following:

1. The best of the simple capacity-oriented heuristic and the product-oriented heuristic in the purely stochastic approach.

2. The best of the capacity-oriented strategy and the product-oriented strategy in the Stochastic Dynamic Programming approach.

3. The optimal strategy for the Deterministic Dynamic Programming approach (without final cost function).

4. The best of the capacity-oriented strategy and the product-oriented strategy in the Deterministic Dynamic Programming approach.

The C or P in Table 4.10 indicates whether the best strategy (or heuristic) is found for the capacity-oriented or for the product-oriented approach.
To find an optimal strategy in the Deterministic Dynamic Programming approach, we did not use an extra cost function on the final inventory positions, since we have seen that such a cost function is not useful in most situations (see subsection 4.6.2). However, since we combine capacity-oriented strategies always with a final cost function, this may result in a better performance of the capacity-oriented heuristic in the Deterministic Dynamic Programming approach for some situations in Table 4.10.

If we overview Table 4.10, we can derive the following conclusions:

-The Stochastic Dynamic Programming approach often yields a good strategy. However, the computational effort in case we account for the distribution of demand around the forecast is high, whereas one can find strategies that are almost as good (or sometimes better) if we either assume the forecast to be perfect (and thus find the Deterministic Dynamic Programming approach) or neglect the forecast (and thus find the purely stochastic approach).

-In most situations, the choice between a capacity-oriented and a product-oriented approach is independent of the way the information about future demand is used. In situations where there are differences, it can be checked, at hand of previous Tables, that there are only minor differences between the performance of the capacity-oriented and the product-oriented strategies. This allows us to apply the criterion of the purely stochastic approach (namely to use the heuristic for which the lower bound is the highest) to determine whether it is best to use a capacity-oriented strategy or a product-oriented strategy.

With these conclusions, we finish the discussion of the single-phase model where demand is partly known in advance. We have compared the capacity-oriented and the product-oriented strategies for different approaches to deal with the information that is available about future demand. It has proven that the criterion for choosing between the two types of strategies is insensitive for the way the information about future demand is treated.
Notice that we have also seen in Section 4.6.1 that the optimal capacity-oriented strategy and the optimal product-oriented strategy are almost just as good in many situations (an exception was the situation with a low utilization rate where the product-oriented strategy performed better). This suggests that there is a large overlap between situations where a capacity-oriented approach leads to good results and situations where a product-oriented approach leads to good results. The difficulty, however, is to find a good strategy within a given approach. We have derived a criterion to choose between the so-called simple heuristics (see Section 4.6.1). This criterion could be applied not only for the purely stochastic approach, but also for other approaches to deal with the information that is available about future demand. This enables us to choose between the capacity-oriented and the product-oriented approach, if we restrict ourselves to simple heuristics within each approach.
We also investigated whether it was easy to find a more advanced heuristic whithin each class that performed better than the simple heuristic. For the capacity-oriented approach, this has proven to be easy. A more advanced product-oriented heuristic, in which the delay in the leadtime for a product due to the interference with other products on the capacity was modelled as a stationary effect, proved to perform

poor. This suggests that it may be difficult to determine a good product-oriented approach in situations where the capacity restriction is tight (notice that the criterion indicates whether the capacity restriction is so tight that one might better use the capacity-oriented approach). The capacity-oriented approach, however may be used in a wider range of situations, if one uses the more advanced capacity-oriented heuristic instead of the simple capacity-oriented heuristic.

In the following Chapters, we will use these results, when considering a model in which the products are no longer identical in the long-run.

Chapter 5. Non-identical products.

5.1 Introduction.

In the previous Chapters, we have considered a single-phase multi-product planning problem with identical products. In this Chapter, we want to discuss how the results of capacity-oriented and product-oriented strategies can be applied in case of non-identical products. Let us first define what we mean by non-identical products.
The products are characterized by:

- the demand process
- the inventory holding cost and the stock-out cost
- the production rate (i.e. the amount that can be produced per unit of time).

We will consider products with different demand processes in this Chapter. We will assume, however, that the inventory holding cost, the stock-out cost and the production rate are the same for all products. Though this assumption may seem restrictive, it can be argued that it covers many situations. Graves [24] used the following reasoning to show that the assumption is realistic: "Consider the inventory holding and backorder costs; these costs are frequently taken to be proportional to the value of a product. In a single machine environment the value of a product consists of the cost of the input plus the value added during the processing. Given that all products require the same or similar inputs, it is reasonable to suppose that the machine processes the inputs at a relatively steady dollar rate, independent of the product. Furthermore, if the rate of value added by the machine were not nearly constant across products, it could be argued that the

machine capacity is not being used efficiently. Hence, the products should have similar value, and hence similar inventory holding and backorder costs per unit of production."
Consequently, after rescaling the production rates of the products to a common value, it seems reasonable to assume that the products have identical cost parameters. The reasoning of Graves is based on a situation with linear inventory holding and stock-out costs and we will consider a more general cost function (p(i)) on the inventories. However, this more general cost function may still be assumed to be the same for all products, after rescaling the products to a common production rate.

The approach in case of non-identical products is discussed using the example of a single-phase model with purely stochastic demand. This model is described in Section 5.2. In section 5.7, we will also make some remarks on how to treat the case where demand is partly known beforehand.

The approach is based on the results of previous Chapters which indicate that:

- the simple product-oriented heuristic is good in case of a weak capacity-restriction.
- the simple capacity-oriented heuristic is good in case of a strict capacity and only few products.

For the product-oriented approach, decomposition over the products is required. Therefore, the product-oriented heuristic may be expected to perform well for non-identical products too if the capacity restriction is weak.
If the capacity restriction plays an important role, we will divide the products into two groups namely a group of fast-movers and a group of slow-movers. As we will discuss, the slow-movers have priority on the capacity so that the capacity restriction for slow-movers is only weak. Since there will be only few fast-movers, each of these groups fulfills the requirements for applying a simple heuristic. We will discuss this distinction into different groups in Section 5.3. The form of the simple heuristics for this model with non-identical products, is discussed in Sections 5.4 and 5.5. In Section 5.6, we will illustrate

the approach at hand of an example. In the Sections 5.7 and 5.8, we will discuss whether it is possible to extend the approach to more general situations and we will make some remarks on the quality of the approach.

5.2 Description of the model.

There are N products that all require the same limited resource (compare the model of Chapter 3, see Figure 3.1).
The demand for product j follows a Poisson process with parameter λ_j, with $\lambda_j > \lambda_{j+1}$, for all j = 1, 2, ..., N-1. Let $\lambda := \sum_{j=1}^{N} \lambda_j$. There is no correlation between the demand for different products.

Let $I_j(t)$ denote the inventory of product j at time t. We consider the same cost rate as in Chapter 3, namely $p(i_j)$ is the cost rate for product j if the realisation of the inventory for that product equals i_j.

The purpose of control is to minimize the (expected) cost per unit of time. Unlike the situation of previous Chapters, we no longer assume that there is a stochastic process that generates production opportunities. In this Chapter, we will model the capacity as a service mechanism with service-time depending on the batch-size. The reason for using a service mechanism now, is that the batch-sizes for different products may be different.
A production run of size q will occupy the capacity for a negative-exponentially distributed time, with mean $q \cdot (1/\mu)$, after which the capacity is free again so that a new production run can be started.

In order to simplify the production situation, we assume that there are only two different batch-sizes, namely q_f and q_s (with $q_f > q_s$). The larger batch-size is used for products with a high average demand. We will return to this in the next Section.

Once a production run has been finished, the batch arrives at the stock-points after a production flowtime ℓ. From the time that the run is finished, the batch will be added to the inventory position, denoted by $Ipos_j$ (see Figure 5.1). Note that it would have been possible to

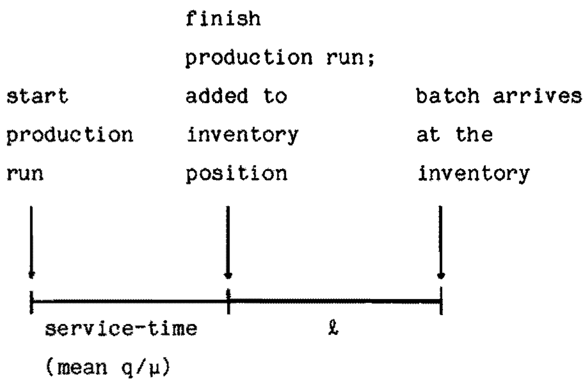


Figure 5.1. Throughput-time of a production order of size q.

define the inventory position such that the batch is already added as soon as the run has been started, but since we cannot produce before the batch is finished these definitions are equivalent.
As in Chapter 3, we will define the shifted cost rate $L(ip_j) := E[p(I_j(t+\ell))|Ipos_j(t)=ip_j]$ as the expected cost rate at time $t+\ell$, if the inventory position of product j equals ip_j at time t (in order to calculate $L(ip_j)$, consider (3.5)). The control of the inventory positions will now be directed at minimizing the shifted cost over time (in Chapter 3, we have discussed that this is equivalent to minimizing the cost rate $p(i_j)$ over time for the inventories).

5.3 Fast-mover/slow-mover approach.

If we assume that the products rarely interact on the capacity, the determination of a good heuristic is simple. In that case there is no need to buffer against uncertainties with respect to the availability of capacity for individual products. Consequently, it is almost optimal to assume that the capacity is always available for each individual product. This heuristic, we have called the simple product-oriented heuristic. To determine the heuristic we decompose with respect to the products, which gives us N one-dimensional problems that have to be solved. This solution is straightforward and yields a heuristic that may be expected to perform well (see the discussion of the slow-movers in Section 5.4).

Next consider the situation in which the capacity restriction plays an important role. The most straightforward way to buffer against uncertainties in the availability of the capacity is by introducing the aggregate inventory position as a measure for the amount of capacity that is stored in the system. The difficulty in this situation, where we have large differences between the products, is that a given inventory position is not equally effective as a capacity buffer in the short-run for each product:
In the long-run increasing the inventory by a quantity x, for any product, corresponds to decreasing the requirement for capacity with an amount (x/μ). In the short-run, however, there are obvious differences between the products: If there is hardly any demand for a given product, creating a large inventory for that product is completely inefficient in the short-run. Although in the long-run we may be sure that the inventory will be used, in the short-run it is just a waste. In general, it can be said that a given inventory position is more effective in the short-run if there is a high demand for the product (in the short-run).
Therefore we distinguish between so-called fast-movers and slow-movers. It is generally so that there are relatively few fast-movers, while each individual fast-mover has a big claim on the capacity. The slow-movers have only a minor claim on the capacity, but there are many slow-movers.

We introduce a value N_f such that the products 1 to N_f are the fast-movers and the rest are the slow-movers.
In Section 5.8, we will return to the choice of N_f.

A buffer against uncertainties, with respect to the availability of the capacity, will now be built up only in the inventories for fast-movers. For the slow-movers, such a buffer will not be built up. Therefore, it seems reasonable, if a slow-mover and a fast-mover compete for the allocation of capacity, to give the slow-mover priority, because a safety-stock has been built up for the fast-movers to protect against these capacity shortages.
Batch-sizes may interfere with the performance of this priority rule: In case the batch-size for slow-movers is very small (preferably lot-for-lot), this priority rule seems reasonable (compare the well-known Shortest Processing Time rule). However in case the batch-size for slow-movers is large, the following problem occurs. Each time we produce a slow-mover, part of the batch is used to reset the inventory

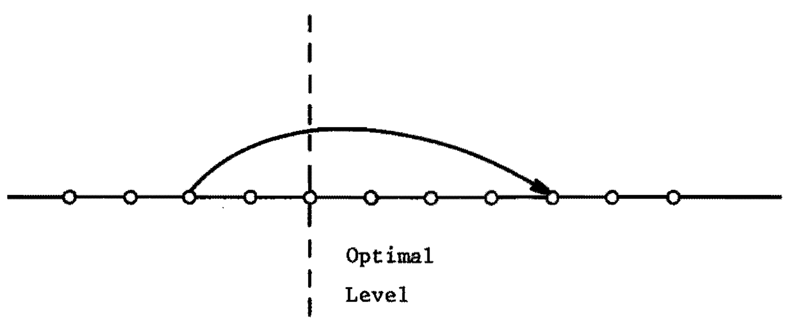


Figure 5.2. Batch production (q=6).

of the slow-mover to its "optimal" level, while the rest of the batch is used to reduce the number of set-ups for slow-movers in the long-run (see Figure 5.2). This last part of the batch will increase the costs in the short-run. During the production of this last part of the batch it will be far from optimal to let the fast-mover queue for the capacity. Therefore, the priority rule is most efficient if the batch-size for slow-movers is small. However, when choosing small batch-sizes for slow-movers, the capacity may be used inefficient due to change-over times between products. Obviously a trade-off has to be made:

choosing high batch-sizes for slow-movers may lead to less change-overs, but it also results in a more irregular availability of the capacity for the other products. This trade-off has to be made on the same level where the availability of the resources can be adjusted (we will return to this in Section 5.8). We assume that the fast-movers are produced in batches of size q_f, whereas the slow-movers are produced in batches of size q_s. In Section 5.7, we will discuss the situation where the batch-sizes need not be constant among fast-movers and slow-movers.

We have now discussed what happens if a slow-mover and a fast-mover apply simultaneously for allocation of capacity. What we have not discussed, yet, is what happens if a slow-mover applies at the time the capacity has already been allocated to a fast-mover. In this case of negative-exponentially distributed processing times it would be optimal to interrupt the production of the fast-mover at such an event. Yet, it seems more realistic not to interrupt the production run for a product (non-preemptive situation), since usually the reason for using batches in the first place is that there are change-over costs and times. On the same level where batch-sizes are fixed, a decision has to be taken about preemption or non-preemption. We assume that the slow-movers have a non-preemptive priority on the capacity.

Once we have determined the priority rule, the determination of a good heuristic, using the results of the previous Chapters, is straightforward. There are many slow-movers, but, since they have priority on the capacity, the capacity restriction is weak when we consider the slow-movers only. As we have seen, in this situation, the simple product-oriented heuristic, that decomposes with respect to the products, is almost optimal. We will return to this heuristic for slow-movers in Section 5.4.
For the fast-movers, we are left with a much tighter capacity restriction. However, we have chosen the number of fast-movers to be small. For a situation with identical products, we have seen that the simple capacity-oriented heuristic, in which the decision whether or not to produce depends on the aggregate inventory position as a measure for the amount of stored capacity, performs well under these conditions. In Section 5.5, we will return to this heuristic.

5.4 Slow-movers.

As mentioned previously, the slow-movers have priority on the capacity. Therefore, the capacity restriction for slow-movers is only weak. For situations with identical products, we have seen in the previous Chapters that the simple product-oriented heuristic performs well under such circumstances. The performance of this heuristic can be estimated by solving one-dimensional optimization problems. In case of non-identical products, a product-oriented strategy is a strategy in which a production run for a slow-mover j is triggered if and only if the realisation of the inventory position for that slow-mover is less than or equal to a certain predetermined level β_j.

If more than one slow-mover triggers a production run, then we need a rule to choose between them. In the previous Chapters, we have seen that a reasonable rule is to allocate the production run to the products in such a way that the first costs, that are influenced by this allocation will be minimized. In Chapter 4, we have compared this allocation rule (the so-called Myopic rule) to a more advanced rule (the Value Function rule) in a situation where the products are non-identical in the short-run. The results indicated that there is little improvement when using a more advanced allocation rule. Since the Myopic rule is the simplest, we will apply this rule here as well. In the Myopic rule, the production run is allocated to the slow-mover j for which expression (5.1) is minimal (ip_j denotes the realisation of the inventory position for product j).

$$\sum_{d=0}^{\infty}\left[p(ip_j+q_s-d)-p(ip_j-d)\right]\cdot P_j(d) \qquad (5.1)$$

where $P_j(d)$ is the probability that the demand for product j over the production time plus the production flowtime equals d. Note that this demand consists of two parts. Firstly, there is the demand during the production time. Since the production time is negative-exponentially distributed, the demand during the production time is geometrically distributed (with parameter $\frac{\lambda_j}{\lambda_j + \mu/q_s}$). The second part consists of demand during the production flowtime ℓ. This second part has a Poisson distribution with parameter $\ell \cdot \lambda_j$. Combining these two, we get the following expression for the demand of product j during the throughput-time:

$$P_j(d) = \sum_{k=0}^{d} \frac{\mu/q_s}{\lambda_j + \mu/q_s} \cdot \left\{\frac{\lambda_j}{\lambda_j + \mu/q_s}\right\}^k \cdot \frac{(\ell\lambda_j)^{d-k}}{(d-k)!} \; e^{-\ell\lambda_j} \qquad (5.2)$$

To find the simple product-oriented heuristic, the interaction with other products is neglected. Define $\hat{H}^{(d)}_{\beta,j}(ip)$ as the steady-state probability that, if we apply a product-oriented heuristic with critical level β for product j and under the above assumption (no interaction), the inventory position of product j equals ip. The cost of product j (if we apply this heuristic under the above assumption) is given by (5.3).

$$\sum_{ip} L(ip) \cdot \hat{H}^{(d)}_{\beta,j}(ip) \qquad (5.3)$$

where L(ip) is defined as the shifted cost rate (see Section 5.2).

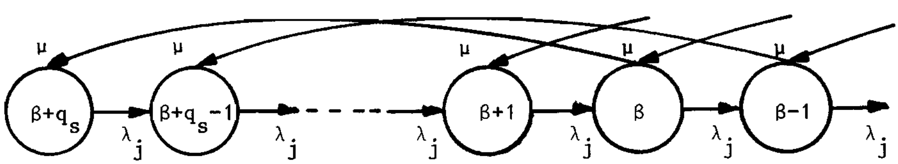


Figure 5.3. Possible transitions of the inventory position for a slow-mover j.

To determine $\hat{H}^{(d)}_{\beta,j}(ip)$, consider the possible transitions between inventory positions for a given slow-mover j, as depicted in Figure 5.3, under the assumption that there is no interference.

These possible transitions appear to be the same as the possible transitions in Section 3.8 (Figure 3.4). Therefore the same approach can be used here in order to find the simple product-oriented heuristic.

Note that, as before, it is possible to determine a more-advanced product-oriented heuristic by estimating the delay in the availability of the capacity due to other products (using some queueing analysis). Such an approach has been worked out in Section 4.6.

5.5 Fast-movers.

The capacity-aspect of the system will be considered in the control of fast-movers. Consequently, the decision whether or not to produce has to depend on the total inventory position, aggregated to stored capacity. Since the "production rate" (the amount of a product that can be produced per unit of time) is the same for all fast-movers, the best measure for the amount of stored capacity is found by taking the sum of the individual inventory positions. Therefore, we define the aggregate inventory position at time t as $Ipos(t) := \sum_{j=1}^{N_f} Ipos_j(t)$. Capacity-oriented strategies are strategies in which a production run is triggered if and only if the realisation of the aggregate inventory position is less than or equal to a certain predetermined level β.

If a production run is started, then (as for the slow-movers) this run will be allocated to the products using the Myopic rule (5.1).

In order to be able to make a reasonable choice of β, we will consider the following three points:

1. the cost rate as a function of the aggregate inventory position
2. the demand process for fast-movers
3. the disturbance of the capacity-availability for fast-movers due to slow-movers.

5.5.1 The cost rate as a function of the aggregate inventory position.

The actual costs that correspond to a given aggregate inventory position depend on the distribution of the aggregate over the individual products. Therefore the conditional steady-state probabilities (given the aggregate) have to be calculated in order to find the exact cost rate of the aggregate inventory position. In Chapter 3, we discussed the problem of finding such a conditional distribution, and we have seen that, as long as N_f is not too large, a good approximation of this aggregate cost rate is given by $\tilde{g}(ip) := N_f \cdot L(ip/N_f)$. This means simply assuming that all inventory positions can be kept equal.
Note that in Chapter 4, we considered a more advanced approximation of the conditional costs too. The approximation that is used for the aggregate cost rate is unimportant with respect to the size of the resulting problem (as long as the cost rate depends only on the aggregate inventory position). Since the approximation $\tilde{g}(ip)$ is the most straightforward and has proven to give good results if N_f is not too large, we prefer this approximation. The capacity-oriented strategy that we find with this approximation is referred to as the simple capacity-oriented heuristic.

5.5.2 The demand process for fast-movers.

The demand process for fast-movers is the superposition of N_f independent Poisson processes. Therefore, again, it is a Poisson process with parameter $\tau_f := \sum_{j=1}^{N_f} \lambda_j$.

5.5.3 The disturbance of the capacity-availability for fast-movers due to slow-movers.

In order to investigate the disturbance of the capacity-availability for fast-movers due to slow-movers, let us first consider a single slow-mover.

As soon as q_s times a demand for this slow-mover occurs, it will trigger a production run (consider the strategies chosen for the slow-movers). Since the demand for slow-mover j follows a Poisson process with parameter λ_j, the interarrival time between successive "triggermoments" for this slow-mover is Erlang-distributed with q_s phases and parameter λ_j.

The process of triggermoments for all slow-movers is the superposition of $N-N_f$ individual trigger processes. Since the number of slow-movers $(N-N_f)$ is large, this process will be approximated well by a Poisson process (for an extensive study on the superposition of non-identical processes, the reader is referred to Khintchine [33]). The parameter in this process is:

$$\tau_s := \sum_{j=N_f+1}^{N} \frac{\lambda_j}{q_s} \tag{5.4}$$

A possible approach would be just to estimate the expected delay of production runs for a fast-mover, due to the disturbance of the capacity, and to use an adjusted leadtime. However, there will be much variation around the average size of the delay. This is due to the queueing of production orders, not only during the time another slow-mover is processed but also during the time a fast-mover is processed. Because of this effect, it may not be suitable to introduce an adjusted leadtime (compare Section 4.6). Therefore, we have used another approach for choosing β in this situation.

The state of the "fast-mover system" at a given point in time can be described by the following tuple:

$$(ip, n_s, r) \tag{5.5}$$

where ip is the realisation of the aggregate inventory position, n_s is the number of unfinished production orders for slow-movers (including the one that presently occupies the capacity) and r denotes the present status of the capacity as follows:

$$r = \begin{cases} 0 & \text{means the capacity is free} \\ 1 & \text{means the capacity is occupied by a fast-mover} \\ 2 & \text{means the capacity is occupied by a slow-mover} \end{cases}$$

It will be clear that if we know the steady-state probability of each state (ip, n_s, r), related to any given strategy, β, and hence we know the marginal distribution of ip, it is possible to determine the approximate, average cost for any choice of β (by using the approximate cost function on the aggregate inventory position as discussed in Section 5.5.1). Therefore we can determine the simple capacity-oriented heuristic by minimizing these average costs over β.

We will now describe a theoretical analysis, based on the above state description, that leads to the simple capacity-oriented heuristic. This analysis is rather technical. For readers who are not interested in the specific derivation, it suffices to say that the analysis is the matrix-analogon of the analysis in Section 3.8.

5.5.4 Theoretical Analysis.

To simplify the analysis, we will bound the number of unfinished production orders for slow-movers that require allocation of the capacity at any moment, say $n_s \leq m$. If there are m unfinished production orders for slow-movers in the system, the arrival of new customers will not result in a transition to a new state. Notice, that since the slow-movers have priority on the capacity, we may expect that this approximation is reasonable, for sufficiently large values of m.

If the system is in equilibrium, not all combinations of ip, n_s and r are possible:

- as long as ip > β (and ip $\leq \beta+q_f$), the capacity cannot be occupied by fast-movers. Consequently, we have only the states (ip,0,0) and (ip,n_s,2) for all $1 \leq n_s \leq m$. Notice that if a production order for a slow-mover arrives, while the capacity is available, this order will occupy the capacity immediately. Therefore, states (ip,n_s,0) are reached with probability zero for $n_s > 0$.

- if ip $\leq \beta$, it is impossible to find the capacity available, because immediately a production run for a fast-mover is started. Therefore r=0 is not possible.

Hence, the states of the system for fast-movers, are as shown in Figure 5.4.

$\beta+q_f,0,0$	$\cdots$	$\beta+1,0,0$	$\beta,0,1$	$\beta-1,0,1$	$\cdots$
$\beta+q_f,1,2$	$\cdots$	$\beta+1,1,2$	$\beta,1,2$	$\beta-1,1,2$	$\cdots$
$\vdots$		$\vdots$	$\vdots$	$\vdots$	
$\beta+q_f,m-1,2$	$\cdots$	$\beta+1,m-1,2$	$\beta,m-1,2$	$\beta-1,m-1,2$	$\cdots$
$\beta+q_f,m,2$	$\cdots$	$\beta+1,m,2$	$\beta,m,2$	$\beta-1,m,2$	$\cdots$
			$\beta,1,1$	$\beta-1,1,1$	$\cdots$
			$\vdots$	$\vdots$	
			$\beta,m-1,1$	$\beta-1,m-1,1$	$\cdots$
			$\beta,m,1$	$\beta-1,m,1$	$\cdots$

Figure 5.4. The possible states for the fast-mover system in equilibrium.

If we want to calculate the steady-state probabilities, the first problem that we meet is that the times between transitions depend on the state of the system. Lippman [36] has shown that this problem can be overcome easily in this situation with exponential distributions, by introducing transitions that do not result in a change of state. Therefore, in any state we can assume that transitions are due to:

1. arrival of demand for a fast-mover; rate τ_f (= $\sum_{j=1}^{N_f} \lambda_j$)
2. finishing of a fast-mover job; rate $\mu_f := \mu/q_f$
3. finishing of a slow-mover job; rate $\mu_s := \mu/q_s$
4. arrival of a slow-mover job; rate τ_s (= $\sum_{j=N_f+1}^{N} \lambda_j$)

where, for example, the finishing of a fast-mover job, if no such job has been started in the given state, corresponds to a transition from the state into itself.

Note that, since the times between transitions are independent of the state of the system, it suffices to calculate the steady-state probabilities of the state of the system, at times just after a transition, which means that we must consider the embedded Markov chain. Then a given transition corresponds with probability

- $p_{af} := \tau_f/(\tau_f+\mu_f+\mu_s+\tau_s)$ to the arrival of a fast-mover
- $p_{ff} := \mu_f/(\tau_f+\mu_f+\mu_s+\tau_s)$ to the finishing of a fast-mover job
- $p_{as} := \tau_s/(\tau_f+\mu_f+\mu_s+\tau_s)$ to the arrival of a slow-mover job
- $p_{fs} := \mu_s/(\tau_f+\mu_f+\mu_s+\tau_s)$ to the finishing of a slow-mover job.

Define the vector of states $\underline{Iv}_k$ (for $k \geq 0$) as follows (compare Figure 5.4, where we have shifted the bottom states to the left in order to ensure that the length of all vectors is the same, namely 2m+1):

$$
\underline{Iv}_k := \begin{bmatrix} \beta+q_f-k,0,0 \\ \beta+q_f-k,1,2 \\ \cdot \\ \cdot \\ \cdot \\ \beta+q_f-k,m-1,2 \\ \beta+q_f-k,m,2 \\ \beta-k,1,1 \\ \cdot \\ \cdot \\ \cdot \\ \beta-k,m-1,1 \\ \beta-k,m,1 \end{bmatrix} \text{for } 0 \leq k \leq q_f-1 \ ; \ \underline{Iv}_k := \begin{bmatrix} \beta+q_f-k,0,1 \\ \beta+q_f-k,1,2 \\ \cdot \\ \cdot \\ \cdot \\ \beta+q_f-k,m-1,2 \\ \beta+q_f-k,m,2 \\ \beta-k,1,1 \\ \cdot \\ \cdot \\ \cdot \\ \beta-k,m-1,1 \\ \beta-k,m,1 \end{bmatrix} \text{for } k \geq q_f \qquad (5.6)
$$

To make the description of the matrix of transition probabilities easier, we introduce the $(2m+1)\times(2m+1)$ matrices :

- B corresponding with a transition from $\underline{Iv}_k$ to $\underline{Iv}_k$ if $0 \leq k \leq q_f-1$
- A_1 corresponding with a transition from $\underline{Iv}_k$ to $\underline{Iv}_k$ if $k \geq q_f$
- A_0 corresponding with a transition from $\underline{Iv}_k$ to $\underline{Iv}_{k+1}$ if $k \geq 0$
- A_{q_f+1} corresponding with a transition from $\underline{Iv}_k$ to $\underline{Iv}_{k-q_f}$ if $k \geq q_f$

It can be checked that this conforms to the following definitions:

$$
\begin{array}{cc}
 & \begin{array}{ccccccccccccc} 0 & 1 & \cdot & \cdot & \cdot & \cdot & m & 1 & 2 & \cdot & \cdot & \cdot & m \end{array} \\
B:= & \begin{bmatrix}
p_{fs}+p_{ff} & p_{as} & & & & & & & & & & & \\
p_{fs} & p_{ff} & p_{as} & & & & & & & & & & \\
 & p_{fs} & p_{ff} & p_{as} & & & & & & & & & \\
 & & \cdot & \cdot & \cdot & & & & & & & & \\
 & & & \cdot & \cdot & \cdot & & & & & & & \\
 & & & & \cdot & p_{ff} & p_{as} & & & & & & \\
 & & & & & p_{fs} & p_{ff}+p_{as} & & & & & & \\
 & p_{ff} & & & & & & p_{fs} & p_{as} & & & & \\
 & & p_{ff} & & & & & & p_{fs} & p_{as} & & & \\
 & & & \cdot & & & & & & \cdot & \cdot & & \\
 & & & & \cdot & & & & & & \cdot & \cdot & \\
 & & & & & p_{ff} & & & & & & p_{fs} & p_{as} \\
 & & & & & & p_{ff} & & & & & & p_{fs}+p_{as}
\end{bmatrix}
\begin{array}{c} 0 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ m-1 \\ m \\ 1 \\ 2 \\ \cdot \\ \cdot \\ m-1 \\ m \end{array}
\end{array}
\tag{5.7}
$$

$$
A_0 := \begin{bmatrix}
p_{af} & & & & & \\
 & p_{af} & & & & \\
 & & \cdot & & & \\
 & & & \cdot & & \\
 & & & & \cdot & \\
 & & & & & p_{af}
\end{bmatrix}
\tag{5.8}
$$

$$
A_1 := \begin{array}{c}
\begin{array}{ccccccccccccc} 0 & 1 & \cdot & \cdot & \cdot & \cdot & m & 1 & 2 & \cdot & \cdot & \cdot & m \end{array} \\
\left[\begin{array}{ccccccccccccc}
p_{fs} & 0 & & & & & & & & & & & \\
p_{fs} & p_{ff} & p_{as} & & & & & & & & & & \\
 & p_{fs} & p_{ff} & p_{as} & & & & & & & & & \\
 & & \cdot & \cdot & \cdot & & & & & & & & \\
 & & & \cdot & \cdot & \cdot & & & & & & & \\
 & & & & \cdot & p_{ff} & p_{as} & & & & & & \\
 & & & & & p_{fs} & p_{ff}+p_{as} & & & & & & \\
 & p_{ff} & & & & & & p_{fs} & p_{as} & & & & \\
 & & p_{ff} & & & & & & p_{fs} & p_{as} & & & \\
 & & & \cdot & & & & & & \cdot & \cdot & & \\
 & & & & \cdot & & & & & & \cdot & \cdot & \\
 & & & & & p_{ff} & & & & & & p_{fs} & p_{as} \\
 & & & & & & p_{ff} & & & & & & p_{fs}+p_{as}
\end{array}\right]
\end{array}
\begin{array}{c} \\ 0 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ m-1 \\ m \\ 1 \\ 2 \\ \cdot \\ \cdot \\ m-1 \\ m \end{array}
\qquad (5.9)
$$

$$
A_{q_f+1} := \begin{array}{c}
\begin{array}{ccccccccccccc} 0 & 1 & \cdot & \cdot & \cdot & m & 1 & 2 & \cdot & \cdot & \cdot & \cdot & m \end{array} \\
\left[\begin{array}{ccccccccccccc}
p_{ff} & & & & & & p_{as} & & & & & & \\
 & & & & & & & & & & & & \\
 & & & & & & & & & & & & \\
 & & & & & & & & & & & & \\
 & & & & & & & & & & & &
\end{array}\right]
\end{array}
\qquad (5.10)
$$

Define the vector of states $\underline{Iv}$ via $\underline{Iv}^T := [\ \underline{Iv}_1^T, \underline{Iv}_2^T, \ldots\ ]$, where $\underline{Iv}^T$ is the transpose of $\underline{Iv}$. Then, we can write the matrix of transition probabilities for $\underline{Iv}$, denoted by P, as follows:

$$
\begin{array}{l}
\begin{array}{ccccccccc}
\quad 0 & 1 & 2 & \cdot & \cdot & q_f & \cdot & \cdot & \cdot
\end{array} \\
P := \left[\begin{array}{ccccccccc}
B & A_0 & & & & & & & \\
 & B & A_0 & & & & & & \\
 & & \cdot & \cdot & & & & & \\
 & & & \cdot & \cdot & & & & \\
 & & & & B & A_0 & & & \\
A_{q_f+1} & & & & & A_1 & A_0 & & \\
 & A_{q_f+1} & & & & & A_1 & A_0 & \\
 & & \cdot & & & & & \cdot & \cdot \\
 & & & \cdot & & & & \cdot & \cdot
\end{array}\right]
\end{array}
\qquad (5.11)
$$

Let $\underline{\pi}_k$ denote the vector of steady-state probabilities that corresponds to $\underline{Iv}_k$ and define the vector $\underline{\pi}$ via:

$$\underline{\pi}^T := [\ \underline{\pi}_0^T, \underline{\pi}_1^T, \underline{\pi}_2^T, \cdots \quad] \qquad (5.12)$$

The vector of steady-state probabilities satisfies the following set of equations:

$$
\begin{cases}
\underline{\pi} \geq 0 \\
\underline{\pi}^T = \underline{\pi}^T \cdot P \\
\underline{\pi}^T \cdot e_\infty = 1 \qquad \text{where } e_\infty := (1, 1, \cdots \quad)^T
\end{cases}
\qquad (5.13)
$$

This set becomes more obvious if we rewrite $\underline{\pi}^T = \underline{\pi}^T \cdot P$ as:

$$\begin{cases} \underline{\pi}_0^T = & \underline{\pi}_0^T \cdot B & + & \underline{\pi}_{q_f}^T \cdot A_{q_f+1} & \\ \underline{\pi}_k^T = \underline{\pi}_{k-1}^T \cdot A_0 & + \underline{\pi}_k^T \cdot B & + & \underline{\pi}_{k+q_f}^T \cdot A_{q_f+1} & \text{for } 1 \le k \le q_f - 1 \\ \underline{\pi}_k^T = \underline{\pi}_{k-1}^T \cdot A_0 & + \underline{\pi}_k^T \cdot A_1 & + & \underline{\pi}_{k+q_f}^T \cdot A_{q_f+1} & \text{for } \quad k \ge q_f \end{cases}$$

$$(5.14)$$

Neuts [42] proved that the solution of (5.14) for $k \ge q_f$ is of the "geometric type", under certain recurrence conditions, which are satisfied as long as the utilization rate of the capacity is less than one (see e.g. Neuts [42]), and under some irreducibility conditions, corresponding to $\lambda > 0$ and $\tau_s > 0$. This means that for $k \ge q_f-1$ we get (5.15).

$$\underline{\pi}_{k+1}^T = \underline{\pi}_k^T \cdot R \tag{5.15}$$

where the $(2m+1)\times(2m+1)$ non-negative, irreducible matrix R is the minimal solution in the set of non-negative matrices of spectral radius less than one that satisfy equation (5.16) (for a definition of the terms related to the use of matrices, the reader is referred to Seneta [48]).

$$R = R^{q_f+1} \cdot A_{q_f+1} + R \cdot A_1 + A_0 \tag{5.16}$$

Neuts [42] showed also that there is an easy way to compute this matrix R (up to any given degree of accuracy): The sequence of the $(2m+1)\times(2m+1)$ matrices $\{X(u)\}_{u=0}^{\infty}$, as defined in (5.17), is monotonely increasing and converges to R (in [42], Neuts also discussed some other methods for finding R).

$$\begin{cases} X(0) := 0 \\ X(u+1) := X^{q_f+1}(u)\cdot A_{q_f+1} + X(u)\cdot A_1 + A_0 \end{cases} \tag{5.17}$$

Applying this result of Neuts, it follows from (5.15) that

$\underline{\pi}_k = \underline{\pi}_{q_f-1} R^{k-q_f+1}$ satisfies the relationships for $k \geq q_f-1$ in (5.14).

If we apply the equations for $k=0, 1, \ldots, q_f-2$, we can also express subsequently $\underline{\pi}_0, \underline{\pi}_1, \ldots, \underline{\pi}_{q_f-2}$ in terms of $\underline{\pi}_{q_f-1}$. This gives us:

$$\begin{cases} \underline{\pi}_k^T = \underline{\pi}_{q_f-1}^T \cdot R^{k-q_f+1} & \text{if } k \geq q_f-1 \\ \underline{\pi}_k^T = \underline{\pi}_{q_f-1}^T \left(\sum_{s=0}^{k} R^{s+1} A_{q_f+1} (I-B)^{-1} \cdot \{A_0 (I-B)^{-1}\}^{k-s}\right) & \text{if } 0 \leq k \leq q_f-2 \end{cases} \tag{5.18}$$

The problem that remains, is to determine $\underline{\pi}_{q_f-1}$ so that (5.13) is satisfied. Let us look at each type of equations in (5.13) separately:

a) $\underline{\pi} \geq 0$.

Note that the spectral radius of B is less than one, and that B is irreducible. Therefore $(I-B)^{-1}$ exists and is strictly positive (see Seneta [48]). Using (5.18), now we can see that $\underline{\pi} \geq 0$ if $\underline{\pi}_{q_f-1} \geq 0$. We will return to the possibility of finding such a $\underline{\pi}_{q_f-1}$, in the next paragraph.

b) $\underline{\pi}^T \cdot P = \underline{\pi}^T$.

These equations are depicted more clearly in (5.14). By choosing the $\underline{\pi}_k$'s as in (5.18), all the equations in (5.14) will hold for all values of $k \neq q_f-1$. The only equation that needs to be checked, is:

$$\underline{\pi}^T_{q_f-1} = \underline{\pi}^T_{q_f-2}\cdot A_0 \quad + \quad \underline{\pi}^T_{q_f-1}\cdot B \quad + \quad \underline{\pi}^T_{2q_f-1}\cdot A_{q_f+1}$$

Which is equivalent to:

$$\underline{\pi}^T_{q_f-1} = \underline{\pi}^T_{q_f-1}\left(\sum_{s=0}^{q_f-1} R^{s+1} A_{q_f+1}(I-B)^{-1}\cdot\{A_0(I-B)^{-1}\}^{q_f-s-1}\right) \tag{5.19}$$

Consequently, we need to show that the $(2m+1)\times(2m+1)$ matrix $H := \sum_{s=0}^{q_f-1} R^{s+1} A_{q_f+1}(I-B)^{-1}\cdot\{A_0(I-B)^{-1}\}^{q_f-s-1}$ has a left eigenvector with eigenvalue one, and that this eigenvector is nonnegative. If this is true, we have found $\underline{\pi}_{q_f-1}$ up to a multiplicative constant. Then, we can find this multiplicative constant in the next paragraph.

Notice that H is non-negative and irreducible (since the matrices R, A_{q_f+1} and A_0 are non-negative and $(I-B)^{-1}$ is strictly positive).
Therefore, in order to show that the matrix H has a non-negative left eigenvector with eigenvalue one, say $\underline{u}$, with $\underline{u} \neq \underline{0}$, it suffices to show that H has a non-negative right eigenvector with eigenvalue one (see Theorems 1.5 and 1.6 in Seneta [48]).
Define the $(2m+1)$ vector $\underline{e} := (1,1,\dots,1)^T$, then we will show that $A_0\underline{e}$ (which is strictly positive) is a right eigenvector of H with eigenvalue one. Consequently, we have to show that

$$H\cdot(A_0\underline{e}) = A_0\underline{e} \tag{5.20}$$

Therefore, we first remark that by definition (see (5.7) and (5.9)) it holds that

$$(B+A_0)\cdot\underline{e} = \underline{e} \tag{5.21}$$

which can be written as

$$(I-B)^{-1}A_0\cdot\underline{e} = \underline{e} \tag{5.22}$$

Consequently, (5.20) can be written as

$$\sum_{s=0}^{q_f-1} R^{s+1} A_{q_f+1} \cdot \underline{e} = A_0 \underline{e}$$

which is equivalent to

$$(I-R)^{-1} \cdot (R \cdot A_{q_f+1} - R^{q_f+1} A_{q_f+1}) \cdot \underline{e} = A_0 \underline{e} \tag{5.23}$$

By (5.16), this is equivalent to

$$(I-R)^{-1} \cdot (R \cdot A_{q_f+1} + R \cdot A_1 + A_0 - R) \cdot \underline{e} = A_0 \underline{e} \tag{5.24}$$

Notice that by definition we also have

$$(A_0 + A_1 + A_{q_f+1}) \cdot \underline{e} = \underline{e} \tag{5.25}$$

If we substitute this in (5.24), we find

$$(I-R)^{-1} \cdot (A_0 - R \cdot A_0) \cdot \underline{e} = A_0 \underline{e} \tag{5.26}$$

Notice that (5.26) is equivalent to $A_0 \underline{e} = A_0 \underline{e}$, which concludes our proof.

c) $\underline{\pi}^T \underline{e}_\infty = 1$ with $\underline{e}_\infty := (1,1,\ldots\ )^T$

Applying (5.18), this can be rewritten as

$$\underline{\pi}^T_{q_f-1} \left(\sum_{k=0}^{q_f-2} \sum_{s=0}^{k} R^{s+1} A_{q_f+1} (I-B)^{-1} \cdot \{A_0 (I-B)^{-1}\}^{k-s} + \sum_{k=q_f-1}^{\infty} R^{k-q_f+1}\right) = 1 \tag{5.27}$$

Under b), we have already determined $\underline{\pi}_{q_f-1}$ up to a multiplicative constant, say α (i.e. $\underline{\pi}_{q_f-1} = \alpha\cdot\underline{u}$). We now define α using (5.28)

$$\alpha^{-1} := \underline{u}^T\left(\sum_{k=0}^{q_f-2}\sum_{s=0}^{k}R^{s+1}A_{q_f+1}(I-B)^{-1}\cdot\{A_0(I-B)^{-1}\}^{k-s}+\sum_{k=q_f-1}^{\infty}R^{k-q_f+1}\right) \tag{5.28}$$

It will be clear that α can be defined by (5.28), because the right-hand side of (5.28) is positive and finite: the expression is positive because of the fact that $\underline{u} \geq 0$, $\underline{u} \neq \underline{0}$ and $(I-R)^{-1} = \sum_{k=q_f-1}^{\infty}R^{k-q_f+1}$ is strictly positive, whereas all other matrices are non-negative. The expression is finite because of the existence of $(I-R)^{-1} = \sum_{k=q_f-1}^{\infty}R^{k-q_f+1}$.

So far, we have proved that it is possible to choose $\underline{\pi}_{q_f-1}$ such that the $\underline{\pi}_k$'s defined by (5.18) are indeed the steady-state probabilities. If $\underline{\pi}_{q_f-1}$ is given, these probabilities are easy to calculate because R can be found easily. The vector $\underline{\pi}_{q_f-1}$ equals $\alpha\cdot\underline{u}$, where α follows directly from (5.28) and $\underline{u}$ is the left eigenvector of a given matrix H. This eigenvector can be found by straightforward numerical methods within a limited computational time (see Stoer and Bulirsch [50]).

This enables us to calculate the marginal steady-state probabilities of a given realisation of the aggregate inventory position for fast-movers, when applying a capacity-oriented strategy β. Denote the probability of a realisation of the aggregate inventory position that equals ip as $P_\beta(ip)$. Note that the nature of the process allows the probabilities, when applying the capacity-oriented strategy β-1, to be found via (5.29).

$$P_{\beta-1}(ip) = P_\beta(ip+1) \qquad \text{for all values of ip and } \beta \tag{5.29}$$

This enables us to find an optimum choice of β.

5.6 An example.

This fast-mover/slow-mover approach is illustrated with a specific example having the following characteristics:

- the demand for product j follows a Poisson process with parameter

  $$\lambda_j = \frac{0.9/j}{\sum_{r=1}^{100} 1/r}.$$

- there are 100 products ($N = 100$). Consequently, the total demand follows a Poisson process with parameter $\lambda = 0.9$.

- The production rate μ equals one, which corresponds to a utilization rate of 0.9.

- the products numbered 1, 2 and 3 are considered to be fast-movers and the others are slow-movers. Note that $\sum_{j=1}^{3} \lambda_j = 0.3534$.

- the fast-movers are produced in batches of size 2 ($q_f = 2$) and the slow-movers are produced lot-for-lot ($q_s = 1$).

- if a production run is finished, it arrives immediately at the stock-points ($\ell = 0$).

- the cost rate is chosen as $p(i) := i^+ + 3 \cdot i^-$.

The simple product-oriented heuristic, that is used for the slow-movers, indicates that a production run for a slow-mover should be triggered as soon as the inventory is less than or equal to -1 ($\beta_j = -1$ for all $j \geq 4$).

Then, the expected cost of product j, assuming no interference between products, becomes $3\cdot\frac{\rho_j}{1-\rho_j}$, where $\rho_j = \lambda_j/\mu$ is the ratio of time that the capacity is occupied by slow-mover j.
The approximate expected total cost for slow-movers then becomes 1.77.

The simple capacity-oriented heuristic indicates that a production run for a fast-mover should be triggered when the realisation of the aggregate inventory position is less than or equal to $\beta = 10$. The approximate expected cost for fast-movers with this heuristic (determined for m=16) proves to be: 14.14.

We have also simulated the process under the same strategy. Thus, we obtained simulated costs for both the group of fast-movers and the group of slow-movers when applying the proposed heuristic. Two simulation runs gave the following results:

	first run	second run
simulated cost slow-movers:	5.93	5.94
simulated cost fast-movers:	14.85	15.59

Looking at the results, it may be concluded that the actual cost for fast-movers can be approximated reasonably well. A better approximation for the actual cost of the fast-movers can be found by using a more advanced approximation of the aggregate cost rate than in Section 5.5.1. This has been investigated for a periodic review model in Chapter 4, where good results were obtained.
The performance of the slow-movers has been approximated badly. To estimate this performance, two assumptions have been made:

- different slow-movers never interfere with each other on the capacity
- a slow-mover never interferes on the capacity with a fast-mover.

Note that since the priority for slow-movers is non-preemptive, also the second assumption needs not hold.

The simplicity of the process for slow-movers (due to the fact that they are produced lot-for-lot) allows to find a simple approximation of the cost that is based only on the second assumption. That means that this new approximation explicitly takes the interference between slow-movers into account. The new approximation is based on the fact that the number of unfinished orders for slow-movers can be described as an M/M/1 queue: interarrival-times between unfinished orders are negative-exponentially distributed with parameter $\sum_{j=4}^{100} \lambda_j$ (≈ 0.5466) and the service-time is negative-exponentially distributed with parameter $\mu=1$. Consequently, the expected total cost for slow-movers can be approximated by $3 \cdot \frac{\sum_{j=4}^{100} \lambda_j/\mu}{1 - \sum_{j=4}^{100} \lambda_j/\mu} \approx \underline{3.62}$.

Although this approximation is better, it still is far from the actual cost. The reason for this is that the interference between slow-movers and fast-movers plays an important role. A rough approximation of this delay due to the slow-movers can be found as follows:
The utilization rate for fast-movers is about 0.3534. An arriving order for a slow-mover has therefore a probability 0.3534 to find a fast-mover job in process. The expected delay due to fast-movers may be approximated by $0.3534 \cdot q_f/\mu = 0.7068$. During this delay, the inventory for the slow-mover is -1, so that the increase in the total costs due to the delay can be approximated by $3 \cdot 0.7068 \approx 2.12$. This would lead to an estimated cost for slow-movers of $\underline{5.74}$, which is near to the actual cost for slow-movers.

The above approximations are only meant to show how the assumptions that underlie the simple product-oriented approach play a role. If one wants to determine a more general heuristic, one may use the results of Williams [xx], who showed how one may take into account the interference between the products in order to find a more advanced product-oriented heuristic.

5.7 Applicability of the fast-mover/slow-mover approach.

In this Section, we want to discuss whether it is possible to use the proposed fast-mover/slow-mover approach in situations where:

- demand is partly known beforehand
- the classification of products as fast-mover or slow-mover is dynamic
- different batch-sizes are used among the fast-movers and among the slow-movers.

Firstly, we want to look at the problems that arise if demand is partly known in advance. These problems stem from the fact that whether a given inventory position is effective as a capacity buffer in the short-run depends on the forecast for future demand (compare Chapter 4). However as long as there are some products that always are fast-mover (other situations are discussed below), we can still apply the same fast-mover/slow-mover approach: We distinguish some fast-movers in which we will put our capacity buffer. The other products will be given priority on the capacity. Product-oriented strategies as discussed in Chapter 4, can be used to control the slow-movers. To find a capacity-oriented strategy, one may use a critical level rule as discussed in Section 5.5, if demand for fast-movers is (almost) purely stochastic, or one may use a rolling schedule approach with a finite planning horizon as in Chapter 4 in case a good forecast for demand is available. In a rolling schedule approach, it is possible to consider the exact capacity usage of the slow-movers in the short-run (a plan for slow-movers has already been found) and next prepare a production plan for the fast-movers based on an "optimization" over the finite horizon.

Now consider a situation where the classification of products as fast-mover or as slow-mover is dynamic. This may for example be due to seasonality of demand or to obsolescence of a fast-mover.
The problem, if a new classification is made, is that a given inventory that is built up becomes ineffective to cope with capacity problems in the short-run. Consequently, such a new classification will only have minor effects if the stocks are low. Sometimes, however, a fast-mover

may become obsolete, for which inventory has been stored in order to be able to cope with capacity problems. In case there is no slow-mover for which demand increases at the same time, this will lead to a decrease in the demand for capacity, so that the fact that the inventory becomes ineffective as a buffer has little effect.
A common situation, however, is where the demand for a slow-mover increases at the same time that the demand for a fast-mover decreases (for example if the original fast-mover is replaced by a new product

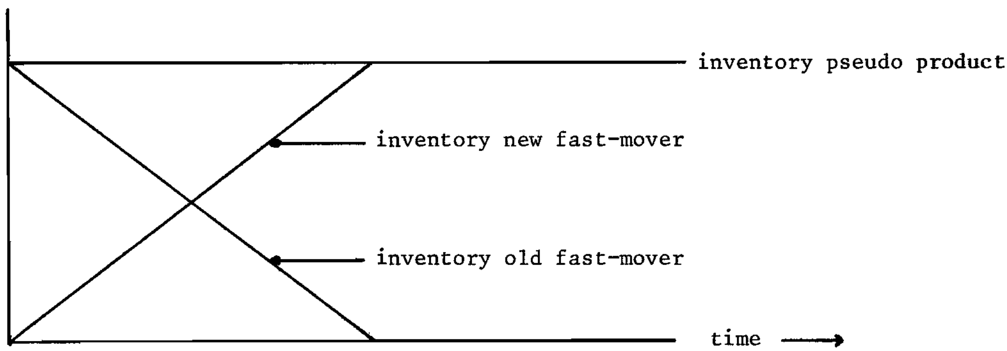


Figure 5.5. The distribution of the inventory of the pseudo product between the old fast-mover and the new fast-mover.

with slightly different characteristics). If the replacement of the fast-mover by the slow-mover is gradually, or if the replacement can be predicted well in advance, it is sensible to shift the inventory that is available for the original fast-mover to the new fast-mover. With the fast-mover/slow-mover approach, this can be done by introducing a "pseudo product". The pseudo product is an aggregate of the original fast-mover and the new fast-mover. The demand on the capacity due to this pseudo product will (hardly) be affected by the obsolescence of the original fast-mover. Only the distribution of the aggregate inventory of the pseudo product among the original and the new fast-mover has to change as the obsolescense of the fast-mover approaches (see Figure 5.5).

Finally, we want to discuss a situation where different batch-sizes, that are still assumed fixed at the level of Material Coordination, are used among the fast-movers and among the slow-movers.

It will be clear that when decomposing the products, it is not important whether all products have the same batch-size, or not. Therefore, choosing different batch-sizes for the slow-movers will have no effect on the approach. Consequently, we only have to consider the effect of using different batch-sizes for the fast-movers. Since the production rate for each fast-mover is the same, a given inventory position will represent the same amount of stored capacity, independently of the fast-mover. Therefore, the capacity-oriented approach should depend on the aggregate inventory position defined as the sum of all individual inventory positions. A difficulty appears now: if we decide to start a production run for the group of fast-movers, the transition of the aggregate inventory position will depend on the choice of a product within the group of fast-movers. If the batch-sizes of the fast-movers do not differ much, it seems reasonable to use the expected batch-size as a common batch-size, when deciding whether or not to start a production run for fast-movers. If the differences between the batch-sizes increase, this approach can be improved by assuming that, after the decision to start a production run for fast-movers, there is a given probability of a certain realisation of the batch-size. It can be proved that the same approach introduced by Neuts [42], and used in Section 5.5, can be applied if the batch-size is stochastic. Notice that also in this case the decision to start a production run is decomposed from the decision of the run-size (we will return to this in Chapter 6).
Notice that in case Material Coordination has the flexibility to choose run-sizes, depending on the actual status of the system, the approach presented here, cannot be applied straightforwardly any more. Again a capacity-oriented approach and a product-oriented approach can be distinguished. The actual run-size will depend on the availability of the capacity in the capacity-oriented approach and on the status of individual products in the product-oriented approach.

5.8 Optimality of the fast-mover/slow-mover approach.

When considering the fast-mover/slow-mover approach, the question comes to mind how good this approach is compared to other possible approaches. The answer to this question is not easy to give. Within the class of situations where the batch-sizes are as given and the priority rule is such as proposed in the previous Sections, it may be expected that the fast-mover/slow-mover approach performs well. This, at least, is indicated by the results of the previous Chapters, that show that a simple product-oriented heuristic performs well if the capacity restriction is weak and that a simple capacity-oriented heuristic performs well in case this restriction is tight (as long as the number of products is not large). Although these results are obtained for situations with identical products, we may expect them to hold for non-identical products as well, especially since they have proven to hold for a situation where the products are not identical in the short-run (see Chapter 4).

The question, however, is whether it is sensible to choose the batch-sizes and the priority rule as in the previous Sections. This decision has to be taken on the level where the availability and the requirement for the resource are balanced (the level of Tactical Planning in the framework of Chapter 2). Notice that also the classification of products as fast-mover or slow-mover (the choice of N_f) plays a role in balancing these two.

It is interesting to compare the research of Williams [55] at this point. Williams considered a model in which he distinguished products that are made-to-stock and products that are made-to-order. The products that are made-to-order have a non-preemptive priority on the capacity (Williams also considered other priority rules, but this rule performed best). His research is concerned with the following points:

1. Which products should be made-to-stock and which products should be made-to-order.
2. What demand should be accepted for products that are made-to-order.
3. How should one choose the batch-sizes for products that are made-to-stock.
4. What is the effect on the safety stocks for make-to-stock products due to make-to-order products.

The products that are made-to-stock can be compared with our fast-movers and the products that are made-to-order with our slow-movers (with batch-size equal to one).

For several choices with respect to the first three points, Williams has evaluated the performance of, what we call, Material Coordination. His approach was product-oriented. He tried to estimate the delay in the availability of the capacity for individual products that is due to the interference between products on the capacity. The analysis, presented in this Chapter, makes it possible to use a capacity-oriented approach when evaluating a given set of choices with respect to the first three points.

Chapter 6. A simple example.

6.1 Introduction.

In this Chapter, we want to consider an example of how to use the results of this study in practical situations. Therefore, we will look at an existing situation that contains aspects other than the ones in the models of the previous Chapters. For example, in this Chapter, a role will be played by change-over times, interference between production plan and maintenance plan, advertising, choice of batch-size and availability of raw materials. In the example, there is one obvious bottle-neck in the production process.

We will not discuss how the results of this text can be applied in situations with more than one bottle-neck in the production process, since this requires further research. Yet, the results of this study indicate that it may be advantageous to introduce the capacity-aspect of inventories in such situations as well.

In Section 6.2, we will describe the situation that we want to consider in detail. The proposed structure for Production Control in this situation will then be sketched in Section 6.3. In Section 6.4, we will discuss a specific part, namely Material Coordination and go into its interface with other levels of control. We will finish this study with some conclusions in Section 6.5.

6.2 The situation.

The production situation, considered in this Chapter, is the manufacturing of plastic utensils. Diagrammatically, we have described the production process in Figure 6.1.

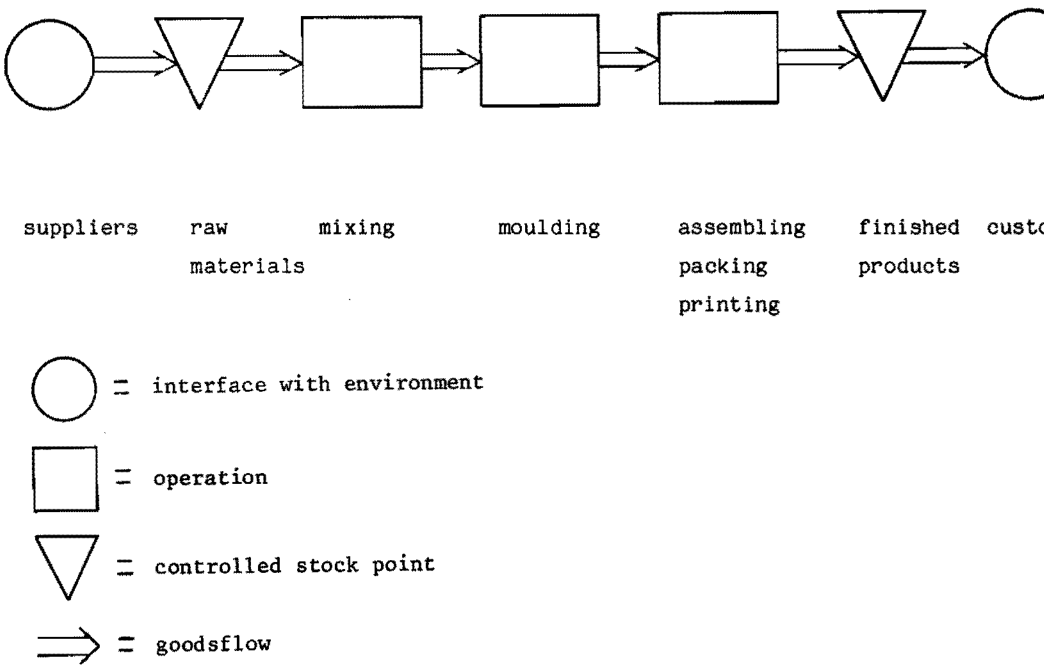


Figure 6.1. Manufacturing phases in the plastic products factory.

To make the description of the situation clear, we will discuss the characteristics of the demand, the production process and the supply of raw materials in distinct subsections.

6.2.1 Demand.

The set of products, with about 800 items, is divided into 200 families. The items may differ in colouring, printing or assembling.

For example, a specific form of basket (completely determined by the mould that is used) will be referred to as a family, whereas red baskets, green baskets or baskets with different handles will be viewed as different items within that family.

The customers are mainly large warehouses. There are about 1500 of such warehouses that each generate a relatively small part of the demand. The demand is, therefore, relatively smooth but difficult to forecast in detail. Consequently, the demand per item can be treated as almost purely stochastic.

For some families, there is a seasonality in the demand. These families have about 20 % of the total annual sales. However, this percentage changes throughout the year: 30 % in the first half year and 4 % in the second half (see Figure 6.2).

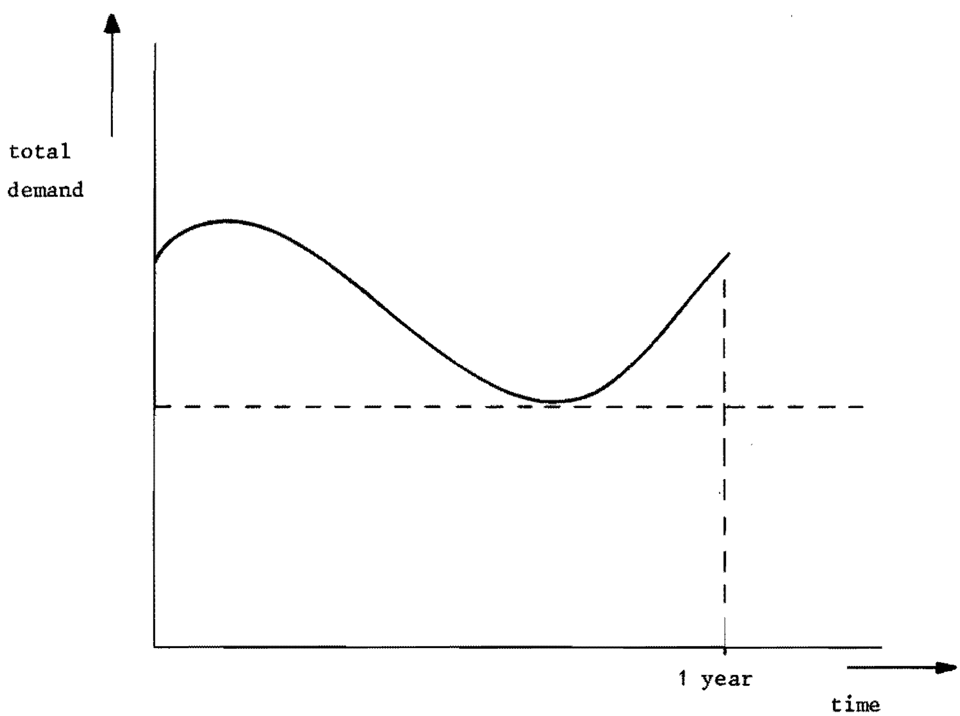


Figure 6.2. The seasonality of demand.

The demand may be influenced by starting advertising campaigns. The preparation time for such a campaign is about a month. On the other hand, demand will also be influenced by advertising campaigns of competitors. Note that these advertising campaigns influence the quality of the forecast for demand.

Finally, we want to mention the risk of obsolescence. The mean life cycle of a family is about 10 years. For items the life cycle is much shorter, namely about 1 year. Usually, however, when an item that is part of a given family ends its life-cycle it is replaced by another item that is part of the same family: think for example of changing, slightly, the colour of a basket.

6.2.2 Production process.

In the production process, the moulding stage has a central place. The mould that is chosen, determines which family will be produced. Before this moulding stage, the plastic pellets are mixed and the pigment is added. This mixing is a straightforward process. The only difficulty with mixing is that, when one wants to change the colour, it only gradually shifts from one colour to another. Thus a lot of scrap is produced. This leads to change-over costs and change-over times if one wants to manufacture another item within the same family. We will return to this when discussing the moulding stage of the process. Mixing is relatively easy and can be done by the same operator that controls the moulding-machine.

We might add that the process that is sketched in Figure 6.1 is an extreme. Not all finished items have to go through the whole traject. On the other extreme, there are items that only require moulding, without mixing, and packing.
All items, however, go through the moulding stage. The role the assembling/packing/printing stage takes in the process is similar to the role of the mixing stage. There is a large excess of capacity at this stage. The reason for holding this excess of capacity is that the products that are manufactured in the moulding stage are difficult to keep clean and they are voluminous. Therefore, a large amount of work-in-process on the production floor is undesirable. Instead, the

products that leave the moulding stage are immediately processed to finished products.

The bottle-neck in the production process is the moulding stage: There are 9 capacity groups of moulding-machines that can be used. Each group consists of several moulding-machines. The moulding-machines within the same capacity group have about the same production characteristics. The moulding-machines within different capacity groups differ in the weight which they can handle. "Heavy" moulding-machines are used most efficiently when handling large items, whereas "light" moulding-machines can only handle small items. Change-overs between families for the light moulding-machines take about half an hour, whereas for the heavy moulding-machines it may take 6 hours.
The change-overs for items within the same family take about 10 minutes for the light moulding-machines and 40 minutes for the heavy moulding-machines.

The manufacturing time for a single item ranges from 8 seconds on the light moulding-machines to 30 seconds on the heavy moulding-machines. Production is in batches.

Sometimes, a batch is rejected. This occurs randomly to 2 % of the batches. Other sources for uncertainty are worker-absentheism (15 %) or breakdowns of the moulding-machines (4 %).

Maintenance on a moulding-machine requires 5 % of the production time. If a mould is broken, it may take upto 15 weeks before it is repaired.

6.2.3 Supply of raw materials.

The main raw materials are plastic pellets. There are not many different types of plastic pellets (about 20). These pellets represent about 70 % of the total value of the raw materials. Other raw materials are pigments, labels, and boxes (600 different items).

The average supply time for the plastic pellets is about a half to one week. For other raw materials, like the metal handles for baskets, the supply time may be 12 weeks.

The total value of the raw materials in finished items is about 30 %. Therefore, a simple way of control for raw materials seems appropriate.

6.3 Production Control.

In Chapter 2, we have mentioned that there are two different reasons for the complexity of Production Control. On the one hand the production process itself may be complex and on the other hand the activities, that should be part of Production Control, have different ranges and levels of detail. Consequently, we have used the concept "Production Unit" and we have defined various "Levels of Control" in Section 2.2.1, respectively 2.2.2.

In this situation, the introduction of Production Units is straightforward. There is only one bottle-neck in the production process, namely the moulding stage. Both mixing and finishing are relatively simple. Therefore, we define the aggregate process (see Chapter 2) as in Figure 6.3. In order to stress the central place that moulding holds in the production process, mixing and finishing (assembling/packing/printing) are modelled as production flowtimes within Moulding.

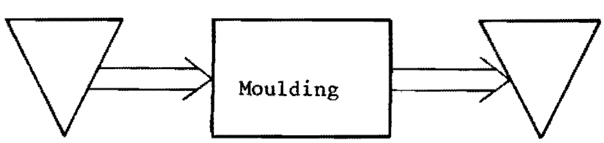


Figure 6.3. Central place of Moulding.

The levels of control, that we propose for this situation, are:

-Strategic Planning

-Master Planning

-Material Coordination

-Scheduling.

In this Section, we will discuss each of these levels roughly. In Section 6.4, we will then discuss the level of Material Coordination in more detail.

Strategic Planning.

On the highest level of Production Control, we have pictured Strategic Planning. At this level, decisions are to be taken that will influence the long-term behaviour of the plastic products factory. This level includes decisions like entering a new market, buying a new stock-house, or changing the work-force levels drastically.
An important decision with respect to lower levels of control, is how much influence customer orders have on the control of the production process (see e.g. Burbidge [15] and van Hees [29]). Burbidge [15] mentions the following three examples of systems that deal with customer orders in different ways: make-to-order systems, stock controlled systems and programme controlled systems.
In this situation with a smooth demand that cannot be forecasted very accurately, and a relatively long life cycle for the families (ten years), we propose to use a stock controlled system.

Master Planning.

Master Planning falls in the framework of Chapter 2 under the heading of "Tactical Planning".
Master Planning has to balance the requirement and the availability of resources in the production process. Therefore, Master Planning should have a planning horizon that covers at least a whole season (plus the time it takes to implement the decisions that are made by Master Planning), which is one year. Over this horizon, the behaviour of individual items presents much uncertainty (e.g. the average life cycle of an item is about one year). Therefore, Master Planning will, in our

view, be concerned exclusively with families (notice that the average life cycle of a family is about ten years).
Over the planning horizon, the demand forecast for families will contain information about the seasonality pattern. The resolution level of this information will be about a month. In order to be able to react on new information about the seasonal patterns, such a Master Plan should be made, say, every three months.

In Figure 6.4, we have sketched the inputs and the outputs (which we will discuss below) of Master Planning.

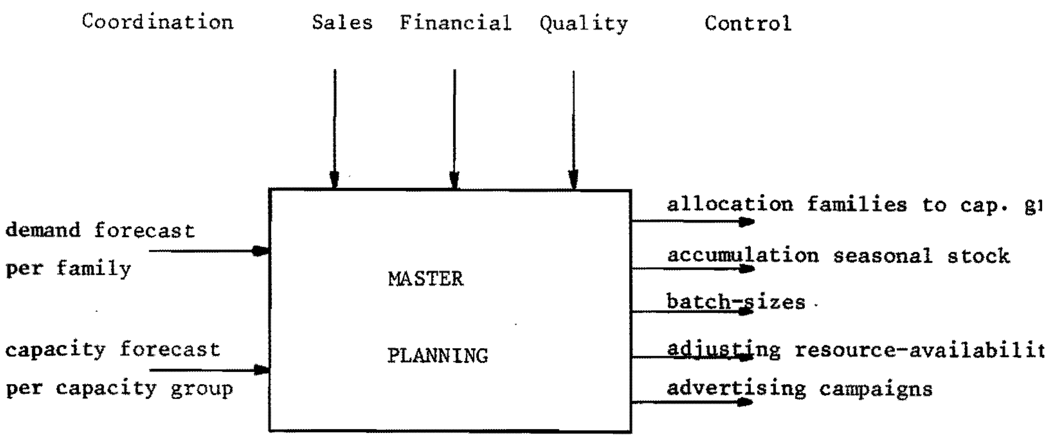


Figure 6.4. Inputs and Outputs of Master Planning.

As we have mentioned, the capacity groups differ with respect to the weight and size of products they can handle efficiently. Although some families can be manufactured on different capacity groups efficiently, we propose to make an allocation of the families to the capacity groups on this level. This allocation should be kept to at lower levels, in principle, for a whole season. Only if problems become urgent, the flexibility to shift a family to another capacity group may be used.
At the level of Master Planning, however, the families are allocated to the capacity groups, so that for each capacity-group, the availability and the requirement of capacity can be balanced. Obvious possibilities to control this relation are to start advertising campaigns (in order to stimulate the demand) or to adjust the availability of resources (e.g. hire extra personel). However, both

possibilities only have a limited applicability. Advertising campaigns have only effect in the short term, and it is much too expensive to adjust the availability of resources often. Since the capacity, in this example, is not so flexible that it can be varied in order to have sufficient capacity in the high season without having too much capacity in the low season, one has to "store" capacity in the low season. Consequently, one accumulates stock in the low season to cope with the capacity problems that would occur (otherwise) in the high season. The stock that is accumulated for this purpose will be referred to as "seasonal stock". How much seasonal stock must be accumulated depends on the available capacity and on how efficiently this capacity will be used. How efficiently a given capacity can be used, depends on the batch-sizes. Consequently, the decision on the accumulation of seasonal stock should be combined with restrictions on the batch-size per family and on the decisions about adjustment of the capacity. It would be possible to leave the decision on the exact run-sizes for families to lower levels of control (e.g. Material Coordination), and only set restrictions on the run-sizes for the families. However, de Bodt and van Wassenhove [14] have shown that little is gained by coupling the run-size decision with the decision on when to start a production run, in case demand cannot be forecasted accurately. In this situation demand, even for families, cannot be forecasted well in detail. This leads us to fixing the run-sizes (for families) already on the level of Master Planning, which makes the control on lower levels more easy. To have an idea of the type of situation, one should think of run-sizes of about one week.
Notice that Master Planning only considers families. How the run-size for a family is disaggregated into run-sizes for individual items is left to lower levels of control.

After the batch-sizes per family have been fixed, Master Planning decides how much seasonal stock must be accumulated in the low season for each capacity-group. This decision is integrated with the adjustment of the resources and the starting of advertising campaigns. An example of the type of model that can be used to balance the investment in seasonal stock and the investment in extra capacity is put forward by Hax and Meal [28], who have developed a "Seasonal Planning Subsystem" for a specific situation with seasonal demand. An analogous system can be developed for the plastic products factory.

Note that these two can be balanced more easily if one has insight in the effect of changing the capacity on the performance of control at lower levels. Since a capacity-oriented approach on the level of Material Coordination is based on a comparison of the availability of capacity and the requirement for capacity, it will be more easy to evaluate the effect of varying the availability of capacity (we will return to this in Section 6.5).

Note that the timing of orders is left to Material Coordination. Master Planning only generates norms with respect to seasonal stock in each period. We assume below, that Master Planning determines what cost is incurred if Material Coordination deviates from the norms with respect to seasonal stock.

Finally, we want to mention that on this level of Production Control, there is also coordination with the other control processes in the organisation. For example, together with Quality Control a maintenance plan for the equipment has to be constructed and Financial Control has to agree on the level of capital that is tight up in work-in-process and in stocks. As far as the coordination with Sales Control is concerned, we have so-called "structural coordination" on this level (see Chapter 2). That means that the coordination with Sales Control is in aggregate terms (like customer service-rate and demand patterns).

Material Coordination.

Master Planning has imposed certain norms on the seasonal stock per capacity group. These norms can be used as a target for Material Coordination. Material Coordination has to reach these targets. If Material Coordination deviates from these targets, a certain cost is incurred.
Besides reaching the required seasonal stocks, a task of Material Coordination is to realise a certain customer service-rate (see Figure 6.5).

Since we have allocated the families to the capacity-groups on the level of Master Planning, Material Coordination can consider the different capacity-groups separately.

Material Coordination for each capacity-group will be split into two levels.
On the first level, only families are considered. Using short term forecasts for demand of families and information about the seasonality

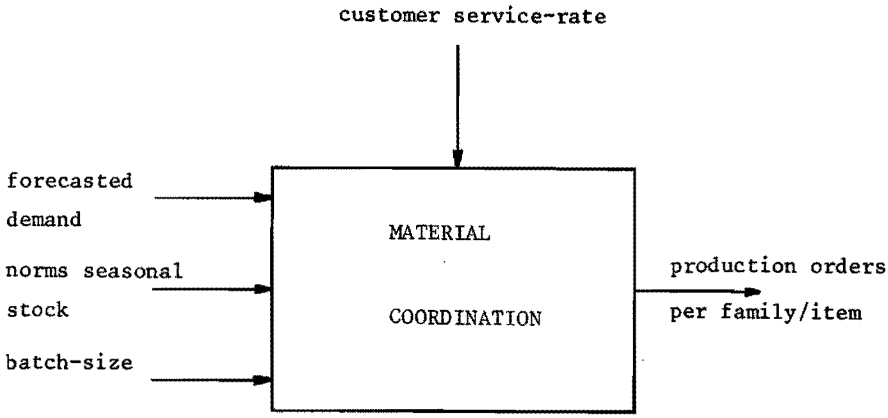


Figure 6.5. Inputs and outputs per capacity-group for Material Coordination.

pattern, it is decided to start production runs for families (the run-size has been fixed by Master Planning).
On the second level, the production run for a family is disaggregated into production quantities for individual items.
Since items within the same family require approximately the same raw materials, set-up costs, processing times and storage space, Material Coordination should emphasize the first level. Then, disaggregation may be straightforward, e.g. by equalisation of run-out times, as proposed by Hax and Meal [28]. Remark that they defined the run-out time of a product as the expected time until the inventory of the product drops below the safety stock of that product. A common way to define the run-out times is as the time until a stock-out occurs for the product.
We will return to the first level of Material Coordination, in Section 6.3.

Material Coordination also has to ensure that the raw materials are available. However, as we have seen, the raw materials are relatively cheap and are used in bulk quantities (at least for the largest part, namely the plastic pellets). Consequently, a simple method for the control of the stocks for raw materials seems adequate. For example, a Reorder Point System (see Section 2.4.1) can be used in this context.

Scheduling.

Scheduling has to assign production orders, that are generated by Material Coordination, to the moulding-machines within a given capacity group. Also operators will be allocated to the moulding-machines. We will not discuss Scheduling any further in this context. On the level of Material Coordination, it is required to have a model of the performance of the Production Units. For a discussion about the performance-measurement of a Production Unit, the reader is referred to Bertrand and Wortmann [9].

6.4 Material Coordination.

In this Section, we will consider Material Coordination in some more detail.
In the previous Section, when describing a design for Production Control for the plastic products factory, we have seen that Material Coordination has to release family-orders for each of the nine capacity groups. Since, at a higher level of control, the families have been assigned to the capacity groups, there are nine separate planning problems: one for each capacity group.
For each planning problem, the batch-sizes of the families and the availability of resources have been fixed. Material Coordination has to decide on the production levels in order to guarantuee the customer service-rate and the accumulation of seasonal stocks, that have been set by Master Planning.

In order to be able to use the results of previous Chapters, we will discuss how demand is experienced on this level, what cost function is used and what the service mechanism looks like for Material Coordination.

The demand in each planning problem (one for each capacity group) is autonomous. This demand is difficult to forecast in detail. Only over a short horizon a detailed forecast is made, whereas in the long term only information is available about the seasonal pattern per family. Consequently, it seems reasonable to consider only a short planning period, in which the forecast for demand is updated frequently (at this level some operational coordination with Sales Control is possible). This may mean that one uses a rolling schedule with a short horizon, or that one determines critical levels for the inventories. For these critical levels, one may think of levels per product (to ensure a certain customer service-rate) or a level on the aggregate (to ensure that the seasonal stock is reached). This level on the aggregate should then be determined dynamically.

As discussed already, Material Coordination must provide a certain customer service-rate and it must reach a certain seasonal stock.
In order to ensure a certain customer service-rate, we may introduce inventory holding costs and stock-out costs for the inventory of individual families (compare Tinarelli [52] and Schwarz [47]).
The seasonal stock is meant to be able to cope with capacity problems in the high season. Therefore, it seems reasonable to measure the seasonal stock, for each capacity group, in aggregate terms. In Chapter 5, we have seen that such stock, that is meant to buffer against capacity problems, should be stored in the fast-movers for each capacity-group. The slow-movers, then, get priority on the capacity. In this situation with a risk of obsolescence, it is sensible to build up high stocks only for families with a low risk of obsolescence. To measure whether enough seasonal stock has been accumulated, a cost rate on the aggregate inventory position of the fast-movers may be used. This cost rate follows from the analysis of seasonal stocks at the level of Master Planning, where stock norms are set and a cost for deviating from these norms is determined. Notice that these stock norms have to be chosen dynamic in this case with a seasonal demand.

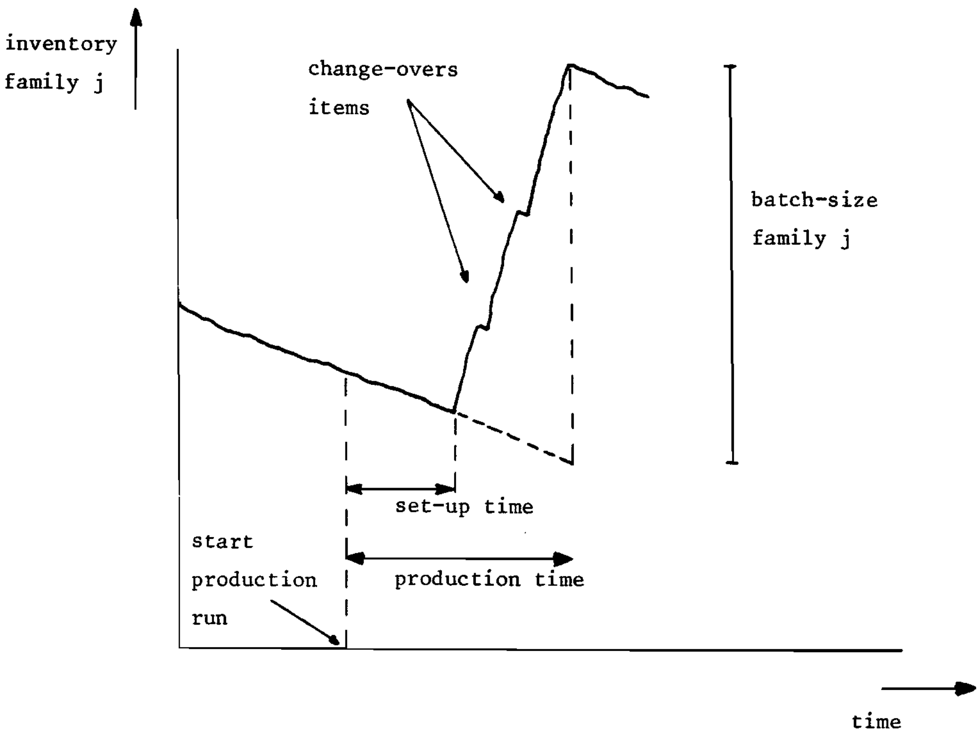


Figure 6.6. Inventory pattern for family j.

Now that we have discussed the demand and the cost function for Material Coordination, we will discuss the service mechanism. If a production run for family j is started, we have to know how this influences the inventory for that family. Therefore, we have sketched an example of the inventory pattern for family j, over a period of time in which a production run for the family is started (see Figure 6.6). It should be realised that the production time for a batch for family j consists of several parts:

1. set-up time for the family (=change-over between families)
2. change-over times between items within the family
3. actual manufacturing time for the batch.

If we denote the production time for family j by T_j, then T_j will be stochastic. This stochasticity is due to the following points:

- the number of change-overs between items depends on how the batch for the family will be disaggregated on a lower level of Material Coordination.
- the maintenance and breakdowns of moulding-machines.
- the uncertainty in the manufacturing time of a single item.
- the uncertainty in the yield of the process.
- the worker-absentheism.

The finished items are added to the inventory somewhere between the start and the end of the production run (duration: T_j). Since the increase in inventory will be irregular, and the actual pattern depends on the change-overs that are decided upon later, it will be difficult to estimate the right incremental pattern. A simple approximation of this incremental pattern is given by the dotted line in Figure 6.6. This dotted line corresponds to the assumption that the whole batch will enter the inventory at the end of the production run. The actual inventory will be higher throughout the production run, which reduces the number of actual stock-outs. This effect may be accounted for in the cost function related to the inventories of the products. As another extreme, one might consider the model in which the batch is added to the inventory at the start of the production run. Other approximations would have been possible. However, the results of previous Chapters can be applied most straightforwardly if we assume the batch to arrive at the inventories as a whole. Notice that the actual costs may be expected to lie between the two mentioned extremes (adding the batch to the inventories at the start of a production run or at the end of a production run).

We have now described the service mechanism, the demand and the cost rate. This enables us to apply the results of the previous Chapters in order to propose a Material Coordination System for the plastic products factory. As was mentioned, the different capacity groups can be treated separately on this level of Material Coordination. Therefore, we will discuss the Material Coordination System for only one of the capacity groups.

In the low season, we start by distinguishing a group of fast-movers and a group of slow-movers. This decomposition is not only based on the expected demand patterns, but also on the risk of obsolescence for each

family. The seasonal stock is built up for the fast-movers, whereas the slow-movers will have priority on the capacity.
The classification of a family as either a fast-mover or as a slow-mover does not change over the season, unless obsolescence of a fast-mover requires it.

In order to ensure that the raw materials are on the workfloor in time, and that the right moulds are available, Material Coordination plans always two production runs, which take about a week, ahead (on each capacity-group). That means that Material Coordination sets a time for the start of the first production run and Material Coordination decides which family will be produced after this run is finished.
In order to ensure that seasonal stocks are accumulated, we use a cost rate to the aggregate inventory position for the fast-movers that puts a penalty on deviating from the seasonal stock, as has been set by Master Planning.

To constitute a plan for each slow-mover, a simple product-oriented approach is used, which leads to the analysis of one-dimensional optimization models (one for each slow-mover). Via simple methods, we then coordinate the production plans for slow-movers. Thus, we find that the capacity that is left for the fast-movers is more tight. Therefore, we use a capacity-oriented approach in order to constitute a production plan for the fast-movers (or to determine critical levels), taking into account the remaining pattern of capacity-availability. Consequently, we aggregate over the fast-movers.

6.5 Conclusions.

Material Coordination as described here, is relatively simple. Only one-dimensional optimization models have to be analyzed. This will make it easy to implement the proposed method of Material Coordination.

The introduction of an explicit measure for the amount of capacity that is buffered in the inventories (the aggregate inventory position for fast-movers) has three advantages:

Firstly, safety stocks can be accumulated more efficiently. The difference between product-oriented uncertainty and capacity-oriented uncertainty will be used when buffering against uncertainty. The results of the previous Chapters indicate that this leads to efficient buffer stocks.

Secondly, the information that Material Coordination requires with respect to individual families only has a short-term character, whereas over a longer horizon only aggregate information is necessary (when a cost function is determined for the aggregate inventory for fast-movers, that should represent the cost of capacity-problems for the future). This information is relatively reliable. If a product-oriented approach would also be used for fast-movers, then detailed information would be required for fast-movers over a long period in order to avoid capacity problems.

Finally, we want to mention the advantage of introducing an explicit measure for the amount of buffered capacity with respect to the coordination of Material Coordination and Master Planning.
Master Planning has to ensure that the resources are obtained and used effectively and efficiently. This will be more easy, since it can explicitly be measured what effects changing the resource-availability has on the performance of Material Coordination. This also makes budgeting and performance-measurement for Material Coordination more easy.
If a product-oriented approach to designing Material Coordination were used in case the capacity restriction plays a role, a separate aggregate model would be necessary on the level of Master Planning for these purposes, with all inherent problems for coupling this with the level of Material Coordination.

The above mentioned, three advantages of introducing a measure for the amount of stored capacity at the level of Material Coordination, may be expected not only to hold for the example that has been treated here, but in a more general class of situations, where capacity restrictions play an important role.

References.

[1] R.L. Ackoff, A concept of Corporate Planning, J. Wiley and Sons, New York, 1970.

[2] J.E. van Aken, On the control of complex industrial organisations, Ph.D. Thesis, Stenfert Kroes, Leiden/Boston/London, 1978.

[3] R.N. Anthony, Planning and Control Systems: A framework for analysis, Boston, 1965.

[4] S. Axsäter and H. Jönsson, Aggregation and disaggregation in hierarchical production planning, European Journal of Operational Research, vol 17 (1984), pp. 338-350.

[5] K.R. Baker, An analysis of terminal conditions in rolling schedules, European Journal of Operational Research, vol 7 (1981), pp. 355-361.

[6] P. van Beek, An application of decomposition and dynamic programming to the N-product, 1-machine problem, Methods of Operations Research, volume 41 (1981), pp. 63-67.

[7] W.L. Berry, T.E. Vollmann and D.C. Whybark, Master Production Scheduling: Principles and Practice, APICS, 1979.

[8] J.W.M. Bertrand, A hierarchical approach to structuring the goods flow control in multi-product multi-phase production systems, appeared in "Modelling Production Management Systems" by P. Falster and R.B. Mazumder (eds), Elsevier , 1985.

[9] J.W.M. Bertrand and J.C. Wortmann, Production Control and Information Systems for component manufacturing shops, Ph.D. Thesis, Elsevier, 1981.

[10] J.W.M. Bertrand and J. Wijngaard, The structuring of production control systems, TH-report BDK/ORS/84/10, 1984.

[11] P.J. Billington, J.O. McClain and L.J. Thomas, Mathematical programming approaches to capacity-constrained MRP systems: Review, Formulation and Problem reduction, Management Science, vol 29 (1983), pp. 1126-1141.

[12] G.R. Bitran and A.R. von Ellenrieder, A hierarchical approach for the planning of a complex production system, appeared in "Disaggregation: Problems in Manufacturing and Service Organizations" by L.P. Ritzman, et al. (eds), Martinus Nijhoff Publishing, Boston/ The Hague/ London, 1979.

[13] G.R. Bitran and A.C. Hax, On the design of hierarchical production planning systems, Decision Sciences, vol 8 (1977), pp. 28-55.

[14] M.A. de Bodt and L.N. van Wassenhove, Experimental and analytic evaluation of cost increases due to demand uncertainty in MRP lot sizing with and without buffering, Working paper 81-81, University of Leuven, 1981

[15] J.L. Burbidge, The principles of production control, McDonnald Evans Ltd., London, 1971.

[16] A.J Clark and H. Scarf, Optimal policies for a multi-echelon inventory problem, Management Science, vol 6 (1960), pp. 475-490.

[17] R. van Dierdonck and W. Bruggeman, Integration problems in materials management: an overview, paper presented at the ORSA-TIMS conference in Lausanne, july 1982.

[18] P.P.J. Durlinger, A Trade-off module (in Dutch), TH-Report BDK/KBS/84/06, Eindhoven University of Technology, 1984.

[19] S.E. Elmaghraby, The economic lot scheduling problem (ELSP): Review and extensions, Management Science, vol 24 (1978), pp. 587-598.

[20] A. Federgruen and P. Zipkin, Approximations of dynamic, multilocation production and inventory problems, Management Science, vol 30 (1984), pp. 69-84.

[21] J.W. Forrester, Industrial Dynamics, J. Wiley and Sons, New York, 1961.

[22] H. Gabbay, Optimal Aggregation and Disaggregation in hierarchical planning, appeared in "Disaggregation: Problems in Manufacturing and Service Organizations" by L.P. Ritzman et al. (eds), Martinus Nijhoff Publishing, Boston/ The Hague/ London, 1979.

[23] J.R. Galbraith, Designing complex organisations, Addison-Wesley, London, 1973.

[24] S.C. Graves, The multi-product production cycling problem, AIIE Transactions, vol 12 (1980), pp.133-240.

[25] J.H. Greene, Production and Inventory Control: systems and decisions, R.D. Irwin Inc., Homewood, 1974.

[26] J. van der Griend, Optimalisation of functions of one variable (in Dutch), Ph.D. Thesis, Leiden, 1978.

[27] G. Hadley and T.M. Whitin, Analysis of Inventory Systems, Prentice-Hall Inc., Englewood Cliffs, 1963.

[28] A.C. Hax and H.C. Meal, Hierarchical integration of production planning and scheduling, appeared in "Studies in Management

Sciences", vol 1: logistics by M.A. Geisler (ed), North-Holland/TIMS, 1975.

[29] R.N. van Hees, Organisatiestructuur en integrale besturing als basis voor voorraadbeheersing (in Dutch), Organisatie en integrale besturing, vp45n, organisatie & efficiency, informatiecentrum, Philips, Eindhoven.

[30] C.C. Holt, F. Modigliani, J.F. Muth and H.A. Simon, Planning Production, Inventories, and Work Force, Prentice-Hall, 1960.

[31] H. Jönsson, Simulation studies of hierarchical systems in production and inventory control, Ph.D. thesis, Linköping University, 1983.

[32] U.S. Karmarkar, A hierarchical approach to multilocation inventory systems, appeared in "Disaggregation: Problems in Manufacturing and Service Organizations" by L.P. Ritzman et al. (eds), Martinus Nijhoff Publishing, Boston/ The Hague/ London, 1979.

[33] A.Y. Khintchine, Mathematical methods in the theory of queueing, appeared in "Griffin's statistical monographs & courses" by M.G. Kendall (ed), London, 1960.

[34] G.E. Kimball, General principles of Inventory Control, unpublished paper, Arthur D. Little Inc., Cambridge Massachussettes, 1955.

[35] G. Liesegang, Aggregation bei linearen Optimierungsmodellen: Beitrage zur Konzipierung, Formalisierung und Operationalisierung, Ph.D. Thesis, Köln, 1981.

[36] S.A. Lippman, Applying a new device in the optimization of exponential queueing systems, Operations Research, vol 23 (1975), pp. 687-710.

[37] J.F. Magee, Production Planning and Inventory Control, McGraw-Hill, New York, 1958.

[38] J. Manz, Zur Anwendung der Aggregation auf mehrperiodische lineare Produktionsprogrammplanungsprobleme, Ph.D.-thesis, Köln, Europäische Hochschulschriften: Reihe 5, 1983.

[39] J.O. McClain and J. Thomas, Horizon effects in aggregate production planning with seasonal demand, Management Science, vol 23 (1977), pp. 728-736.

[40] H.C. Meal, A study of multi-stage production planning, appeared in "Studies in Operations Management" by A.C. Hax (ed), North Holland, Amsterdam, 1978.

[41] T.E. Morton, Universal planning horizons for generalised convex production scheduling, Operations Research, vol 26 (1978), pp. 1046-1058.

[42] M.F. Neuts, Matrix-geometric solutions in stochastic models: an algorithmic approach, The Johns Hopkins University Press, Baltimore, 1981.

[43] H.L.W. Nuttle and J. Wijngaard, Planning horizons for manpower planning: a theoretical analysis, OR Spektrum, vol 3 (1981), pp. 153-160.

[44] J. Orlicky, Material Requirement Planning, McGraw-Hill, New York, 1975.

[45] L.P. Ritzman, L.J. Krajewski, W.L. Berry, S.H. Goodman, S.T. Hardy and L.D. Vitt (eds), Disaggregation: Problems in manufacturing and service organizations, Martinus Nijhoff Publishing, Boston/ The Hague/ London, 1979.

[46] S.M. Ross, Applied probability models with optimization

applications, Holden-Day, San Fransisco/ Cambridge/ Londen/ Amsterdam, 1970.

[47] B.L. Schwartz, A new approach to stockout penalties, Management science, vol 12 (1966), pp. B538-544.

[48] E. Seneta, Non-negative Matrices, George allen & Unwin Ltd, London, 1973.

[49] S. Stidham, Jr., A last word on L = λW, Operations Research, vol 22 (1974), pp. 417-421.

[50] J. Stoer and R. Bulirsch, Introduction to numerical analysis, Springer Verlag, Berlin/ Heidelberg/ New York, 1976.

[51] J.P.J. Timmer, W. Monhemius and J.W.M. Bertrand, Production and inventory control with the Base Stock System, paper presented at the ORSA-TIMS conference in Copenhagen, june 1984.

[52] G.U. Tinarelli, On the stockout cost implicit in fixing the level of service, paper presented at ORSA-TIMS conference in Lausanne, july 1982.

[53] J. van der Wal, Stochastic Dynamic Programming, Ph.D. thesis, Mathematical Centre, Amsterdam, 1981.

[54] D.C. Whybark and J.G. Williams, Material Requirements Planning under uncertainty, Decision Sciences, vol 7 (1976), pp. 595-606.

[55] T.M. Williams, Special products and uncertainty in production/inventory systems, European Journal of Operational Research, vol 15 (1984), pp. 46-54.

[56] D.A. Wismer (ed), Optimization methods for large-scale systems ... with applications, McGraw-Hill, New York, 1971.

[57] J. Wijngaard, Decomposition for dynamic programming in production and inventory control, Engineering and process economics, vol 4 (1979), pp. 385-388.

[58] J. Wijngaard, On aggregation in production planning, Engineering Costs and Production Economics, vol 6 (1982), pp. 259-265.

[59] J. Wijngaard, Stationary Markovian Decision Problems, Ph.D. thesis, Eindhoven University of Technology, 1975.

[60] E. Zabel, Some generalisations of an inventory planning horizon theorem, Management Science, vol 10 (1964), pp. 465-471.

[61] P.H. Zipkin, Simple ranking methods for allocation of one resource, Management Science, vol 26 (1980), pp. 34-43.

[62] P.H. Zipkin, Exact and approximate cost functions for product aggregates, Management Science, vol 28 (1982), pp. 1002-1012.

On the capacity-aspect of inventories.

Summary.

In this study, we have considered the Material Coordination level of Production Control. On this level different Production Units are discerned and the flow of material over these Production Units is coordinated (compare well-known Material Coordination Systems, like Material Requirements Planning, the Reorder Point System or the Base Stock System).

The performance of a given Material Coordination System depends not only on the characteristics of the Material Coordination System, but also on the characteristics of the environment it should work in. Therefore, a trade-off has to be made between investments that are necessary to reduce uncertainty and investments to be able to cope with existing uncertainty (e.g. safety stocks or flexible resources).
In this text, we have concentrated on the way one can create safety stocks to protect against uncertainty on the level of Material Coordination, efficiently. The results of this text may then be used in making this more general trade-off.
One may distinguish two fundamentally different approaches to create such stocks, namely:

- the product-oriented approach. In this approach, first the delivery patterns for individual products are translated to production patterns, using standard throughput-times for orders. Next, the different production patterns are coordinated over a shorter horizon, taking the capacity restrictions into account. Uncertainty in the required delivery pattern and the availability of capacity, can be attacked in the first step, so per product.

-the capacity-oriented approach. In this approach, first a production level plan is made, possibly combined with a capacity adjustment plan. This requires aggregation of delivery patterns and inventories to capacities. Next, over a shorter horizon, the aggregate production plan is disaggregated to individual products. Uncertainties in the aggregate delivery patterns and in the availability of capacities can be accounted for in the first step, which leads to aggregate safety stocks.

In this text, we have compared both approaches at hand of the single-phase multi-product model (with one clear capacity bottle-neck). The reason to consider this model is that it is the most straightforward starting point for the analysis of the weak and strong points of the approaches.

After discussing a framework for Production Control in which the level of Material Coordination can be embedded, we have compared the two different approaches for some single-capacity models. First, we have considered a single-phase model in which both the demand for individual products and the availability of the capacity are purely stochastic. We have compared the two approaches by using simulation experiments. This has led us to an operational criterion for choosing between the two approaches. This criterion has also been checked in a situation where demand is partly known beforehand and the availability of the capacity is deterministic. We have compared different ways of using the information that is available about future demand in this model, and it proved that the choice between the capacity-oriented approach and the product-oriented approach was not influenced by this. Therefore, the criterion could be used again.
Also, we have considered a model, in which the demand rate varies widely among the products. It has been suggested to use a capacity-oriented approach for the fast-movers in such situations and a product-oriented approach for the slow-movers, after giving the slow-movers priority on the capacity.
Finally, we have discussed a case of a plastic products factory. At hand of this case, we have shown how the results of this study can be used in a practical situation.

Het capaciteitsaspect van voorraden.

Samenvatting.

In dit proefschrift beschouwen we het niveau van Materiaal Coördinatie binnen produktiebeheersing. Op het niveau van Materiaal Coördinatie worden diverse produktie-eenheden onderscheiden in het produktieproces, en moet de goederenstroom over deze eenheden gecoördineerd worden. Enkele bekende Materiaal Coördinatie Systemen zijn "Material Requirements Planning", het "Reorder Point System" en het "Base Stock System".

De doelmatigheid van een gegeven Materiaal Coördinatie Systeem hangt niet alleen af van het systeem zelf, maar ook van de omgeving waarbinnen het systeem moet werken. Daarom dient een afweging gemaakt te worden tussen investeringen die nodig zijn om te zorgen dat de onzekerheden van de omgeving gereduceerd worden en investeringen die nodig zijn om Materiaal Coördinatie te beschermen tegen de gevolgen van de bestaande onzekerheden (bv. door het creëren van veiligheidsvoorraden of het werken met flexibele capaciteiten).
In dit proefschrift, beschouwen we de vraag hoe men efficient veiligheidsvoorraden kan creëren op het niveau van Materiaal Coördinatie. De resultaten van dit proefschrift kunnen zodoende gebruikt worden bij het maken van de genoemde afweging.
Men kan twee fundamenteel verschillende aanpakken onderscheiden bij het creëren van dergelijke voorraden, namelijk:

- de produkt-georiënteerde aanpak. In deze aanpak worden eerst de afleverpatronen voor individuele produkten vertaald naar produktiepatronen. Hierbij wordt gebruik gemaakt van standaard doorlooptijden voor orders. In een tweede stap worden de verschillende produktiepatronen gecoördineerd, waarbij de capaciteitsrestricties in beschouwing worden genomen. De horizon

in deze tweede stap is meestal kleiner dan in de eerste stap. Onzekerheid, zowel met betrekking tot de vereiste afleverpatronen als met betrekking tot de beschikbaarheid van capaciteiten, dient te worden opgevangen in de eerste stap. Dit leidt tot veiligheidsvoorraden per produkt.

- de capaciteit-georiënteerde aanpak. In deze aanpak wordt eerst beslist hoe men de capaciteiten zal gebruiken. Dit wordt beslist op grond van een vergelijking van de beschikbaarheid van de capaciteiten en de vraag naar capaciteiten. Daartoe dienen de afleverpatronen en de voorraden geaggregeerd te worden naar capaciteiten. Na het opstellen van een aggregaat produktiepatroon, wordt dit patroon over een kortere horizon gedisaggregeerd naar individuele produkten. Onzekerheid met betrekking tot aggregaat afleverplannen en met betrekking tot beschikbaarheid van capaciteiten moet worden opgevangen in de eerste stap. Dit leidt tot aggregaat veiligheidsvoorraden.

In dit proefschrift, hebben we beide aanpakken met elkaar vergeleken aan de hand van het één-capaciteit meer-produkten model (met één capaciteit bottle-neck). We hebben dit model beschouwd, omdat het het meest simpele model is waarin men onderscheid kan maken tussen de beide verschillende aanpakken voor het creëren van veiligheidsvoorraden.

Nadat we een raamwerk voor produktiebeheersing hebben beschreven, waarbinnen het niveau van Materiaal Coördinatie een plaats inneemt, hebben we de beide aanpakken vergeleken voor enkele verschillende voorbeelden van het één-capaciteit meer-produkten model. Eerst hebben we het model beschouwd, waarin zowel de vraag naar produkten als de beschikbaarheid van de capaciteit volledig stochastisch is. Met behulp van simulatie hebben we de twee genoemde aanpakken vergeleken. Dit heeft ertoe geleid om een operationeel criterium voor de keuze tussen beide aanpakken op te stellen. Dit criterium hebben we vervolgens gebruikt in een model waarin de vraag, gedeeltelijk, van te voren bekend is en waarin de beschikbaarheid van de capaciteit deterministisch is. De resultaten tonen aan dat de keuze tussen de capaciteit-georiënteerde en de produkt-georiënteerde aanpak niet afhangt van de wijze waarop men gebruik maakt van de informatie die

beschikbaar is over de toekomstige vraag. Dit maakt het mogelijk om het criterium ook voor dit model te gebruiken.
Ook hebben we een model beschouwd, waarin de gemiddelde vraag sterk fluctueert over de verschillende produkten. Voor een dergelijke situatie hebben we voorgesteld om de snellopers te beheersen met behulp van een capaciteit-georiënteerde aanpak en de langzaamlopers te beheersen met behulp van een produkt-georiënteerde aanpak (terwijl de langzaamlopers voorrang krijgen op de capaciteit).
Tenslotte, hebben we een case beschouwd van een fabriek, waarin plastic voorwerpen geproduceerd worden. De analyse van deze case toont aan hoe de resultaten van dit proefschrift in praktijksituaties gebruikt kunnen worden.

Curriculum Vitae.

De schrijver van dit proefschrift werd op 1 september 1959 te Mheer geboren. In 1977 behaalde hij het diploma Gymnasium-β aan het St. Adelbertcollege te Wassenaar en begon hij zijn studie in de Wiskunde aan de Rijksuniversiteit te Leiden. Hij slaagde voor het kandidaatsexamen "Wis- en Natuurkunde met Sterrenkunde" in 1979. Van september 1979 tot juni 1981 was hij werkzaam als student-assistent bij het Instituut voor Toegepaste Wiskunde en Informatica van de Rijksuniversiteit te Leiden. In juni 1981 sloot hij een afstudeeronderzoek af bij Prof. dr. A. Hordijk, en slaagde hij (cum laude) voor het doctoraalexamen Wiskunde met bijvak Informatica.
Op 8 juni 1981 trad hij in dienst van de Onderafdeling der Bedrijfskunde van de Technische Hogeschool te Eindhoven, waar hij onder leiding van Prof. dr. J. Wijngaard gewerkt heeft aan het onderzoek waarvan de resultaten in dit proefschrift zijn vastgelegd.

STELLINGEN

behorende bij het proefschrift

ON THE CAPACITY-ASPECT OF INVENTORIES

van

R.P.H.G. Bemelmans

28 mei 1985

I

Het aanleggen van capaciteit-gerichte voorraden past niet binnen het kader van de MRP-filosofie.

II

Beschouw het volgende probleem (zie Dantzig et al. [1]):
Een reder mag kiezen welke havens zijn vrachtschip aandoet en in welke volgorde. Door een reis van haven i naar haven j verdient hij a_{ij} guldens. De reis duurt t_{ij} dagen (inclusief laden en lossen). Stel er zijn n havens met $|a_{ij}| \leq \psi$ en $|t_{ij}| \leq \tau$ voor alle (i,j). Om de winst per dag te minimaliseren, suggereert Lawler [4] een algoritme dat $O(n^3 \log n)$ tijd vergt. Gebruikmakend van een resultaat van Karp [3] is het mogelijk een algoritme te vinden dat $O(n^3)$ tijd vergt (zie Hordijk en Bemelmans [2]).

[1] G.B. Dantzig, W. Blattner and M.R. Rao, Finding a Cycle in a Graph with Minimum Cost to Time Ration with Application to a Ship Routing Problem, appeared in "Theory of Graphs" by P. Rosenstiehl (ed), Dunod, Paris, and Gordon and Breach, New York, 1976.

[2] A. Hordijk and R. Bemelmans, interne notitie, Rijksuniversiteit te Leiden, 1980.

[3] R.M. Karp, A characterization of the minimum cycle mean in a Digraph, Memorandum No. UCB/ERL M77/47, Electronics Research Laboratory.

[4] E.L. Lawler, Combinatorial Optimization: Networks and Matroids, Holt-Rinehart-Winston, New York, 1976.

III

X Door bij het bespreken ven een raamwerk voor "Master Production Scheduling" voorbij te gaan aan het verschil tussen "structurele" en "operationele" coordinatie, worden praktische problemen geintroduceerd.

[1] W.L. Berry, T.E. Vollmann and D.C. Whybark, Master Production Scheduling: Principles and Practice, APICS, 1979.

[2] J.W.M. Bertrand and J. Wijngaard, The structuring of production control systems, rapport Technische Hogeschool Eindhoven, BDK/ORS/84/10, 1984.

IV

Het standaardpakket CAN-Q, waarmee men doorlooptijden in productieprocessen kan bepalen, is slechts in zeer speciale gevallen toepasbaar.

[1] J. Solberg, CAN-Q User's guide, Purdue University, 1980.

V

Bij de beschrijving van Cox-verdelingen, wordt vaak gebruik gemaakt van het beeld van een aantal exponentiele fasen in serie, waarbij na iedere fase met een bepaalde kans (afhankelijk van de fase) geloot wordt of de volgende fase nog doorlopen wordt. Indien men echter gebruik wil maken van het feit dat de Cox-verdelingen dicht liggen in de verzameling van verdelingen die alleen gewicht leggen op de niet-negatieve rechte, dan moet de kans in het lotingsmechanisme complex gekozen worden, hetgeen de beschreven interpretatie bemoeilijkt.

[1] D.R. Cox, A use of Complex probabilities in the theory of Stochastic Processes, Proceedings of the Cambridge Philosophical Society, vol 51 (1955), pp. 313-319.

VI

Dat het verstandiger is een doorzichtige regel te gebruiken dan een (vermeend) optimale regel door te drukken, wordt bewezen door de problemen rond de invoering van de ~~tweeverdienersregeling~~. voordeurdelers

VII

Als de "struggle for life" een essentieel element is van iedere evolutie, dan moeten we erg oppassen met het zoeken van contact met buitenaardse beschavingen.

VIII

Zoals het beste orkest niet bestaat uit een samenstel van de beste solisten, zo bestaat het beste hierarchische gestructureerde beheersingssyteem niet uit een samenstel van de beste beheersingsregels op ieder niveau.

IX

Nu medewerkers aan Universiteiten en Hogescholen met de term (hoofd) docent worden aangeduid, is het de hoogste tijd om de didactische bekwaamheden van deze medewerkers te ontwikkelen.

X

Iemand die een "practicality gap" wil vullen, wordt vaak gezien als een "practical joker".