Logist. Res. (2009) 1:163–172
DOI 10.1007/s12159-009-0017-6

ORIGINAL PAPER

Planning and scheduling production systems from a logistics perspective

Timm Gudehus · Herbert Kotzab

Received: 3 June 2008 / Accepted: 12 August 2009 / Published online: 10 September 2009

Abstract Production systems are special performance systems which generate physical goods from input material. Therefore, the methods of system analysis, the limit performance and queuing laws and the strategies, which have been developed for general performance systems such as logistics systems, can be applied to production systems. This article describes strategies and standard procedures for production planning and scheduling, which result from the logistic approach.

Keywords Production logistics · Planning procedures · Scheduling strategies · Dynamic scheduling

1 Introduction

Since the 1950s, academic research has examined the field of production planning, scheduling and control from theoretical, methodological, practical, and empirical aspect (see e.g. [21]). Literature offers insights into the topic based on case studies representing different types of production companies, such as garment manufacturers, beverage production, pharmaceutical industry, mechanical or electrical engineering, or for different type of manufacturing and/or planning systems such as automated manufacturing systems or hierarchical planning systems (e.g. [3, 12, 13, 14, 19, 23]). Research also looks at the production planning problem whether it occurs for one or multiple products respectively in one or multiple periods (e.g. [5]). A major interest lies in the optimisation of production systems (e.g. [1, 21]).

From a logistical point of view, production systems are part of the general supply network of industry, trade and consumers. They are central units of the business networks of manufacturing companies and transform input material into physical goods. Accordingly, production logistics is dealing with planning, organising and scheduling of production orders and assembling orders [9]. As logistic networks supply the input and distribute the output of production systems, production and logistics are closely interrelated. Production planning without taking logistics into account is as incomplete as logistics without considering production.

This paper shows that many strategies, methods and procedures of logistics, in particular the limit performance laws, can be applied also for production planning and scheduling [1, 2, 11]. As compared to other production planning frameworks (see e.g. [16]) the proposed procedure has been derived as well from theoretical considerations as from practical insights stemming from successful implementation and experiences from several industries such as tobacco, chemical, beverages, plastic, metal, machines and car industry.

First, an overview to modes and types of production systems is given. This is followed by the presentation of the general limit performance law which holds for any performance station. By application of the limit performance law to elementary production stations, strategies and rules for planning, scheduling and optimisation of production systems are derived. The next section presents procedures for production planning and scheduling from a strategic and tactical point of view. Furthermore, links to scheduling

T. Gudehus
Strandweg 54, Hamburg, Germany
e-mail: Tgudehus@aol.com

H. Kotzab (✉)
Department of Operations Management,
Copenhagen Business School, Frederiksberg, Denmark
e-mail: hk.om@cbs.dk

of procurement and dispatch are discussed and strategies for the avoidance of bottlenecks are developed. The paper ends with a summary and an outlook for future research.

2 Modes and types of production

Production systems are multi-stage networks of elementary production stations which are either directly linked by conveyors and transport means or indirectly connected via intermediate buffers and storages. A single production station transforms input, such as raw material, parts, pre-products or modules, with the help of people, machines and other resources into products. The production units PU_r and quantities m_r [PU_r/PO] of the products P_r, $r = 1,2, \ldots, N_P$, and the required completion dates are given by production orders [PO]. A production order inflow λ_{PO} [PO/PE] requiring the mean order quantities m_r for products P_r causes a partial production demand $\lambda_r = m_r \cdot \lambda_{PO}$ [PU_r/PE].

The task of production planning and scheduling (PPS) is to enable and to ensure the execution of incoming orders in due time at lowest costs. The solution of this task depends on the production mode and on the type of the production system [6, 8, 10–12, 15]. The basic production modes are:

- Cyclic production: In a discontinuous process, discrete product units, which can be single pieces or filled packages, bottles or load units, are generated within certain cycle times.
- Process production: In a steady process, continuous goods, such as gas, liquids, bulk or long material, are produced during longer time without interruption.

The discrete results of a cyclic production are immediately conveyed to other production stations, directly delivered to customers or kept for short time in a buffer or for longer time in a store. The continuous output of process production is generally buffered in tanks, silos or stores for mass products, from which succeeding production, consumption, packing or bottling stations are supplied.

The different possibilities of coupling single production stations define the basic production types [17, 20, 22]:

- In workshop production, the single production stations are separated by order buffers and material buffers. The workshop stations generate, parallel or in sequence, the required goods in isolated part processes.
- In line production, several production stations are directly connected in sequence without buffers and supplied from outside with input material via intermediate buffers. They generate products in successive process steps.
- In network production, a buffer-less main performance chain generates the final products stepwise from input material, modules and parts, which are directly provided by supply chains without intermediate buffers.

Workshop production is more adequate for small orders with changing products, whereas line production becomes economic with larger orders for the same product. However, by reduction of setup times, standardisation of components and professional variant management, line production can also be made effective for smaller lots of changing products. Hence, the traditional equalisation of workshop production with small lots and of line production with large lots is no longer valid [24].

Nowadays, many different combinations and hybrid forms of the basic production types can be found in business practice. Some industries, such as the automotive industry or computer assembling, have aimed for an extended network production with just-in-time-provision of parts and modules. However, extreme JIT-production networks are generally unreliable and not cost-optimal [9].

3 Production capability

From a logistic aspect, a production system is a special performance system. It is a network of connected elementary production stations which are arranged in compounded stations, function modules or production plants. The single production stations generate from input material due to orders and commands tangible outputs, such as material, industrial products or consumer goods. For this task the production stations require resources, such as persons, areas, buildings, machines and equipment. The technological production process can be an extraction, manufacturing, refinement, machining, assembling or a filling process [19]. Production planning and scheduling has to ensure that the production orders are executed due to the promised delivery dates at lowest costs. For this purpose, the capability of all performance stations, which are necessary for the order execution, must be known.

An elementary production station PS(n, m), which transforms n partial input flows λ_q [MU_q/PE] of materials M_q, $q = 1,2, \ldots, n$, with material units MU_q into m partial output flows λ_r [PU_r/PE] of products P_r, $r = 1,2, \ldots, m$, with production units PU_r, is an irreducible performance station of type (n, m) and order $n + m$ as defined in [11]. It is not decomposable without loosing its function. The capability of such a station is determined by its limit performances, i.e. by the maximal input, output and/or throughput. They depend on the performance process and on the number of functions a station can execute. Hence, the formula for limit performances, the allocation, operating and scheduling strategies and the general limit performance and queuing laws of logistics can be transferred to production systems [2, 11].

3.1 Limit performances and switch times

For discrete goods, the input to a production station is specified by a bill of material, for continuous goods by a recipe. The bill of material respectively the recipe determines the quantities m_{qr} [MU$_q$/PU$_r$] of input materials M$_q$ measured in material units MU$_q$, which are necessary to generate one output unit PU$_r$ of product P_r. Using the bill of material, requirement planning (MRP) can calculate the input flows λ_q [MU$_q$/PE] of a production station from the output flows λ_r [PU$_s$/PE] by the sum:

$$\lambda_q = \sum_r m_{qr} \cdot \lambda_r \quad [\mathrm{MU}_q/\mathrm{PE}]. \tag{1}$$

The output flows λ_r are determined by the current production performances or by future requirements. They result either from the expected customer orders or from the supply orders of succeeding production stations, which are derived from customer orders with the corresponding bills of materials.

The production capability, i.e. the maximal periodical output of products P_r, is determined by the limit performances and switch times of the elementary performance stations of a production network:

- The partial production limit performance μ_r [PU$_r$/PE] is the maximal production performance if only product P_r is produced.
- The product switch time T_{rs} [PE] is the minimal time needed for changing or switching from product P_r to product P_s.

The product switch times are elements of the switch time matrix T_{rs}. They are setup times or changeover times, which in some cases can be longer than the pure technical setup times [11].

The preparation time until the first start of the station for orders of product P_r is the initial setup time. The initial setup time and the interruption time, that is needed regularly after a maximal output $m_{r\max}$ in order to clean and maintain the station, can be taken into account as eigen-set up time T_{rr} [PE] which are written in the diagonal of the switch time matrix.

The switch time multiplied with the cost rate of the station plus the costs for start-up losses, e.g. for waste, gives the switch costs, respectively the setup costs. They are of central importance for cost optimal scheduling of orders and inventories in supply chains. Analogous to the relation for the limit performance of a source [9, 11], the partial limit performances of a cyclic production are given by:

$$\mu_r = c/\tau_r(c) \quad [\mathrm{PU}_r/\mathrm{PE}]. \tag{2}$$

Herein, $\tau_r(c)$ is the cycle time for producing or the service time for processing a charge, lot or batch of c output units. For batch size $c = 1$, it is a single unit production and for batch sizes $c > 1$ a batch wise production. In a batch wise production, fixed quantities c of the same product are produced, processed and completed in one process cycle. The charge or batch size c is determined by the capacity of the production station, such as a melting furnace, kiln or electro-coating station, and by the capacity of the employed load carriers.

3.2 Limit performance law of production

The total utilisation of a production station is the sum of the productive utilisation ρ_{prod} for the generation of useful output and the unproductive utilisation ρ_{unpr} for setup und switching. The partial utilisations of a station with the partial outflows λ_r and limit performances μ_r are $\rho_r = \lambda_r/\mu_r$. A switching frequency ν_{rs} with switch times T_{rs} causes the switching occupation $\rho_{\mathrm{rs}} = \nu_{\mathrm{rs}} \cdot T_{\mathrm{rs}}$. The sum

$$\rho_{\mathrm{tot}} = \rho_{\mathrm{prod}} + \rho_{\mathrm{unpr}} = \sum_r \lambda_r/\mu_r + \sum_{r,s} \nu_{\mathrm{rs}} \cdot T_{rs} \quad [\%] \tag{3}$$

which includes all regular interruptions for cleaning and maintenance that happen with a cleaning frequency $\nu_{\mathrm{rr}} = \lambda_r/m_{r\max}$ and eigen-setup time T_{rr}, is the total utilisation of a production station [2, 10]. As the total utilisation cannot exceed 100%, i.e. due to $\rho_{\mathrm{tot}} \le 100\%$, relation (3) leads to the limit performance law of irreducible production stations

$$\rho_{\mathrm{prod}} = \sum_r \lambda_r/\mu_r \le 1 - \rho_{\mathrm{unpr}} = 1 - \sum_{r,s} \nu_{rs} \cdot T_{rs}. \tag{4}$$

From this limit performance law follow the availability principles:

- The productivity of a station, i.e. its availability for productive output, decreases with longer switch times and with increasing switching frequency.
- If the switch times between the products differ, the productivity depends also on the sequence of the product changes.

If the order inflow and/or the cycle times are stochastically fluctuating, an order waiting queue builds up in front of the production station. As demonstrated by Gudehus [10, 11], a stochastic waiting queue increases over proportionately with the utilisation (3) and grows infinitely when the utilisation approaches 100%. This causes waiting times at the bottleneck stations of a production chain or network, and extends the production lead times. When the total utilisation exceeds 100%, a systematic waiting queue grows and the order waiting time increases continuously until the total utilisation drops again below 100%. During the overload time the respective production station is a critical bottleneck.

3.3 Production lead times

The production lead time is the order lead time or delivery time for incoming orders from entering the production system to final execution. The minimal lead time for a product P_r with quantity m_r is the sum of a switch time and the order execution time:

$$T_{\mathrm{LTmin}}(m_r) = T_{rs} + m_r/\mu_r \quad [\mathrm{PE}]. \tag{5}$$

The switch time T_{rs} [PE] depends on the last product P_s. For $s = r$, it is the initial setup time. For $s \neq r$, it is the switch time from product P_s to product P_r. With limit performance μ_r [PU/PE], the execution time for a quantity m_r [PU] is m_r/μ_r [PE]. If the production is followed by a ripening process, the lead time is elongated by the ripening time. From relation (5) follow the lead time rules:

- The minimal production lead time increases with order quantity and decreases with higher limit performance.
- Long lead times of large orders can be shortened by execution in part orders on parallel stations or partly later on the same station.

The minimal lead time (5) can be achieved as long as the production station is free and no other orders with higher priority are waiting. If the station is occupied and other orders are waiting, the lead time is elongated by an order waiting time, which is the sum of the switch times $T_{\mathrm{ss}-1}$ and the execution times $T_s = m_s/\mu_s$ of the queuing orders for products P_s with quantity m_s. As ripening does not affect the running production, the order waiting time is not elongated by the ripening times of preceding orders.

The sum of the minimal lead times of the incoming order for product P_r and of all waiting orders with priority is the current production lead time:

$$T_{\mathrm{LTr}} = \sum_{s=1}^{r} (T_{\mathrm{ss}-1} + m_s/\mu_s) \quad [\mathrm{PE}]. \tag{6}$$

The first part of this sum is the setup time sum, the second part the execution time sum. From relation (6) follow the lead time principles:

- The order lead time of a production station depends on the number and size of the waiting orders with higher priority.
- The production lead time is elongated by higher switching frequency, i.e. by the number of waiting orders, and by longer switch times.
- With different switch times, the production lead time depends on the sequence, in which the waiting orders are executed.

Shorter production lead times can be achieved by part orders with smaller production lots. However, part orders increase the switching frequencies and reduce the availability for productive utilisation. The opposite influence of smaller lot sizes and higher switching frequencies on lead times and productivity causes the fundamental goal conflict between efficient production and short delivery times.

4 Production planning

The objective of production planning is the efficient and timely provision of resources and material for a future demand [9, 10]. The tasks of long-term planning are to set up the production network, to organise and optimise the production processes, to develop production strategies and to organise production scheduling. The tasks of medium- and short-term planning are advanced resource planning and material requirement planning [4, 17, 22, 26, 27].

Long-term production planning is performed regularly once a year or as often as required by management. A proven procedure of long-term planning is:

1. Optimisation of the existing production network, respectively, design of a new production network, i.e., selection of production technology; design of production stations; delimitation of parallel and consecutive production areas or the design of production lines and networks, which are directly connected by conveyors and transport means or decoupled by intermediate buffers and storages.
2. Documentation of network structure, specification of limit performances of available production stations and transport connections, and determination of capacities of the necessary buffers and storage systems.
3. Segmentation of the final and intermediate products into kanban-articles and parts, storekeeping parts and articles and order-articles or parts [10]:
    - Kanban-articles and parts are discrete mass products of low-value with continuous demand. As their consumption must not be registered on a single item basis, they can flow without IT-assistance due to the self-regulating pull-principle from entrance stations, production stations or storages directly to the consumption stations.
    - Storekeeping articles and parts are standard products with continuous demand and positive storekeeping profit. They can be produced or procured

anonymously and kept on stock. Their production orders result from dynamic inventory scheduling.
   - Order-articles and parts are the remaining final and intermediate products that are produced on the basis of customer orders

4. Definition of standard production chains and trees: Selection and design of cost optimal production chains or trees for the different product groups.
5. Specification and documentation of the production, performance and storage stations, following the order execution from the order penetration limit downstream to the exit station.
6. Organisation of production scheduling: Development of production strategies. Definition of the tasks of production scheduling. Synchronisation of periods, dates and strategies of production scheduling with order scheduling and logistic scheduling.

By these steps of long-term planning, the network structure and the process chains of production are established. The results are documented in structure charts and in process plans. For example, Fig. 1 shows the production structure and Fig. 2 the standard production chains of a metal processing company.

A process plan determines the sequence and connection of production and performance stations, specifies the technical processes and fixes the minimal lead times for representative product groups with standard order quantities. From the partial lead times of the stations, a standard delivery time can be calculated for the different process plans taking into account the mean transport and waiting times, which are expected at maximal tolerable utilisation.

The punctuality of production, i.e. the probability to keep standard delivery times, depends on the current demand and the variability of the orders quantities and lead times.

If a certain process step can be executed by several production stations, the corresponding process plan offers different options, but leaves open which one should be used. A standard process plan also does not fix the starting date and the finishing date. Medium- and short-term production planning is performed regularly in fixed planning cycles, e.g. weekly or monthly. Well proven steps of short-term planning are (see also [16, 28, 29]):

1. Advanced resource planning (ARP) and material requirement planning (MRP): The weekly or monthly demand for final products is forecasted or planned for the next 6 or 12 months. In a retrograde procedure, starting with the outflows of the exit stations, the partial product flows (1) of the upstream stations are calculated with the bill of material and allocated to the standard process chains or trees.
2. Announcement and check of resources and material: The results of ARP and MRP are announced to production and external suppliers, in order to check the required resources and capacities, and approve feasibility.
3. Identification, adjustment and elimination of bottleneck stations: Potential bottlenecks are stations that operate in peak times close to 100% utilisation. For critical bottlenecks, the expected utilisation exceeds 100% during unacceptable long time. In order to achieve a balanced utilisation, the limit performances of potential bottlenecks are adjusted. As far as possible, critical bottlenecks are eliminated.
4. Adjustment of operating and scheduling times: Working hours and operating times of the production stations are adjusted and synchronised to the expected demand.

The selection of the execution stations in case of several options and the determination of the execution dates for current orders are tasks of production scheduling. Only for major projects and for pre-production for sales actions and bottleneck phases, production planning fixes starting and finishing dates.

5 Production scheduling

5.1 Objective and tasks of production scheduling

The main objective of production scheduling is to ensure the cost optimal execution of current production orders in due time. For this purpose, order scheduling prioritises incoming external orders and disaggregates them due to appropriate scheduling strategies into external procurement orders for suppliers, internal production orders and delivery orders. The internal orders are transferred to the production, stores and commissioning systems, where they are executed. Additional tasks are procurement and dispatch scheduling [25].

The orders can be executed with promised punctuality, if production planning has provided the necessary resources and material. Essential for business success is the compatibility of the central order scheduling strategies with the strategies of inventory scheduling, order picking, procurement and production. Further critical factors are adjusted period lengths and synchronised starting times of operation and scheduling.

The goals of production scheduling are to ensure efficient order execution and to keep the agreed lead times. For this purpose, the current orders are allocated—based on adequate production strategies—to the available resources of production.

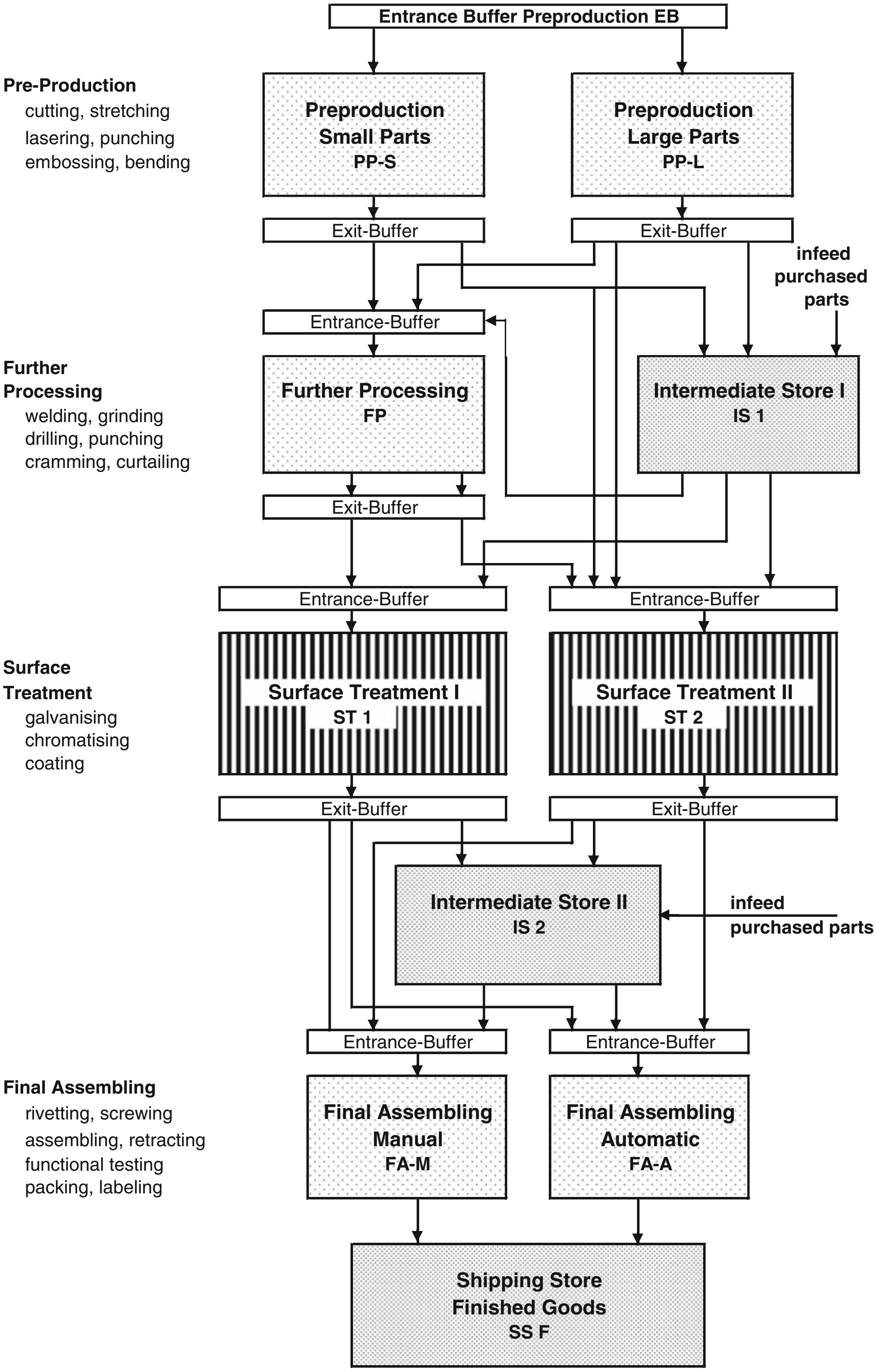


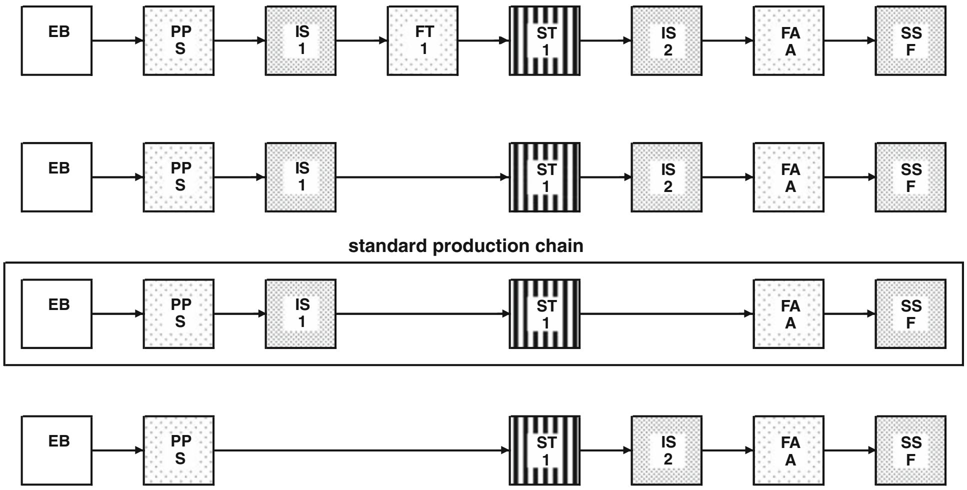
Fig. 1 Production stations and network structure of a metal processing company (Bottleneck stations: surface treatment)

5.2 Stock production or order production

Storekeeping articles can be delivered ex stock or produced to order. From the cost opportunity of storekeeping [9] result the delivery rules for storekeeping articles:

- Small orders with quantities less than half of the optimal replenishment quantity and smaller than the current stock should be delivered ex stock.
- Major orders with quantities larger than half of the optimal replenishment quantity or larger than the

Fig. 2 Standard production chains of a metal processing company (production structure, see Fig. 1, Bottleneck stations: surface treatment)



current stock should be produced to order if the required delivery time is longer than the production lead time.

- Major orders with delivery time shorter than the lead time are split into a make to stock-share, which is half the size of the optimal replenishment quantity and delivered immediately from stock, and a make to order-share that is produced to order and delivered later.

In the long run, these delivery rules achieve a cost-efficient share p of orders for order production and a remaining share $1-p$ for production on stock in a self-regulating way.

5.3 Dynamic production scheduling

Production scheduling can be performed statically in longer fixed scheduling cycles, e.g. weekly or monthly, or dynamically depending on the required punctuality or on single events, such as arrival of an urgent order or unexpected interruptions. For example, daily scheduling is necessary, if day-accurate delivery is required. Provided, demand does not exceed the production capacity and the required materials are available in due time, lowest production costs and punctuality can be achieved through the following standard production scheduling procedure:

1. For all production orders, which have arrived until the scheduling date, the latest starting date at the critical order penetration station is scheduled backward from the delivery date with the minimal lead times (5) of the involved stations.
2. The orders are sorted in urgency sequence with ascending starting dates. The orders with due starting date are scheduled with first priority for the next production period.
3. All due orders concerning the same product P_r are collected in series production orders $\mathrm{SO_r}$ with quantities m_r. They generate the required output flows $\lambda_r = m_r \cdot \lambda_{\mathrm{SOr}}$.
4. The production orders resulting from step 3 are assigned to the cost optimal standard process chain or tree.
5. Starting with the final station and progressing retrograde upstream to the order penetration points, the input flows (1) for the single stations of the production chain or tree, which are the output flows of the proceeding stations, are calculated with the help of the bill of material from the output flows.
6. For these output flows, the utilisations (3) of the production stations are calculated from the limit performances and switch times. In this way, the potential bottleneck, which is the station with highest utilisation, is identified.
7. If the switch times differ for different products, the switch time sum for the potential bottleneck station is minimised by optimal order sequencing.
8. If the total utilisation of the potential bottleneck station is less than the technical availability, e.g. $\eta_{\mathrm{avail}} = 95\%$, steps 3–6 are repeated after adding orders with descending urgency until the bottleneck utilisation reaches the technical availability or until all orders are scheduled.
9. If the total utilisation of the bottleneck station is above availability and its operation time cannot be extended, the station is a current bottleneck, which prevents the punctual execution of all due orders. In this case, less critical due orders have to be removed one by one, following appropriate priority rules, and the steps 3–6 are repeated until utilisation is below availability.
10. With the orders of the bottleneck station, the upstream stations are scheduled due to the pull-principle, the downstream stations due to the push-principle.
11. Finally, the resulting production orders are transferred to the internal production stations and the procurement orders to the suppliers and storage stations.

The steps of the standard production procedure are performed after the last scheduling period has ended and before the next period starts. This can be done manually by people only in longer scheduling cycles for small numbers of stations, orders and products. For shorter cycles, many orders, longer production chains and extended production networks, it requires suitable production planning and scheduling software.

5.4 Additional rules and strategies

When implementing the above dynamic production scheduling procedure, the following rules and strategies should be additionally taken into account:

- The length of the scheduling periods is determined by the required punctuality, e.g. when aiming at day-accurate delivery daily scheduling is necessary [8, 9].
- The scheduling periods of production should not be longer then the periods of order scheduling.
- If the limit performances refer to hours and switch times are measured in hours, formula (5) gives the lead time in hours.
- For calculating the utilisation (3), the daily production demand and flows have to be divided by the actual number of operating hours per day. This holds for fixed and for varying operating times.
- A breathing factory with flexible working hours adapts the operating times dynamically to the current performance demand and can prevent bottlenecks to a certain extend.
- Not only un-started orders but also the un-completed orders of the previous day are newly scheduled.

If a production network offers several production stations or chains for the same order with different costs, minimal costs are achieved by

- Cost optimal allocation where each order is allocated to the production station or chain, which can execute it at lowest order costs.

The order costs are the sum of the quantity-independent setup cost and the quantity-dependent execution costs. If the order costs for the optional stations are equal or unknown, optimal utilisation of parallel stations is achievable by

- Dynamic balanced allocation where the orders are allocated to the lowest utilised station. When the mean utilisation of all running stations exceeds 90% for a certain time, another station is added, if it is available. If the overall utilisation falls below $((N-1)/N)\cdot 80\%$, the station with the lowest utilisation is closed.

With dynamic adding and closing of stations, the mean utilisation is kept between 80 and 90% and longer waiting queues and waiting times are avoided.

6 Procurement and dispatch scheduling

Procurement scheduling and dispatch scheduling are closely connected with production scheduling: Procurement scheduling takes care of the cost optimal and punctual provision of material and parts from external suppliers. It receives orders from production and inventory scheduling. Dispatch scheduling is responsible for the cost optimal and punctual delivery of the finished orders to customers.

Basic strategies of procurement are procurement on stock and procurement on order, either advanced or just-in-time (see e.g. [7, 9, 18]). Dispatch scheduling can be performed by a scheduling centre, a local dispatch department or a logistic service provider who is responsible for storing, commissioning and delivery. In order to avoid self-optimisation and to ensure maximal consolidation and cost optimal dispatch, clear dispatch rules should guide the operations of the dispatch department or logistic service provider. Approved dispatch rules are:

- Several single customer orders that are due at the same delivery date are consolidated in a single shipment and delivered in the most cost-efficient manner.
- If customers agree upon fixed delivery dates, such as a delivery on every second day or once a week, all deliveries are collected in an intermediate buffer until the due date is reached and delivered as a whole.
- Further consolidation is achieved by collection of shipments destined for different customers in a region or country that can be shipped on the same day.

The consolidated shipments are picked up either by a parcel service provider or a freight forwarder in full or part truck loads and transferred to the next transhipment point. Shipments of sufficient quantity can be transported directly to the transhipment station in the target region. This is especially efficient for shipments to distant countries.

Order scheduling also decides whether articles produced on order should be shipped directly ex factory or via a dispatch logistic centre. This decision depends on the order content, the logistic costs and on the technical equipment in the production and in the logistic centre. If the proper equipment is available, standard rules for ex factory shipment are:

- Shipment ex factory is necessary for urgent single-item orders or part order deliveries, if the delivery date can only be kept by direct delivery.

- Direct delivery ex factory is more economic for large, heavy and bulky parts, machines and aggregates, which are packed for shipment at the end of the production.
- Direct delivery ex factory is faster and more economic for large quantities, which can be shipped in full pallets as part or full loads on trucks or trains.

In the last case, producing into the transport means is optimal. Standard rules for shipment via a logistic centre are:

- Delivery via a logistic centre is necessary for the complete delivery of multi-position orders with store-keeping articles and/or order articles from several production areas.
- Dispatch via a logistic centre is cheaper for small quantity orders due to more efficient storing, commissioning and packaging.
- A logistic centre enables consolidation of shipments to different customers in the same region of destination.

These dispatch rules as well as the standard production strategies can be realised by decision-based scheduling software, which proposes the cost optimal execution and delivery mode for each order.

7 Bottleneck strategies

Future peaks in demand and potential bottlenecks are identified in advance during the medium-term planning for the forecasted demand. The total demand $\lambda_{\mathrm{PS}}(t)$ for a production station PS in a future period t is the sum of the demand of all articles or products P_r which will be executed by this station:

$$\lambda_{\mathrm{PS}}(t) = \sum_r \lambda_r(t) \quad [\mathrm{PU/PE}]. \tag{7}$$

For the advanced identification of bottlenecks, the demand structure, given by the structural weights $g_r = \lambda_r/\lambda$, and the mean switch time loss per production unit τ_{STm} of the past can be assumed to hold also in the future. With these values the effective limit performance of a production station PS is:

$$\mu_{\mathrm{PS}} = \left(\sum_r g_r/\mu_r + \tau_{\mathrm{STm}}\right)^{-1} \quad [\mathrm{PE/PU}]. \tag{8}$$

In periods where the total demand does not exceed the effective limit performance, i.e. for $\lambda_{\mathrm{PS}}(t) < \mu_{\mathrm{PS}}$, all orders can be executed within the standard lead times. If for a bottleneck phase of length T_{BN}, the demand exceeds the limit performance (8), the production becomes a critical bottle neck. During the periods t of the bottleneck phase T_{BN} an order backlog accumulates which is equal to the excess demand:

$$D_{\mathrm{ex}} = \sum_{t \in T_{\mathrm{BN}}} (\lambda_r(t) - \mu_{\mathrm{PS}}) \quad [\mathrm{PU}]. \tag{9}$$

If the demand in previous periods is less then the limit performance, the order backlog can be reduced or even avoided by advanced production of bottleneck articles. Suitable scheduling software recognises and announces bottleneck phases, calculates the accumulating excess demand (9) and generates pre-production orders using the excess capacity in periods with low utilisation.

The disadvantage of pre-production for bottleneck phases is a higher stock with all its costs and risks. This leads to the pre-production recommendation:

- The scheduler, not the scheduling program should decide on advanced production or procurement in agreement with the sales department and management.

For different articles that are produced by the same bottleneck station holds the pre-production selection rule:

- Articles with the highest and most reliable sales frequency should be pre-produced first and kept on stock for the bottleneck phase.
- If two articles have equally high and reliable sales frequency, the one with the lower value should be pre-produced first in order to save inventory interest.

These selection rules prevent a production of non-saleable and high value articles and long storing times in case, that the demand does not reach the expected peak or even drops. The best selling articles can be produced in very cost-efficient lot sizes during periods of low utilisation.

If pre-production is impossible or insufficient and the production demand exceeds the limit performance for more than one period, the respective production station becomes a critical bottleneck. During this time, order backlogs and longer lead times are unavoidable. In such a case, the limited resources should be allocated to the customers by the scheduler, not by the program. In order to avoid conflicts and unfounded decisions, it is recommended to agree in advance about appropriate bottleneck allocation rules such as first come first served, profit margin, absolute profit, urgency, customer importance, out-of-stock costs or related to the regular demand. The fairest allocation is last rule, i.e., the limited production output is allocated in relation to demand of the regular customers in normal times.

If production supplies several consumption stations of an extended supply network, central scheduling is essential during bottleneck periods in order to keep the allocation rules. It holds generally that:

- Local scheduling by de-central consumption stations, e.g. by sales outlets or small customers, is satisfactory

if the products are distributed via a store-keeping logistic centre, as long as the production capacities are sufficient.
- Central scheduling is necessary for order products and if shortages and bottlenecks are expected.

Pre-production is not feasable, if the bottleneck phase lasts for a long time. Possibilities in such situation are increasing the prices, outsourcing of critical production or investment into new production facilities.

8 Conclusion

The purpose of this paper was to present how useful strategies, rules and standard procedures for production planning, scheduling and optimisation can be derived from the logistic aspect. They result from the general limit performance laws of logistics and offer feasible solutions for many practical production problems. It would be an interesting task for research to test the limit performance laws for the different production systems by stochastic simulation as it has been done successfully for logistic systems [1, 11].

References

1. Agrawal A, Minis I, Nagi R (2000) Cycle time reduction by improved MRP-based production planning. Int J Prod Res 38(18):4823–4841
2. Arnold D, Furmans K (2008) Materialfluss in Logistiksystemen, 4th edn. Springer/VDI, Berlin
3. Ashayeri J, Selen W (2003) A production planning model and a case study for the pharmaceutical industry in the Netherlands. Int J Logist Res Appl 6(1/2):37–41
4. Chopra S, Meindl P (2006) Supply chain management, strategy, planning & operation, 3rd edn. Pearson, Upper Saddle River
5. Christou I, Lagodimos A, Lycopoulou D (2007) Hierarchical production planning for multi-product lines in the beverage industry. Prod Plan Control 18(5):367–376
6. Comelli M, Féniès P, Tchernev N (2008) A combined financial and physical flows evaluation for logistic process and tactical production planning: application in a company supply chain. Int J Prod Econ 112(1):77–95
7. Dellaert N (1989) Multi-item production control for production to order. Eng Costs Prod Econ 17(1-4):167–173
8. Forsyth H, Porter K (2000) Production planning and control improvements in a small UK garment manufacturer; a case study. Prod Plann Control 11(1):617–625
9. Gudehus H, Kotzab H (2009) Comprehensive logistics. Springer, Berlin
10. Gudehus T (2006) Dynamische Disposition, Strategien zur optimalen Auftrags- und Bestandsdisposition, 2nd edn. Springer, Berlin
11. Gudehus T (1999/2005) Logistik, Grundlagen – Strategien – Anwendungen, 3rd edn. Springer, Berlin (the limit performance and queuing laws described here have been developed by T. Gudehus 1975–1976 and first published in ZOR Zeitschrift für Operations Research Band 20 1975, B37-61, B127-B160 and B207-B525)
12. Günther H-O, Tempelmeier H (2007) Produktion und Logistik, 7th edn. Springer, Berlin
13. Gunasekaran A, Goyal S, Virtanen I, Yli-Olli P (1993) Production planning and control in automated manufacturing systems. Prod Plann Control 4(1):2–16
14. Halsall D, Muhlemann A, Price D (1994) A review of production planning and scheduling in smaller manufacturing companies in the UK. Prod Plann Control 5(5):485–493
15. Hopp W, Spearman M (2000) Factory physics. McGraw-Hill, New York
16. Jodlbauer H (2008) Produktionsoptimierung: Wertschaffende sowie kundenorientierte Planung und Steuerung, 2nd edn. Wien, New York
17. Kern W (ed) (1979) Handwörterbuch der Produktionswirtschaft. Poeschel, Stuttgart
18. Ketzenberg M, Metters R, Semple J (2006) A heuristic multi-item production with seasonal demand. IEE Trans 38(3):201–211
19. Liberatore M, Miller T (1985) A hierarchical production planning system. Interfaces 15(4):1–11
20. Slack N, Chambers S, Johnston R (2004) Operations management, 5th edn. In: Harlow et al (eds) Financial time. Prentice-Hall, Englewood Cliffs
21. Sprague L, Ritzmann L, Krajewski L (1990) Production planning, inventory management and scheduling: spanning the boundaries. Manag Decis Econ 11(5):297–315
22. Stadtler H (2000) Supply chain management and advanced planning: concepts, models, software, and case studies. Springer, Berlin
23. Sumichrast R, Burch E (1985) Understanding hierarchical production planning and scheduling models. Ind Manage 27(3):6–11
24. Tempelmeier H (2005) Material-Logistik: Modelle und Algorithmen für die Produktionsplanung und -Steuerung in Advanced Planning-Systemen, 6th edn. Springer, Berlin
25. Vollmann T, Berry L, Whybark C (1997) Manufacturing planning and control systems, 4th edn. Irwin, Homewood
26. Wagner H, Whitin T (1958) Dynamic version of the economic lot size model. Manage Sci 5(1):89–96
27. Ware N, Fogarty D (1990) Master schedule/master production schedule: the same or different. Prod Inventory Manage J 31(1):34–38
28. Zäpfel G (2000) Strategisches Produktions-Management, 2nd edn. München, Oldenbourg
29. Zäpfel G (2001) Grundzüge des Produktions- und Logistikmanagement, 2nd edn. München, Oldenbourg