Load Dependent Lead Times – From Empirical Evidence to Mathematical Modeling

Julia Pahl, Stefan Voß, David L. Woodruff

Summary:
As organizations move from creating plans for individual production lines to entire supply chains it is increasingly important to recognize that decisions concerning utilization of production resources impact the lead times that will be experienced. In this paper we give some insights into why this is the case by looking at the queuing that results in delays. In this respect, special mention should be made that it is difficult to experience related empirical data, especially for tactical planning issues. We use these insights to survey and suggest optimization models that take into account load dependent lead times and related "complications."

Keywords:
Supply Chain Management, Load Dependent Lead Times, Lead Times, Tactical Planning, Aggregate Planning

1 Introduction

Let us define the lead time as the time between the release of an order to the shop floor or to a supplier and the receipt of the items. Lead time considerations are essential with respect to the global competitiveness of firms, because long lead times impose high costs due to rising WIP (work in process) inventory levels as well as larger safety stocks caused by increased uncertainty about production prerequisites and constraints. Despite this, considerations about load dependent lead times are rare in the literature. The same is valid for models linking order releases, planning and capacity decisions to lead times, and take into account factors influencing lead times such as the system workload, batching and sequencing decisions or WIP levels.

Present practice for manufacturing supply chains is dominated by the use of material requirements planning (mrp) with its inherent problems. Many companies do not use adequate planning tools at all. Accordingly, problems arise when fixed, constant or "worst case" lead times are assumed at an aggregate planning level, e.g., to have enough "buffer time" to securely meet demands. In order to meet due dates there is also a tendency to release jobs into the system much earlier than necessary, leading to very high WIP levels and, therefore, longer queuing (waiting times) causing even longer lead times. This overreactional behavior becomes a self-fulfilling prophecy and is addressed in the literature as the *lead time syndrome* which results from the fact that the relationship between WIP, output, workload and average flow times is ignored (Zäpfel & Missbauer, 1993; Tatsiopoulos & Kingsman, 1983). Moreover, most mrp and enterprise resource planning (ERP) models implement sequential planning algorithms which neither consider uncertainties nor resource and production flow constraints of raw material, WIP and finished goods inventory (FGI), leading to suboptimal or infeasible production plans (Caramanis & Ahn, 1999).

Another fundamental problem of manufacturing and production planning models is the omission of modeling nonlinear dependencies, e.g., between lead times and the workload of a production system or a production resource. This happens even though there is empirical evidence that lead times increase nonlinearly long before resource utilization reaches 100% (Asmundsson et al., 2003; Karmarkar, 1987); see Figure 1. This may lead to significant differences in planned and realized lead times. There is a lack of models allowing the analysis of behavior of lead times and WIP levels considering the facility workload under variable demand patterns like seasonal demand. In addition, it seems likely that queuing tends to be correlated so that a machine failure at one point of the system will cause queuing at other stations which leads to the presumption that lead time distributions tend to be fat-tailed and skewed. However, to the best of our knowledge there is no comprehensive (empirical) work on this topic currently available. Furthermore, it seems that there is no model which analyzes load dependent lead times in the context of stochastic demand evidently prevailing in practice.

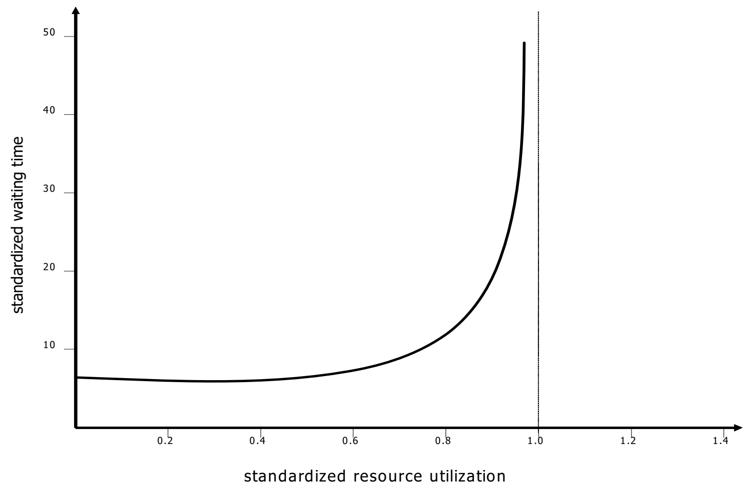

Figure 1: Nonlinear Relationship between Waiting Time and Resource Utilization (Voß & Woodruff, 2003: 162)

It is necessary to examine the problem of lead time dynamics at individual links to better understand the effects and the modelling requirements and complexities at the aggregate planning level for the entire supply network. The aim of this paper is to demonstrate, based on empirical evidence obtained by a survey and interviews recently executed and briefly sketched in the next section, the need for aggregate planning models with the following features: being able to take into account the nonlinear relation between lead times and workload, while remaining tractable to be adapted to complex production systems and supply chains. The remainder of this paper is organized as follows. In Section 2 we point out the empirical evidence of load dependent lead times by means of the results obtained from interviews and a survey recently executed. Then we survey methods and models dealing with load dependent lead times and examine indirect approaches, aspects of queuing theory, and introduce so-called clearing functions in Section 3. The paper concludes with some remarks and suggestions for future research directions.

2 Load Dependent Lead Times – Empirical Evidence

Production planning is a complex issue especially in the context of variable demand patterns or stochastic demand. In numerous production environments demand quantities are not known at the beginning of the production planning process. As a result it is difficult to create forecasts of (highly) variable demand patterns. Production uncertainties and unforeseen events such as machine breakdowns, unavailability of production resources, illness of workers, etc. raise the instability of the production process with queues building up in front of machines resulting in increased WIP and FGI levels and consequently in raised lead times. Therefore, production processes in such environments tend to become somewhat hectic with overtime in peak situations leading to unbalanced utilization of production resources. This is especially true for the food (or semiconductor) industry which also has to account for various deteriorating rates of their production material, which is another complication issue of (load dependent) lead times. As a result, the forecast quality is very important in tactical (and operational) production planning since it prevents precipitated releases of jobs (orders) in the production process and, therefore, should be linked with aggregate production planning and order release control. Nevertheless, there is a lack of practical and useful tools for tactical production planning which permits companies to account for variable demand and unforeseen events causing load dependent lead times. This is one of the principle outcomes of our survey and interviews with companies from different industrial sectors such as transportation, logistics and inventory, aerospace, industry automation and mineral oil, and the chemical industry.

The study includes enterprises of various sizes producing different types of products with very different product life cycles including base polyols, load cels, indicators/ transmitters, software, IT- and logistic services and satellite launchers. These companies face diverse demand patterns and environmental challenges they have to take into consideration in the overall production planning process, especially with regard to the planning of resources and their utilization levels. Many companies face variable (seasonal) and not easily predictable demand for their foremost products. This seems to be the main uncertainty in the production process, because further potential precarious factors such as, e.g., the cooperations with supply chain partners and delieveries from supply chain partners are not validated as highly impacting the production process. This is due to the fact that, e.g., the launcher business has very long production cycle times (the production of one launcher like Ariane 5 or Vega takes on average 2.5 years). Here, problems concerning the cooperation of supply chain partners are not very critical in terms of time compared to other branches like the automotive industry where JIT production is mainly implemented and late deliveries of components cause the whole production process to stop. Nevertheless, late deliveries of important components for, e.g., a launcher also cause the whole assembly process to stop which leads to great financial losses not only because of idle times of production resources and very expensive WIP waiting in the queue, but also because of the costs for the client associated with the delay (lost profits of satellite services). Other companies do not experience supply chain cooperation problems due to long endurance with few supply chain partners, leading to a stabilized and well defined work flow.

Usually medium or large organisations are characterized by a large number of supply chain partners and more or less complex production processes. The range of surveyed companies and their production systems cover synchronized facilities, job shops and make-to-order-systems with different core objectives in tactical production planning as, e.g., maximizing resource utilization in order to avoid idle times (mostly in synchronized production facilities), minimizing lead times or cycle times (this holds true for make-to-order situations) and minimizing WIP and FGI levels. Only a few companies use specific tools for tactical production planning such as SAP R/3, SAP APO (APO SNP for tactical production planning and APO PP/DS for operational production planning). The survey confirms the prevailing use of mrp-based systems together with estimated lead times (or planned lead times) leading to the problems outlined above. Most of the companies experience rising lead times due to machine breakdowns as well as rising WIP levels and consequent queuing in front of machines. Nevertheless, because of the unavailability of data sets (surveys) which are necessary to execute a detailed empirical analysis, it is not clear whether this occurs before reaching 100% utilization, but despite the lack of information, queuing theory emphasizes the impact of resource utilization on load dependent lead times.

The main goals in tactical production planning of the surveyed companies consist in minimizing lead (or cycle) times, as well as WIP and FGI inventory levels, and maximizing resource utilization in order to avoid idle times. For this purpose some of them use, e.g., some “worst case lead times” in order to have enough buffer time at certain (critical) points in the production system and to secure that demands can be met. Others implement estimates of lead times derived from historical data of the production system, which gets problematic when production processes change. Consequently, the underlying data for estimated or planned lead times is neither reliable nor useful in order to achieve the mentioned objectives. However, companies are aware of the fact that decisions on the workload in the production system (and of single resources), on scheduling and sequencing, and on lot sizing and setup times are key factors influencing (load dependent) lead times. To summarize, they lack models (included in comprehensive, usable and useful software tools) providing them with, e.g., “if-then”-analysis in order to better understand the impact of decisions of resource utilization levels, and not only for one single machine or production resource, but even for the whole supply chain network, and furthermore, in order to permit them to use better estimates of lead times. Until now useful models did not exist which provide production planners with necessary information about the lead times which will be experienced in case of diverse resource utilization levels.

As mentioned above, load dependent lead times are the result of production planning processes and should not be an input factor for production planning and scheduling. Moreover, the surveyed companies state the interest in models which take into account load dependent lead times and their impact on the performance of production. Thus it is necessary to analyse the nonlinear relationship of resource utilization and lead times, as well as influencing factors in more detail in order to integrate them in aggregate production planning. Finally, the integration of supply chain partners in an overall supply chain network tends to precede in the right direction by connecting the participants through information system tools, e.g., the same production planning software or linking them together by add-ons in order to guarantee real time information of their production process and those of the supply chain partners. Nevertheless, this is still an ongoing process.

3 Models Including Load Dependent Lead Times

Load dependent lead times are primarily considered in the framework of capacity planning models and order release control mechanisms. Traditional models aim at "filling time buckets" which represent the available capacity of a production system in discrete time periods, while linear programming models are typically employed using hard capacity constraints which ignore the phenomenon that in asynchronous systems, prevailing in practice, queues build up long before 100% resource utilization is reached. Furthermore, they do not impose costs until the capacity constraint is violated, i.e., the constraint only tightens in case of 100% utilization (Karmarkar, 1989). Moreover, these models neither account for WIP and other lead time related cost factors that increase with queues and delays and accordingly with longer lead times (Karmarkar, 1993; Zipkin, 1986), nor do they include WIP costs and lead time consequences of capacity loading which can have significant effects on the performance of the production system.

3.1 Indirect Approaches

There are several ways to address problems associated with load dependent lead times. Some authors do not directly consider the difficulty of modeling nonlinear dependencies of lead times and workload, but try to solve the problem indirectly by influencing parameters that have an effect on lead times such as decisions on job release policies, influences of the demand side, changes in production plans or by smoothing demand variability, e.g., by implementing a make to stock policy, or shifts (away) from bottlenecks in order to increase capacity. Other approaches concentrate on lot sizing as an influencing factor or on production system characteristics as well as employing queuing theory as an analytical method.

3.2 Aspects of Queuing Theory

Analysis of production system performance and important key factors like throughput, WIP levels and load dependent lead times are frequently executed in the context of queuing theory due to the fact that a large percentage of lead times are waiting times. It has been shown that 90% of the total flow time is due to transit times, where 85% consists of waiting (queuing) time, 3% of quality control, and 2% of transportation time; only 10% is due to value added processing operations (Tatsiopoulos & Kingsman, 1983). Queuing network models highlight the relationship between the capacity, loading and production mix as well as the resulting WIP levels and effect on lead times (Karmarkar, 1987) and provide important information on the causes of congestion phenomena. Furthermore, they show that delays predominantly depend on the service variability, i.e., the processing time of a resource, the variability of the arrival rate of work at a resource and the current workload as well as scale effects with major delays near the maximum capacity usage (Srinivasan et al., 1988).

Congestion phenomena are inherent problems of production systems complicating the planning process. They emerge at different and frequently changing times and places which are hardly predictable. Therefore, it is crucial to better understand the reasons for congestion phenomena like the limited capacity of a machine (resource) to respond to demand variation over time (Lautenschläger, 1999) and to account for them in aggregate planning models. The literature on queuing and congestion phenomena is multitudinous; see, e.g. (Chen et al., 1988; Karmarkar, 1987, 1989; Spearman, 1991; Suri & Sanders, 1993; Zipkin, 1986). Spearman (1991) develops a cyclic closed queuing network model with three parameters, viz. the bottleneck capacity, the “raw processing time” (i.e., increasing failure rate processing time, IFR) and a congestion coefficient which specifies a unique throughput/WIP curve in order to analyze the dependency between mean cycle time (synonymously used for “flow time” in many references) and WIP for the whole production system, i.e., single resources and their processing times are not considered. The model indicates a relationship between mean cycle time and WIP level and can be used to predict the average cycle time in exponential as well as in IFR closed queuing networks. Chen et al. (1988) provide a network queuing model for semiconductor wafer facilities which points out that congestion and delays are due to variability in the operating environment. So this variability has to be smoothed in order to obtain shorter production cycle times.

It is useful to start with a queuing model in order to obtain some approximations for the key parameters or objective functions to be implemented in an aggregate planning model (Buzacott & Shantikumar, 1993).

3.3 Indirect Integration of Load Dependent Lead Times

There are only a few approaches which try to integrate load dependent lead times directly into mathematical programming models. For instance, Zijm & Buitenhek (1996) developed a manufacturing planning and control framework for a machine shop which includes workload oriented lead time estimates. For this purpose they suggest a method that determines the earliest possible completion time of arriving jobs with the restriction that the delivery performance of any other job in the system will not be adversly affected, i.e., that every job can be completed and delivered on time. Their aim is to determine reliable planned lead times based on workload which guarantee that due dates are met and can be implemented at a capacity planning level, serving as input for a final detailed capacity scheduling procedure that also takes into account additional resources, job batching decisions as well as machine setup characteristics. Their framework is partly based on the work of (Karmarkar, 1987; Karmarkar et al., 1985). Missbauer (1998) focuses on the hierarchical production planning concept in which all partial problems can be included (e.g., aggregate production planning, capacity planning, lot sizing, scheduling etc.), but which avoids the problems of a comprehensive model, e.g., problems of data procurement, limited computational storage space, CPU times that are too long for calculation etc.

Graves (1986) studied the dependencies between production capability, variability (uncertainty) of the production requirements, and level of WIP inventory in a tactical planning model and analyzes to which extent the job flow time (or WIP inventory) depends on the utilization of each resource of a job shop or production stage. He further concentrates on analyzing the interrelationship of flow time and production mix. For this purpose he employs a network model where multiple routings of jobs are possible so that the lack of a dominant work flow renders production control, which aims at reducing the variance of planned lead times. In addition, he uses a queuing model that includes flexible production rates of resources which can be set by a tactical planning model in order to smooth the work flow and to avoid the overload of resources. Moreover, he implements a control rule at each resource that determines the amount of work performed during a time period which is a fixed portion α_j of the queue of work at j remaining at the start of the period at a specific resource j: $P_{jt} = \alpha_j Q_{jt} \quad \forall j, t$ with P_{jt} denoting the production of resource j in time period t, α_j a smoothing parameter with $0 < \alpha_j \leq 1$ and Q_{jt} the queue of work at j at the beginning of time period t. This parameter α_j is called "proportional factor" by Missbauer (1998) and "clearing factor" by Graves (1986), because it indicates the quantity of jobs (orders) which can be cleared or finished (and passed to another station) in one time period. Here, the clearing factor implies infinite capacity since the resource is able to complete the fixed portion α_j even when the workload (WIP) is infinitely high. The major drawback of this model is the employment of a linear function and consequently the omission of the nonlinear relationship of WIP and lead times. Nevertheless, Graves (1986) seems to be the first reference accounting for the dependency between lead times and workload and giving a practical aid on how to set planned lead times in, e.g., mrp models considering the workload of the production system.

3.4 Clearing Functions

Taking up the idea of Graves and integrating it in a model with so-called “clearing factors” *α(WIP)* which are nonlinear functions of the WIP yields a clearing function of the following form (Karmarkar, 1989; Srinivasan et al., 1988):

$$Capacity = \alpha(WIP) * WIP = f(WIP)$$

where *f* represents the clearing function which models capacity as a function of the workload. The clearing factor specifies the fraction of the actual WIP which can be completed, i.e., “cleared,” by a resource in a given period of time (Asmundsson et al., 2003). Missbauer (1998) calls this function “utilization function.”

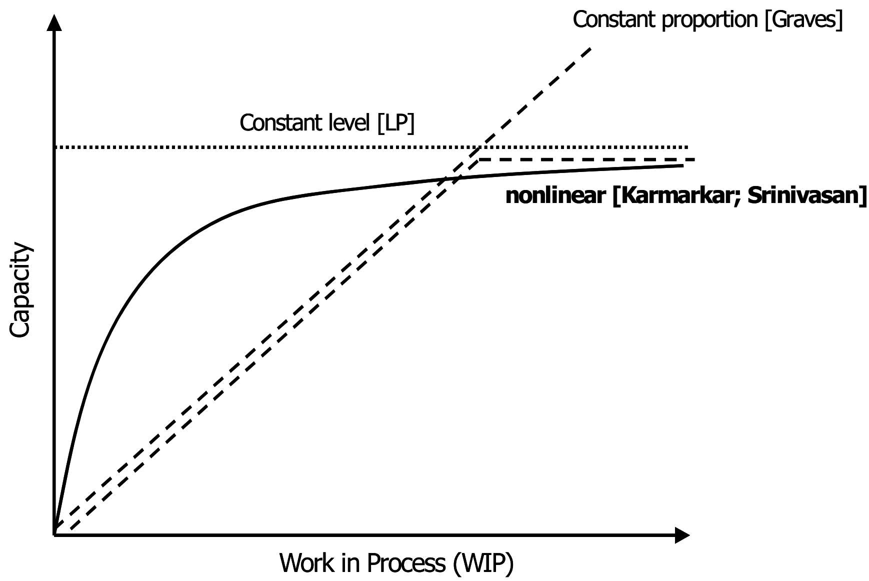

Figure 2: Different Clearing Functions (Karmarkar, 1993)

Figure 2 depicts some possible clearing functions where the constant level clearing function corresponds to an upper bound for capacity as mainly employed by linear programming models. This implies instantaneous production without lead time constraints since production takes place independently of WIP in the production system. The constant proportion clearing function represents the control rule given by Graves (1986) which implies infinite capacity and hence allows for unlimited output. In contrast to the nonlinear clearing function of Karmarkar and Srinivasan et al., the combined clearing function in some region underestimates and in others overestimates capacity. Moreover, the nonlinear clearing function relates WIP levels to output and lead times to WIP levels which are influenced by the workload of the production system (Karmarkar, 1993) and, therefore, is able to capture the behavior of load dependent lead times. By applying Little's Law, the clearing function can be reinterpreted in terms of lead time or WIP turn. Additionally, the slope of the clearing function represents the inventory turn with lead times given by the inverse of the slope (Karmarkar, 1989). Asmundsson et al. (2003) combine aspects of queuing theory with the clearing function concept by employing a clearing function of the above given form and by defining the performance of a resource (work center) as dependent on the workload using a G/G/1 queuing model.

In order to develop the clearing function, two approaches can be found in the literature to date. The first is the analytical derivation from queuing network models and the second an empirical approximation using a functional form which can be fitted to empirical data. Because of the large amount of detail in practical systems the complete identification of the clearing function will not be possible, so we have to work with approximations. Asmundsson et al. (2002) integrate the estimated clearing function in a mathematical programming model where the framework is based on the production model of (Hackman & Leachman, 1989) with an objective function that minimizes the overall costs. It is assumed that backorders do not occur and that all demand must be met on time. We concentrate on this model as an example for the direct integration of load dependent lead times in aggregate production planning models. The model is then stated as follows:

$$Min \sum_t \left[\sum_n \left[\sum_i (\phi_t^n X_{it}^n + \varpi_{it}^n W_{it}^n + \pi_{it}^n I_{it}^n + \rho_{it}^n R_{it}^n) \right] + \sum_j \theta_{jt} Y_{jt} \right]$$

subject to :

$$W_{it}^n = W_{i,t-1}^n - \frac{1}{2}\left(X_{it}^n + X_{i,t-1}^n\right) + R_{it}^n + \sum_{j \in A(n,i)} Y_{jt} \qquad \forall n,t,i$$

$$I_{it}^n = I_{i,t-1}^n + \frac{1}{2}\left(X_{it}^n + X_{i,t-1}^n\right) - D_{it}^n - \sum_{j \in B(n,i)} Y_{jt} \qquad \forall n,t,i$$

$$X_{\bullet t}^n \le f_{nt}\left(W_{\bullet t}^n\right) \qquad \forall n,t$$

$$X_{it}^n, W_{it}^n, Y_{jt}, I_{it}^n, R_{it}^n \ge 0 \qquad \forall n,t,i,j$$

where $\phi_t^n X_{it}^n$ denotes the costs of the total amount of production over the latter half of period t and the first half of period t+1, represented by X_{it}^n, with ϕ_t^n referring to the corresponding unit costs at node n in period t. The WIP costs and the FGI costs of item i at node n at the end of period t are denoted by $\omega_{it}^n W_{it}^n$ and $\pi_{it}^n I_{it}^n$, respectively, with ω_{it}^n and π_{it}^n being the corresponding unit costs and W_{it}^n and I_{it}^n representing the WIP and the inventory, respectively. Likewise, the costs of releases of raw material of item i at node n during period t are represented by $\rho_{it}^{n} R_{it}^{n}$ with unit costs ρ_{it}^{n} and finally the transfer costs on arc *j* during period *t* are given by $\theta_{jt} Y_{jt}$ with corresponding unit costs θ_{jt}. $X_{\bullet t}^{n}$ is the production quantity and $W_{\bullet t}^{n}$ the WIP level summarized over all items *i*.

The first two constraints denote the flow conservation for WIP and FGI, which is different from classical models since inventory levels at each node in the network are connected with the throughput rate. (*A(n,i)* / *B(n,i)* represent a set of transportation arcs contributing to inflow / outflow of item *i* at node *n.*) In contrast to Ettl et al. (2000), the nonlinear dynamic is incorporated in the clearing function and thus not included in the objective function, but modeled as a constraint of the model. Furthermore, the planning circularity which is one of the most significant shortcomings of mrp systems is overcome by not modeling the lead time explicitly in the mathematical program. Consequently, there is no need to employ fixed and / or estimated lead times ignoring the nonlinear relationship between lead times and WIP. Instead, they are calculated using Little's Law:

$$L_{it}^{n} = \frac{W_{it}^{n}}{X_{it}^{n}}$$

where L_{it}^{n} denotes the expected lead time for the last job of item *i* which arrived before the end of period *t*. We are also interested in deriving the lead times for single items *i* in order to consider multiple product types with different resource consumption patterns. For that purpose we assume the standard case of FIFO processing for which the following relationship holds:

$$\frac{X_{it}^{n}}{X_{\bullet t}^{n}} = \frac{W_{it}^{n}}{W_{\bullet t}^{n}}$$

Taking this relation and multiplying the production quantities X_{it}^{n} with their so-called resource consumption factor ξ_{it}^{n} which defines the capacity consumption per unit produced for item *i* at node *n*, we derive a new variable Z_{it}^{n} of the following form:

$$Z_{it}^{n} = \frac{\xi_{it}^{n} X_{it}^{n}}{X_{\bullet t}^{n}} = \frac{\xi_{it}^{n} W_{it}^{n}}{W_{\bullet t}^{n}} \qquad \forall n,t,i$$

By implementing this variable Z_{it}^{n} we obtain the clearing function for each item *i* which is called the partitioned clearing function:

$$\xi_{it}^{n} X_{it}^{n} \le Z_{it}^{n} f\left(\frac{\xi_{it}^{n} W_{it}^{n}}{Z_{it}^{n}}\right) \qquad \forall n,t,i$$

with the following properties of Z_{it}^{n}:

$$\sum_i Z_{it}^n = 1 \qquad \forall i,n,t$$

$$Z_{it}^n \geq 0 \qquad \forall i,n,t$$

The partitioned clearing function is depicted in Figure 3.

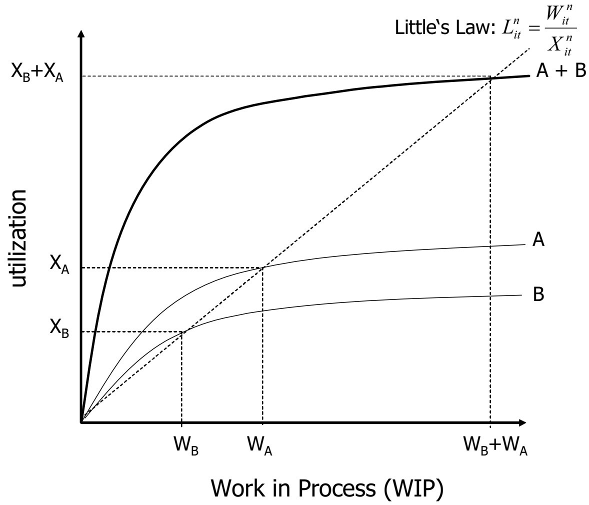

Figure 3: Clearing Function for Products A and B (Asmundsson et al., 2002)

In order to relax the assumed priority rule (FIFO) we only suppose that Z_{it}^n satisfies the properties stated above, but has an arbitrary functional form. With this formulation Asmundsson et al. (2002) succeed in integrating the nonlinear relationship between WIP and lead times in a mathematical model. The second goal is to transform this model in a tractable form which allows even the relatively large planning problems to be dealt with. For this reason we use a linear programming formulation by representing the partitioned clearing function through a set of linear constraints. To be more precise, the clearing function is approximated by the convex hull of straight lines which is possible because of its concavity:

$$f_{nt}\left(W_{\bullet t}^n\right) = \min_c \left\{\alpha_{nt}^c W_{\bullet t}^n + \beta_{nt}^c\right\} \qquad \forall n,t$$

The individual lines of the items are denoted by the index *c*. The β coefficients represent the intersection with the y-axis and indicate the capacity splitting (sharing) across the items while the α coefficients represent the slope of the clearing function. Applying this formulation to the partitioned clearing function leads to the following form:

$$Z_{it}^{n}\left(\frac{\xi_{it}^{n}W_{it}^{n}}{Z_{it}^{n}}\right)=Z_{it}^{n}\cdot\min_{c}\left\{\alpha_{nt}^{c}\frac{\xi_{it}^{n}W_{it}^{n}}{Z_{it}^{n}}+\beta_{nt}^{c}\right\}=\min_{c}\left\{\alpha_{nt}^{c}\xi_{it}^{n}W_{it}^{n}+\beta_{nt}^{c}Z_{it}^{n}\right\}$$

Replacing the former capacity constraint of the original nonlinear mathematical programming model with nonlinear lead time and capacity dynamics gives the complete linearized formulation:

$$Min\sum_{t}\left[\sum_{n}\left[\sum_{i}(\phi_{it}^{n}X_{it}^{n}+\varpi_{it}^{n}W_{it}^{n}+\pi_{it}^{n}I_{it}^{n}+\rho_{it}^{n}R_{it}^{n})\right]+\sum_{j}\theta_{jt}Y_{jt}\right]$$

subject to:

$$W_{it}^{n}=W_{i,t-1}^{\mathrm{n}}-\frac{1}{2}\left(X_{it}^{n}+X_{i,t-1}^{n}\right)+R_{it}^{n}+\sum_{j\in A(n,i)}Y_{jt}\qquad\forall n,t,i$$

$$I_{it}^{n}=I_{i,t-1}^{n}+\frac{1}{2}\left(X_{it}^{n}+X_{i,t-1}^{n}\right)-D_{it}^{n}-\sum_{j\in B(n,i)}Y_{jt}\qquad\forall n,t,i$$

$$\xi_{it}^{n}X_{it}^{n}\le\alpha_{nt}^{c}\xi_{it}^{n}W_{it}^{n}+Z_{it}^{n}\beta_{nt}^{c}\qquad\forall n,t,i,c$$

$$\sum_{i}Z_{it}^{n}=1\qquad\forall n,t$$

$$X_{it}^{n},W_{it}^{n},Y_{jt},I_{it}^{n},R_{it}^{n},Z_{it}^{n}\ge 0\qquad\forall n,t,i,j$$

The approximation of the partitioned clearing function is depicted in Figure 4.

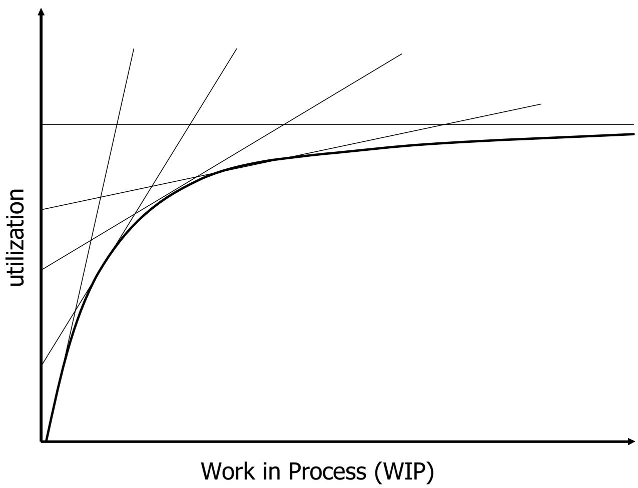

Figure 4: Linearization of the Partitioned Clearing Function

Advantages of this approach lie in the fact that the marginal cost of capacity and the marginal benefit of adding WIP are strictly positive, because of the fact that the constraints are always active as opposed to classical models where, e.g., the capacity constraint is only active at 100% utilization. However, this is only likely to be the fact at the bottleneck of the production system (Asmundsson et al., 2002). In order to examine the relevance and performance of this approach the authors consider an example of a simple single stage system with three products, which gives very good results. Furthermore, the sensitivity of the estimated clearing function to diverse shop floor scheduling algorithms, different demand patterns and techniques of production planning using a simulation model is analyzed. To summarize, the clearing function model reflects the characteristics and capabilities of the production system better than models using fixed planned lead times (like mrp) and derives realistic and robust plans with better on time delivery performance, lower WIP and system inventory (Asmundsson et al., 2003). In addition, the model captures the effects of congestion phenomena on lead times and WIP and, therefore, the fundamental trade-off between anticipatory production to account for possible demand peaks and just in time production to avoid higher costs due to preventable FGI. Finally, the releases generated by the partial clearing function model are smoother and lead to enhanced lead time performance. Moreover, interactions between clearing functions and shop floor execution systems such as the dependency of load dependent lead times on the various priority rules have to be analyzed more closely. This is a circularity, because clearing functions are dependent on the employed scheduling policy and, therefore, on the result of the scheduling algorithm. Moreover, the schedules are dependent on the release schedule and consequently on the planning algorithm.

4 Conclusions

We have seen that considerations on load dependent lead times are rare in the literature to date which is also true for aggregate planning and control models. This is particularly noteworthy, because reflections on lead times are essential with respect to the global competitiveness of firms. Furthermore, we have seen the importance to account for the nonlinear relationship between lead times and workload of production systems and further influencing factors such as product mix, scheduling policies, batching or lot sizing, variable demand patterns, deterioration etc. Analytical (queuing) models emphasize the nonlinear relationship between lead times and workload which is included only in a few mathematical planning models. Additionally, there is a lack of models which analyze load dependent lead times in the context of stochastic demand and uncertainties evidently prevailing in practice. The approach of modeling clearing functions in order to account for load dependent lead times as outlined in this paper is considered very promising and will be implemented in a stochastic framework by using queuing models with the purpose of integrating the problem of variable demand patterns, and in order to analyze the behavior of load dependent lead times. This will be used as a starting point for more sophisticated modelling of production systems where we try to model single production units (resources, workstations, etc.) as queuing models in order to derive their specific clearing functions. Furthermore, it has been stated that load dependent lead times mainly arise due to congestion phenomena which are pervasive problems of production systems, complicating the planning process by emerging at different and frequently changing times and places which are hardly predictable due to various factors like machine breakdowns, variable demand patterns or deteriorating items. For future work we shall develop approaches for aggregate production planning which take empirical values of the probability of machine breakdowns into account as well as the other mentioned causes of congestion phenomena. This can be achieved by applying, e.g., a learning algorithm which allows for learning the behavior of production units (resources or workstations) as well as the overall system behavior, and including this information into aggregate production planning. Information on downtimes is rarely considered or integrated in mathematical models. It is not even considered in the latest and sophisticated supply chain management software like SAP APO. This will be a subject for further research.

5 References

Asmundsson, J., Rardin, R. L., Uzsoy, R. (2002): Tractable Nonlinear Capacity Models for Aggregate Production Planning, Working Paper, School of Industrial Engineering, Purdue University, West Lafayette.

Asmundsson, J., Rardin, R. L., Uzsoy, R. (2003): An Experimental Comparison of Linear Programming Models for Production Planning Utilizing Fixed Lead Time & Clearing Functions, Working Paper, School of Industrial Engineering, Purdue University, West Lafayette.

Buzacott, J. A., Shantikumar, J. G. (1993): Stochastic Models of Manufacturing Systems, Englewood Cliffs, New York.

Caramanis, M. C., Ahn, O. M. (1999): Dynamic Lead Time Modeling for JIT Production Planning, Proceedings of the IEEE International Conference on Robotics and Automation, Detroit, MI, May 10-15, v2, 1450-1455.

Chen, H., Harrison, M. J., Mandelbaum, A., van Ackere, A., Wein, L. M. (1988): Empirical Evaluation of a Queuing Network Model for Semiconductor Wafer Fabrication, Operations Research, 36(2): 202-215.

Ettl, M., Feigin, G. E., Lin, G. Y., Yao, D. D. (2000): A Supply Network Model with Base-Stock Control and Service Requirements, Operations Research, 48(2): 216-232.

Graves, S. (1986): A Tactical Planning Model for Job Shops, Operations Research, 34(4): 522 – 533.

Hackman, S. T., Leachman, R. C. (1989): A General Framework for Modeling Production, Management Science, 35(4): 478-495.

Karmarkar, U. S. (1987): Lot Sizes, Lead Times and In-Process Inventories, Management Science, 33(3): 409-418.

Karmarkar, U. S. (1989): Capacity Loading and Release Planning with Work-In-Process (WIP) and Lead Times, Journal of Manufacturing and Operations Management, 2(2): 105-123.

Karmarkar, U. S. (1993): Manufacturing Lead Times, Order Release and Capacity Loading, in: Graves, S., Rinnooy Kan, A., Zipkin, P. (eds.): Logistics of Production and Inventory, Handbooks in Operations Research and Management Science, Vol. 4, Amsterdam: p. 287-329.

Karmarkar, U. S., Kekre, S., Kekre, S. (1985): Lotsizing in Multi-Item Multi Machine Job Shops, IIE Transactions, 17(3): 290-297.

Lautenschläger, M. (1999): Mittelfristige Produktionsprogrammplanung mit auslastungsabhängigen Vorlaufzeiten ("Tactical Production Planning with Workload Dependent Forward Production Times"), PhD Thesis, TU Darmstadt.

Missbauer, H. (1998): Bestandsregelung als Basis für eine Neugestaltung von PPS-Systemen ("Inventory Control as a Basis for a New Concept for PPS-Systems"), Physica, Heidelberg.

Spearman, M.L. (1991): An Analytic Congestion Model for Closed Production Systems with IFR Processing Times, Management Science, 37(8): 1015-1029.

Srinivasan, A., Carey, M., Morton, T. E. (1990): Resource Pricing and Aggregate Scheduling in Manufacturing Systems, Working Paper, GSIA, 1988 (Revised December 1990).

Suri, R., Sanders, J.L. (1993): Performance Evaluation of Production Networks, in: Graves, S., Rinnooy Kan, A., Zipkin, P. (eds.): Logistics of Production and Inventory, Handbooks in Operations Research and Management Science, Vol. 4, Amsterdam: p. 199-286.

Tatsiopoulos, I.P., Kingsman, B.G. (1983): Lead Time Management, European Journal of Operational Research, 14(4): 351-358.

Voß, S., Woodruff D.L. (2003): An Introduction to Computational Optimization Models for Production Planning in a Supply Chain, Springer, Berlin.

Zäpfel, G., Missbauer, H. (1993): New Concepts for Production Planning and Control, European Journal of Operational Research, 67(3): 297-320.

Zijm, W. H. M., Buitenhek, R. (1996): Capacity Planning and Lead Time Management, International Journal of Production Economics, 46-47: 165-179.

Zipkin, P. H. (1986): Models for Design and Control of Stochastic, Multi-Item Batch Production Systems, Operations Research, 34(1): 91-104.