International Journal of Production Research

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tprs20

The strong effect of flexible capacities on achievable WIP levels and throughput times: a simple model and its implications

Hermann Lödding, Constantin Steffens & Michael Winter

INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH
https://doi.org/10.1080/00207543.2024.2399732

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

The strong effect of flexible capacities on achievable WIP levels and throughput times: a simple model and its implications

Hermann Lödding, Constantin Steffens and Michael Winter

Institute of Production Management and Technology, Hamburg University of Technology, Hamburg, Germany

ABSTRACT
Setting WIP levels at the workstations of a production is an important task: WIP levels affect throughput times as well as utilisation rates. Consequently, they also impact the delivery times to customers and the productivity. Queueing theory provides profound knowledge regarding the factors for determining suitable WIP levels. However, the existing models have been neglecting the important influence of capacity flexibility. This paper suggests an extension of the well-known Kingman equation to model the impact of flexible capacities on the required WIP levels. Simulation experiments have shown that the model is able to accurately forecast the utilisation level of a workstation for different factor levels of capacity flexibility, WIP and load variance. Companies can use the model to set consistent target levels for these factors.

ARTICLE HISTORY
Received 21 November 2023
Accepted 25 July 2024

KEYWORDS
Logistic operating curves; capacity flexibility; queueing theory; throughput times; WIP levels

1. Introduction

One of the most challenging, but also one of the most important decisions companies need to make in Production Planning and Control (PPC) is determining the WIP (work in progress) levels and the throughput times for the workstations: If they are set too high, the resulting long delivery times might lead to a rejection of customers. Conversely, setting the WIP levels and throughput times too low, leads to interruptions in the material flow and thus to costly losses in capacity utilisation and productivity.

Literature provides established models and methods for dimensioning WIP levels and throughput times.

Queueing theory, invented by Danish mathematician Erlang, provides a broad basis of models that quantify the impact of utilisation and interarrival times on the WIP levels and service times (Erlang 1917). Queueing theory is widely applied in different areas such as computing, telecommunication, and transport (Giambene 2021; Cascetta 2009). Moreover, it provides fundamental insights that are useful for production management. A widely known example of this is the well-known Kingman equation describing the influence of load variance on the performance of production (Kingman 1961; Hopp and Spearman 2008).

However, the standard models do not contain variables that consider load or capacity flexibility and implicitly assume rigid capacities and loads (Gross et al. 2008). This is problematic for the setting of throughput times and WIP levels, as one of the core tasks of PPC is to balance load and capacity (Schönsleben 2023; Hopp and Spearman 2008). Nyhuis and Wiendahl (2009) therefore point out that queueing theory overestimates WIP levels and throughput times.

Unlike queueing theory, the *Hanover Funnel Model* has its origins in production management (Bechte 1984) and is in many ways very similar to queueing theory, as it also provides fundamental models on the achievable WIP levels and throughput times. A very widely known model in this context, are the logistic operating curves suggested by Nyhuis and Wiendahl (2009) which depict utilisation and throughput time as a function of the WIP level. The model is based on a deductive-empirical approach combining analytical models with the results of simulation studies.

A fundamental result of the model is that only a fraction of the WIP levels predicted by queueing theory is required with flexible loads and capacities. However, the logistic operating curves suggested by Nyhuis and Wiendahl (2009) also do not provide a quantified relationship between capacity flexibility and required WIP levels.

In summary, (i) there is a gap between the standard models from queueing theory and the logistic operating curves suggested by Nyhuis and Wiendahl (2009), (ii) there is no model quantifying the impact of capacity flexibility on the achievable WIP levels.

CONTACT Constantin Steffens constantin.steffens@tuhh.de Institute of Production Management and Technology, Hamburg University of Technology, Denickestraße 17, 21073 Hamburg, Germany

The aim is to connect the findings from queueing theory and from the funnel model as well as quantifying the impact of capacity flexibility by extending the well-known Kingman equation. Thus, the resulting research question is:

> How can the impact of capacity flexibility on the achievable WIP levels, throughput times and utilisation rates be modelled?

The paper is structured as follows: in Section 2, we will present the theoretical background of our research. The focus will primarily be on the Kingman equation and the logistic operating curves according to Nyhuis and Wiendahl (2009). Section 3 introduces the research method, which is based on a deductive-experimental approach and a four-step procedure. The results are presented in Section 4 by starting with a basic extension of the Kingman equation and including a simulation study to verify the model accuracy. Section 5 discusses the research results by comprising the theoretical and practical implications of the study. Moreover, the limitations are explained and future research is derived from these. Finally, Section 6 summarises the paper's outcomes.

2. Theoretical background

The section on the theoretical background starts with a brief explanation of capacity flexibility (2.1). Afterwards, we present the queueing theory and the Kingman equation (2.2). Lastly, we explain an alternative model for logistic operating curves from Nyhuis and Wiendahl (2009) (2.3).

2.1. Capacity flexibility

Capacity flexibility is the ability to change (i.e. increase or reduce) the capacity of a workstation or a production area. In more detail, it can be described (1) by the extent of the capacity change and the resulting maximum respectively minimum capacity, (2) by the required time to change the capacity (e.g. 3 days for a voluntary shift), (3) by the costs for changing the capacity (Wiendahl and Breithaupt 1998; Lödding 2012; Thürer and Stevenson 2020).

Typical measures for realising a capacity increase are overtime, extra shifts, the short-term hiring of temporal workers and cross-qualified workers. Similar measures exist for capacity reduction (Schönsleben 2023; Lödding 2013).

Capacity flexibility can and should be applied both in capacity planning and in capacity control. Responding to changes in demand (volume, product mix) by using capacity flexibility is well established. Similarly, reacting to production disturbances and accumulating backlogs is a standard in manufacturing. As we will see, the impact of flexible capacities on achievable WIP levels and throughput times has gained much less attention so far (Lödding 2013; Schönsleben 2023).

2.2. Queueing theory and the Kingman equation

Queueing theory models the WIP levels, throughput times, and utilisation rates of queueing systems. It originates at the beginning of the last century with Agner Krarup Erlang's studies of the Copenhagen telephone network (Giambene 2021; Cascetta 2009). There is comprehensive literature on queueing theory, presenting models and fundamental laws in detail, e.g. the impact of different service time or interarrival time distributions on the required WIP levels. Its field of application is much larger than PPC and comprises applications from computer theory to transport system modelling and the management of hospitals (Giambene 2021; Cascetta 2009).

Within the field of manufacturing systems, current research includes aspects such as digital twins (Seok, Cai and Park 2021) and a data-based modelling for the prediction of WIP levels and cycle time distributions (Deenen et al. 2024). Furthermore, machine learning and its use with queueing theory offers research opportunities (Khayyati and Tan 2022). The widely known Kingman equation is particularly suitable for PPC (Kingman 1961). Unlike other approaches from queuing theory, it makes no assumptions about the specific distribution shape of the interarrival time and the work content. It rather estimates the relationship between the mean values of WIP and utilisation solely based on the variances of the interarrival time and the work content. In addition, it is particularly accurate for high utilisation rates (Kingman 1961; Hopp and Spearman 2008).

In the notation of Nyhuis and Wiendahl (2009), the mean WIP required for a defined utilisation level is calculated as follows (see Figure 1 for an illustration):

$$WIPO_m = U_m + \frac{U_m^2}{1 - U_m} \cdot \left(\frac{WC_v^2 + TIA_v^2}{2} \right) \tag{1}$$

$WIPO_m$ mean WIP in number of orders [−]
U_m mean utilisation [−]
WC_v coefficient of variance of the work content [−]
TIA_v coefficient of variance of the interarrival time [−]

Generally, it can be distinguished between the currently processed WIP (active WIP) and the WIP in the queue (buffer WIP) (Nyhuis and Wiendahl 2009). The active WIP is described by the first summand of Equation

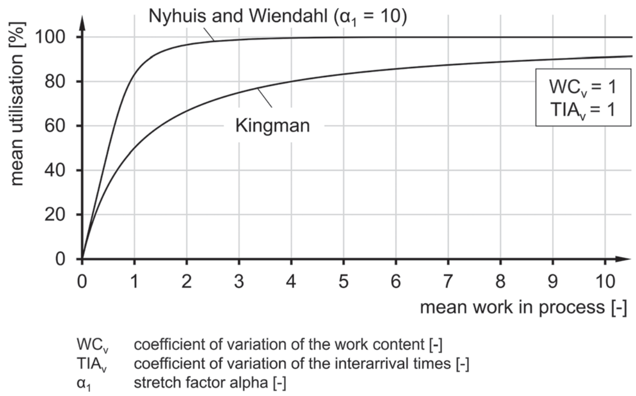

Figure 1. Comparison of logistic operating curves.

1. It increases proportionally to the utilisation. The second summand describes the required average buffer WIP in the queue that increases with (1) higher utilisation and (2) greater coefficients of variance of work contents and interarrival times. In the following, the second summand is denoted as the load variance coefficient. It is sometimes referred to as variability term, e.g. (Hopp and Spearman 2008):

$$LD_{var} = \frac{WC_v^2 + TIA_v^2}{2} \tag{2}$$

LD_{var} load variance coefficient [−]
U_m mean utilisation [−]
WC_v coefficient of variance of the work content [−]
TIA_v coefficient of variance of the interarrival time [−]

An important limitation of the Kingman equation is that it only applies to a single workstation, but not to parallel workstations. However, queueing theory provides a generalisation of the equation for workstations with multiple machines (Hopp and Spearman 2008; Whitt 1993).

In addition, the industry not only needs information on utilisation and inventory, but also on output rates and throughput times. However, this can easily be achieved by using the following relationships. Multiplying the utilisation by the available capacity leads to the output rate of the workstation in h/day (Nyhuis and Wiendahl 2009)[1] that can be converted to orders per day by a division with the mean work content.

$$ROUT_m = U_m \cdot CAPA_m \tag{3}$$

$$ROUTO_m = ROUT_m / WC_m \tag{4}$$

$ROUT_m$ mean output rate [h/day]
$ROUTO_m$ mean output rate in number of orders per day
[−/day] U_m mean utilisation [−]
$CAPA_m$ mean capacity of a workstation [h/day]
WC_m mean work content [h]

A variant of Little's Law (Little 1961) or the funnel formula (Bechte 1984) can be used to calculate the average throughput time:

$$TTP_m = WIPO_m / ROUTO_m \tag{5}$$

TTP_m mean throughput time [day]
$WIPO_m$ mean WIP in numbers of orders [−]
$ROUTO_m$ mean output rate [−/day]

2.3. An alternative model: the logistic operating curves from Nyhuis / Wiendahl

One of the important findings of the Hanover Funnel Model is that queueing theory overestimates the required WIP levels and throughput times due to its implicit assumption of rigid capacities and loads (Nyhuis and Wiendahl 2009).

In his dissertation, Nyhuis (1991) derives an approximation equation for logistic operating curves for production in a deductive-experimental procedure (Nyhuis 1991; Nyhuis and Wiendahl 2009). It is based on extensive simulation experiments with flexible capacities and a load-oriented order release to smooth load variations (Bechte 1984; Yan et al. 2016; Breithaupt, Land, and Nyhuis 2002). Under these conditions, the workstations in the production already reach a defined utilisation level with much lower WIP than with rigid capacities and unsmoothed loads. Figure 1 illustrates how big the differences are.

The example shows that achieving a utilisation of 95 % only requires less than 10 % of the buffer WIP calculated by the Kingman equation. The Kingman equation is therefore hardly suitable for setting WIP levels and throughput times for workstations in production with capacity flexibility.

However, the approximation equation by Nyhuis and Wiendahl (2009) also has its weaknesses: Nyhuis and Wiendahl (2009) explicitly point out the importance of an empirically determined stretch factor α_1 on the logistic operating curves. By setting a standard value of 10 for the alpha stretch factor α_1, the logistic operating curve cannot represent different characteristics of the load and capacity flexibility (Nyhuis and Wiendahl 2009). Nyhuis and Wiendahl (2009) consider its influence to be so elementary that they elevated it to the rank of a basic law of logistics: 'The size of the WIP buffer required to ensure the utilisation of the workstation is mainly determined by the flexibility of the load and capacity' (Nyhuis and Wiendahl 2009). This is confirmed by extensive simulation experiments by Busse (2013). Moreover, systematic simulation experiments by Land et al. (2015) and by Falu and Duffie (2014) support the effectiveness of capacity adjustments.

With respect to load flexibility, we refer to Busse (2013) who examines the impact of load flexibility on the alpha stretch factor α_1. Winter, Luttkau, and Lödding (2021) propose an extension of the Kingman equation for a simple load balancing algorithm that shifts orders between neighbouring periods to smoothen load.

This paper focusses on how to model the impact of capacity flexibility. We explain our modelling approach in Section 3.

3. Research methodology

Section 3 is divided into two subsections: First, we present the four phases of our research approach (Section 3.1). Then, we describe the simulation model which we used to examine the modelling quality (Section 3.2).

3.1. The research process

Our research process followed four phases, from early decisions and premises (phase 1) to evaluation of the approximation equation (phase 4). In each phase, we applied different research methods.

Phase 1: Early decisions and premises – An early and important decision in the research process was to use the well-known Kingman equation as a starting point, mainly because of its simplicity. To extend the WIP buffer related term of the Kingman equation by a factor (CF) is also obvious, because it would be (1) possible to isolate the effect of capacity flexibility on the required WIP buffer, (2) easy to calculate CF from simulation experiments by relating the buffer WIP to the theoretical value suggested of the Kingman equation and (3) easy to interpret the effect of capacity flexibility. The result of phase 1 is the extended Kingman equation, which is presented in Section 4.1.

Moreover, we decided to begin modelling with a single workstation, which is simple and comprehensible and we used only a single method for adjusting capacities.

Phase 2: Understanding the effect of capacity flexibility on the required WIP buffers – During this phase of the research process, we tried to improve our understanding of the underlying principles. Important questions were: (1) Which are the main influencing factors determining the effect of capacity flexibility on the required WIP levels? (2) Which of the influencing factors are especially important and should therefore be mapped by the approximation equation? (3) Which of the influencing factors are less relevant and may be neglected for the approximation? (4) How can the relevant factors be quantified as a basis for the modelling base?

Already in this phase of the research process, we used simulation experiments to test our hypotheses regarding the relevance of some factors. The main results of this can be found in Section 4.2.

Phase 3: Specifying the approximation equation for the capacity flexibility factor CF – The goal of phase 3 was to derive the approximation equation for the capacity flexibility factor CF and consequently also for the logistic operating curves. The approach is depicted in Figure 2:

The basic idea is to compare the results of an approximation equation with those from simulation experiments for the same setting of input data. To ensure that the approximation equation is not only valid for one but a wide range of WIP levels, we compared simulation experiments and approximation equation for different WIP levels. If there was a significant deviation, we analysed the reasons for it and modified the equation or its parameters. To support the process of parameter setting, we applied Matlab functionalities on curve fittings.

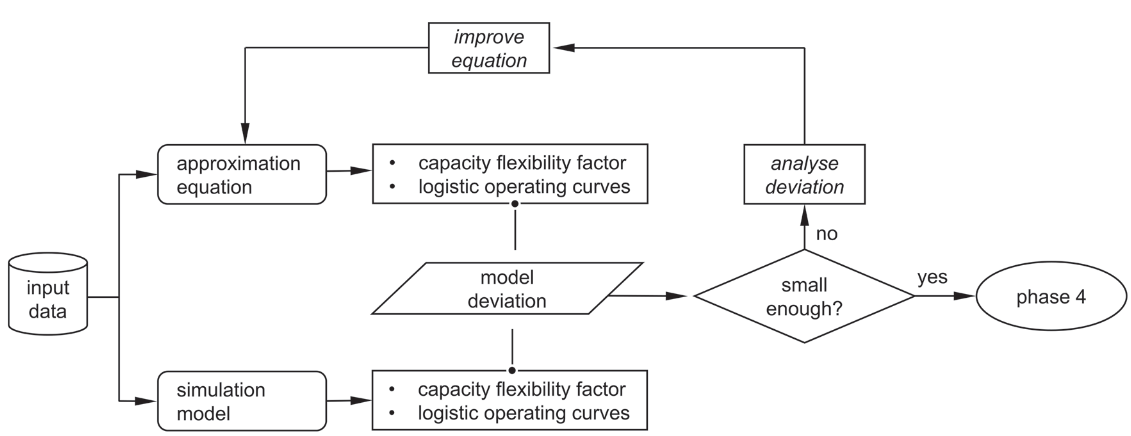

Figure 2. Specifying approximation equation for the capacity flexibility factor.

This phase involved many iterations before we found a sufficient approximation for the capacity flexibility factor (CF), including an iteration back to phase 2 to find out that the effect of the capacity flexibility relates to the load variance. The approximation equation for CF can be found in Section 4.3.

Phase 4: Evaluation of the approximation equation with a full-factorial design of experiments – The final phase of the research process was to verify the approximation equation for the capacity flexibility factor CF with a full-factorial design of experiments and to document the model's accuracy. The results can be found in Section 4.5. We also conducted some experiments in which we varied the period length as a noise factor. The simulation model and the implemented capacity control method is explained in detail in the following section.

3.2. The simulation model and the capacity control logic

The simulation model for a single workstation was created with Tecnomatix Plant Simulation. Each simulation run comprised more than 23,000 orders including a ramp-up and ramp-down phase. 20,000 orders were investigated to calculate one operating point. We used 76 operating points and thus 76 simulation experiments for each logistic operating curve. WIP levels were measured at random times during the investigation period and averaged to calculate the mean WIP level of a simulation experiment. To evaluate different WIP levels and utilisation rules, we varied the interarrival times and the target WIP. Changing the target WIP level resulted in some WIP levels not being realised in the simulation runs. This led to gaps in the simulated logistic operating curves.

For capacity control, we use the method by Petermann (1995), which is described in his dissertation on applying control theory to manufacturing. The method determines the required capacity in a way that a defined target WIP level can be reached at the end of the period ((Petermann 1995) cf. Figure 3).

The required capacity of a period is calculated as follows:

$$CAPA_{req}(P) = \frac{IN(P_{end}) - WIP_{target} - OUT(P_{start})}{P_{end} - P_{start}} \quad (6)$$

$CAPA_{req}$ (P) capacity requirements of a period [h/day]
WIP_{target} target WIP [h]
IN input [h] OUT output [h]
P_{start} start of period [day]
P_{end} end of period [day]

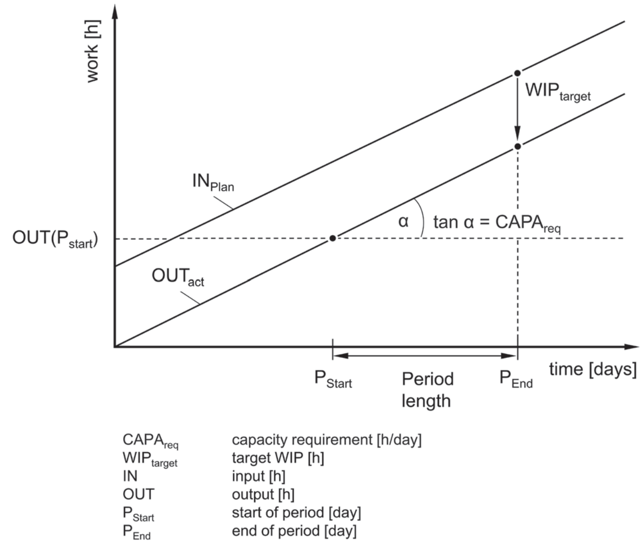

Figure 3. Principle of capacity planning according to Petermann (Petermann 1995).

If the required capacity is between the minimum and maximum capacity, capacity is set to the required capacity. Otherwise, the capacity is set to its maximum value ($CAPA_{required}$ > maximum capacity) or to its minimum value ($CAPA_{required}$ < minimum capacity).

This way of setting capacities positively influences the required WIP levels for a defined utilisation rate in two ways: (1) An increase in capacity reduces WIP build-up during periods with loads greater than standard capacity. This results in a lower average WIP level. (2) A capacity reduction in time periods with a low load leads to lower utilisation losses.

We simulated capacity flexibility by a change in intensity. Therefore, the following applies:

$$WC_{sim} = WC \cdot \frac{CAPA_{standard}}{CAPA_{actual}} \quad (7)$$

WC_{sim} work content in simulation [h]
WC (standard) work content [h]
$CAPA_{standard}$ standard capacity [h/day]
$CAPA_{actual}$ actual capacity [h/day]

With an actual capacity of 1.1 times the standard capacity, a job with a standard time of one hour is completed in under 55 min. Thus, the simulation model calculates at the beginning of each period the required capacity by applying Equation 6, and then sets the actual capacity to either the required capacity if this is possible or otherwise to the maximum or minimum capacity as described above. For each order, the completion time is documented and for each period the utilisation rate is recorded by dividing the period output by the possible period output at the capacity level of the period.

4. Results

Section 4 is divided into five subsections: First, we present the extended Kingman equation (4.1). Then, we describe the effect of capacity flexibility on the required WIP level (4.2). Afterwards, we explain the approximation equation (4.3). Since models simplify the reality, we discuss the neglected influencing variables (4.4). Lastly, we evaluate the approximation equation by conducting simulation experiments with a full factorial experimental design (4.5).

4.1. The extended Kingman equation

The basic idea of our approach is to extend the Kingman equation for rigid capacities by a factor CF. This factor is intended to represent the influence of capacity flexibility on the required buffer WIP (Winter, Luttkau, and Lödding 2021):

$$WIPO_m = U_m + \frac{U_m^2}{1-U_m} \cdot \left(\frac{WC_v^2 + TIA_v^2}{2}\right) \cdot CF \quad (8)$$

$WIPO_m$ mean WIP in number of orders [−]
U_m mean utilisation [−]
WC_v coefficient of variance of the work content [−]
TIA_v coefficient of variance of the interarrival time [−]
CF capacity flexibility factor ∈ [0;1] [−]

This seems appropriate for the following reasons: (1) with CF = 1, the approximate equation corresponds to the well-established Kingman equation for rigid capacities. (2) With CF = 0, the approximate equation corresponds to the ideal curve of Nyhuis and Wiendahl (2009). (3) The reasonable range of values for the empirical factor lies between the values of 0 and 1 and is thus clearly defined. (4) The factor CF can be easily interpreted: A value of CF = 0.5 (or x) implies that the workstation requires half (or x times) the buffer WIP of a comparable workstation with rigid capacities. (5) The equation can be solved as a quadratic equation by mean WIP. The required mean WIP for a desired utilisation is therefore directly calculable.

To simplify the notation, we define an effective load variance coefficient:

$$LD_{var,eff} = \left(\frac{WC_v^2 + TIA_v^2}{2}\right) \cdot CF = LD_{var} \cdot CF \quad (9)$$

$LD_{var,eff}$ effective load variance coefficient [−]
WC_v coefficient of variance of the work content [−]
TIA_v coefficient of variance of the interarrival time [−]
CF capacity flexibility factor [−]
LD_{var} load variance coefficient [−]

Since we will model CF as a function of the mean WIP, we solve Equation 8 for the utilisation. For an effective load variance < 1:

$$U_m = -\frac{WIPO_m + 1}{2\cdot(LD_{var,eff} - 1)} - \sqrt{\left(\frac{WIPO_m + 1}{2\cdot(LD_{var,eff}-1)}\right)^2 + \frac{WIPO_m}{(LD_{var,eff}-1)}} \quad (10)$$

U_m mean utilisation [–]
$WIPO_m$ mean WIP in number of orders [–]
$LD_{var,eff}$ effective coefficient of variance of the load [–]

Note that Equation 10 is not defined for $LD_{var,eff} = 1$. In this case, Equation 10 is equal to the well-known M/M/1 case with exponentially distributed interarrival times and work contents and $U_m = WIPO_m$ / $(1 + WIPO_m)$. For $LD_{var} > 1$ the square root term is added to (and not subtracted from) the first term. This reflects the alternative solution of the quadratic equation (cf. Appendix A for a complete notation of Equation 10).

4.2. Understanding the effect of capacity flexibility on the required WIP level

Equation 10 reduces the complexity of creating logistical operating curves for workstations with flexible capacity by using only an approximation of the CF. This approximation should be based on a good understanding of the influencing variables on the required WIP buffer. To come up with an equation, a measurable variable has to be defined for each influencing factor. To make the approximation applicable for a wide scope of production settings, these variables should be normalised. Subsections 4.2.1–4.2.3 describe the influencing factors that we decided to include in the approximation equation.

4.2.1. Capacity flexibility

Capacity flexibility is the obvious influencing factor on the required WIP buffer. To normalise capacity, we relate capacity to the mean load of a period.

$$CAPA_{rel}(P) = \frac{CAPA(P)}{LD_m(P)} \quad (11)$$

Conversely, the relative load of a period is calculated as:

$$LD_{rel}(P) = \frac{LD_m(P)}{CAPA(P)} = \frac{1}{CAPA_{rel}(P)} \quad (12)$$

$CAPA_{rel}$ relative capacity of a workstation [−]
CAPA capacity of a workstation [h/day]
LD_{rel} relative load of a workstation [−]
LD_m mean load [h/day]

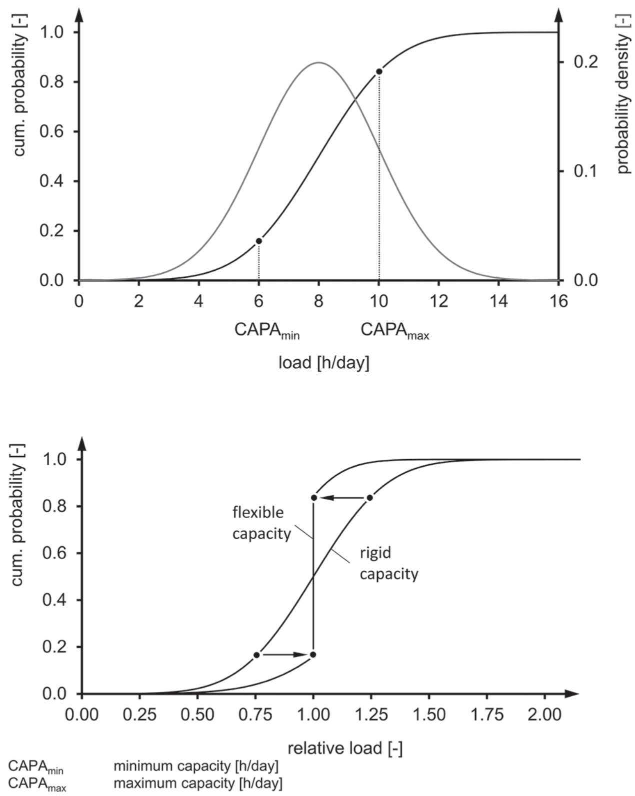

Figure 4. Distribution function of the relative period load with a capacity flexibility.

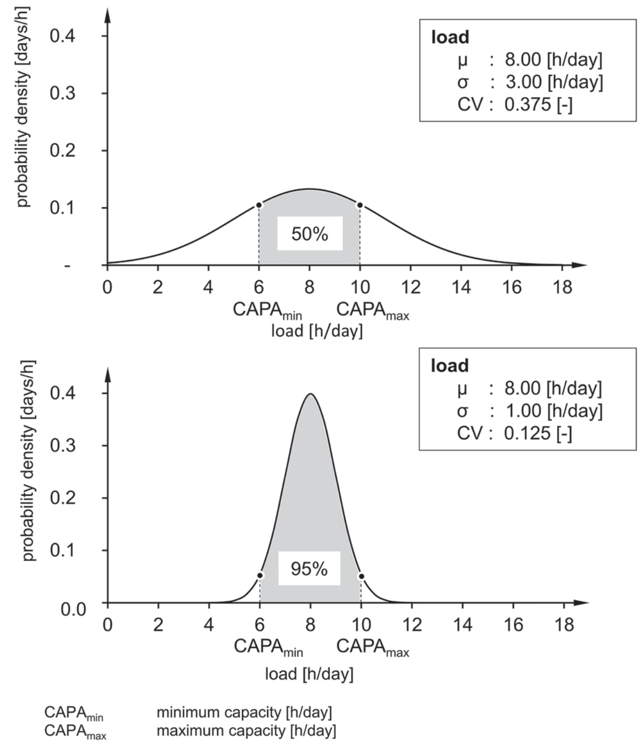

Figure 5. Load variance and capacity flexibility.

As a simplification, we assume that the standard capacity is equivalent to the long-term average of the load across many periods. Consequently, a capacity flexibility of $\pm$ 10 % means that the capacity can be set such that $0.9 \leq CAPA_{rel} \leq 1.1$, where $CAPA_{rel}$ is calculated with the long-term average of the load.

With these definitions, the effect of capacity flexibility on the relative load can be easily depicted. The upper half of Figure 4 shows an example of a normally distributed load distribution with an expected value of 8 h/day and a standard deviation of 2 h/day. The standard capacity in the example corresponds to the expected value of the load, as explained above. The company can freely vary the capacity for the period between a minimum capacity of 6 h/day ($CAPA_{rel} = 0.75$) and a maximum capacity of 10 h/day ($CAPA_{rel} = 1.25$) reflecting a capacity flexibility of $\pm$ 25 %. From the known properties of the normal distribution, about 16 % of the period loads are less than the minimum capacity and 16 % exceed the maximum capacity. In the remaining 68 % of the cases, the period load is between the minimum and maximum capacity.

In the lower part of Figure 4, the horizontal axis represents the relative period load, which is obtained by relating the (absolute) load to the period capacity (Equation 12). Accordingly, its expected value is 1 if the period capacity matches the mean period load. If a company adjusts the capacity to the period load, the relative period load changes, so the relative load can demonstrate the effect of the flexible capacity.

It becomes immediately clear that capacity flexibility can be a very effective means of improving the balance between load and capacity.

4.2.2. Load variance

As we had to learn from the simulation experiments, the impact of capacity flexibility on the required WIP buffer also depends on the load variance: The lower the standard deviation of the load, the greater the probability that a given capacity flexibility is sufficient to match the load.

This is illustrated in Figure 5. Both parts of Figure 5 assume a maximum capacity of 10 hours/day. This results in a relative maximum capacity of 1.25.

In the upper part of Figure 5, the load scatters in a way that the capacity flexibility is sufficient to match 50 % of the period loads. In the lower part of Figure 5, the same capacity flexibility covers a much larger proportion of the period loads (95 %) and can therefore compensate for a larger proportion of the load fluctuations.

To map this relationship, we normalise the maximum relative capacity by the square root of the load variance coefficient LD_{var} as an estimate for the coefficient of

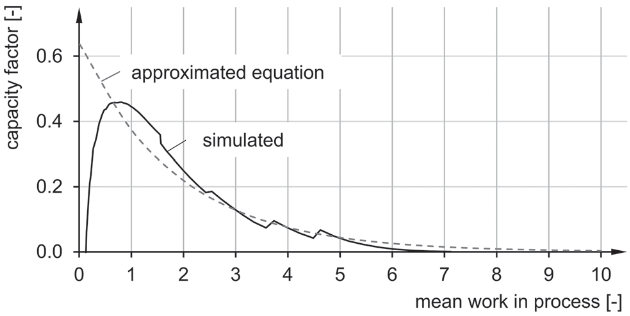

Figure 6. Simulated and approximated values for the capacity flexibility factor CF.

variance (CV) of the load:

$$CAPA_{max,rel,norm} = 1 + \frac{CAPA_{max,rel} - 1}{\sqrt{LD_{var}}} \qquad (13)$$

$CAPA_{max,rel,norm}$ normalised relative capacity at maximum capacity [−]

$CAPA_{max,rel}$ relative capacity at maximum capacity [−]

LD_{var} load variance coefficient [−]

In the upper part of Figure 5, the normalised relative maximum capacity corresponds to 1.67, i.e. the maximum capacity can be equal to the mean + 0.67 standard deviations of the relative load. In the lower part of Figure 5, the maximum capacity corresponds to the mean + 2 standard deviations (normalised relative maximum capacity = 3).

As a result, the maximum capacity is normalised twice: First by reference to the average load, then by reference to the standard deviation of the load.

4.2.3. Utilisation and WIP

According to the standard Kingman model, the required buffer WIP increases disproportionately with the utilisation of the workstation, so that the buffer WIP accounts for the majority of the total WIP at high utilisation levels. We were initially intending for a WIP-independent capacity flexibility factor CF. However, this hypothesis falsified, as shown in Figure 6 which is displaying CF over WIP for a capacity flexibility of ± 25 %.

The CF values decrease with increasing WIP for WIP levels greater than 0.8 orders. In the simulation experiments, CF is decreasing for very low WIP levels. However, this is less relevant for the modelling, because (1) at very low WIP levels the buffer WIP is small compared to the active WIP and (2) utilisation rates are low and therefore less relevant for practice.

4.3. The approximation equation

We approximate the simulated CF curve by a hyperbola of the form $CF = 1/x^n$ as illustrated in Figure 6. x is a function of the normalised relative maximum capacity and n is a function of the mean WIP. The parameters were set as follows:

$$CF = \frac{1}{(1 + (CAPA_{max,rel,norm} - 1) \cdot 0.6)^{2+2.4 \cdot WIPO_m}} \qquad (14)$$

CF capacity flexibility factor [−]

$CAPA_{max,rel,norm}$ normalised relative maximum capacity [−]

$WIPO_m$ mean WIP in number of orders [−]

The equation calculates CF as a function of the WIP. The following procedure can be used to determine logistic operating curves:

- Calculation of the normalised relative maximum capacity (Equation 13)
- Calculation of the CF values for different WIP levels (Equation 14)
- Calculation of the utilisation as a function of the WIP and CF (Equation 10)

For an example of a maximum capacity of 10, the relative maximum capacity is $CAPA_{max,rel} = \frac{10h/SCD}{8h/SCD} = 1.25$. For a coefficient of variance of the load of 0.6, Equation 13 yields a normalised relative maximum capacity of $CAPA_{max,rel,norm} = 1 + \frac{1.25-1}{0.6} \approx 1.417$.
Accordingly, for a mean WIP level of 3 orders, $CF \approx 0.128$. This means that in the example slightly less than 13 % of the Kingman buffer WIP is required to achieve a certain capacity utilisation.

Figure 7 depicts the resulting logistic operating curves with a capacity flexibility of 25 % and a load variance coefficient of 1.0. Comparing the curves shows the high impact of flexible capacities on the achievable WIP and utilisation levels.

4.4. Neglected influencing variables

Models simplify reality. For model application, it is important to know which simplifications can have a particular influence on the accuracy of the results.

(1) *Period length:* The period length influences the dispersion of period loads. In practice, companies usually set capacity for defined periods, e.g. for a day, a week or even a month. The longer the period, the more likely load fluctuations will average each other out. For very long periods, the period loads therefore

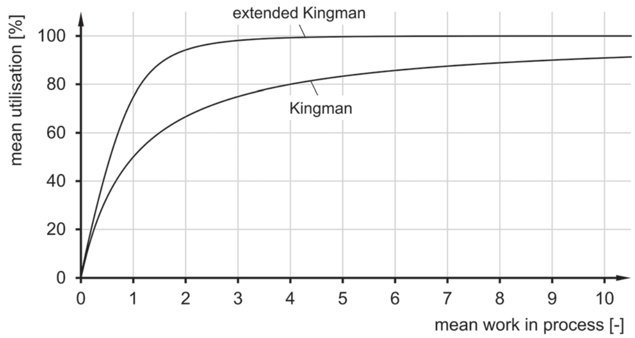

Figure 7. Comparison of the extended and the standard Kingman equation for a capacity flexibility of +/− 25 % and a load variance coefficient of 1.0.

converge towards the expected value of the load, so that capacity flexibility no longer has any influence (cf. also the following influencing variable).

(2) *Excess capacity:* Capacity flexibility greater than the maximum load remains unused and is therefore is not able to further reduce the required buffer WIP. E.g. in the lower part of Figure 5 it is obviously irrelevant whether the maximum capacity flexibility is 12 h/day or 14 h/day or 16 h/day. Period length and excess capacity are interrelated influencing variables, as load fluctuations average out over longer periods, such that a maximum capacity which is important for short periods may become obsolete for longer periods.

(3) *Method of capacity control:* We used a WIP-regulating capacity control in our simulation experiments to derive the operating curve (see Section 3.2). Other capacity control methods exist and will probably have an impact on CF as well.

4.5. Evaluation by simulation experiments

In order to evaluate the accuracy of the approximation equation for the logistic operation curves, we conducted simulation experiments. Therefore, we describe the Design of Experiments (Subsection 4.5.1). We then present the simulation results (Subsection 4.5.2).

4.5.1. Design of experiments

We created a full-factorial experimental design with three factors: relative maximum capacity, load variance and average WIP. Table 1 shows the factors with their factor levels.

We tried to cover a broad spectrum of values from moderate to ambitious, respectively challenging values for the relative capacity and the load variance. The WIP values reflect a relevant range from both: a modelling and application perspective. In total, this results in a design with 54 different experiments.

Table 1. Factors and factor levels of the simulation study.

Factor	Factor levels		
Relative capacity	1.05	1.15	1.25
Load variance $\sqrt{LD_{var}}$	0.3	0.6	1.0
WIP	1 2	3 4	5 10

Note: LD_{var} Load variance coefficient

4.5.2. Simulation results

The complete table of the simulation results and the comparison with the modelled values is depicted in the Appendix B. Figure 8 shows the results of the simulation experiments by superimposing the capacity utilisation measured in the experiments at the different WIP levels on the model-based capacity utilisation shown as a line graph. For each of the three levels of capacity flexibility three logistic operation curves are displayed, one for each load variance level.

The mean absolute deviation across all 54 tests is 0.6 % with a standard deviation of 1.0 %. Eight out of ten experiments, that had the highest deviation between modelling and simulation experiments, were conducted with a mean WIP level of one order. This indicates that the model accuracy at very low WIP levels is not as good as at higher levels (cf. Appendix B).

The impact of flexible capacities is highlighted in Figure 9 which depicts the logistic operating curves for a load variance coefficient of 1.0.

5. Discussion, limitations, and future research directions

Section 5 is divided into three subsections. Subsection 5.1 discusses the theoretical implications of the results whereas Subsection 5.2 explains the practical implications. Finally, future research directions are derived from the limitations (5.3).

5.1. Theoretical implications

(1) The results confirm the large effect of capacity flexibility on the logistic operating curves. The effect is described by potentially very small CF values. As depicted in Figure 6, CF values can easily take on values below 0.1 respectively 0.05. This means that less than 10 % respectively less than 5 % of the buffer WIP suggested by the Kingman equation is required with flexible capacities which confirms the general findings of Nyhuis and Wiendahl (2009).

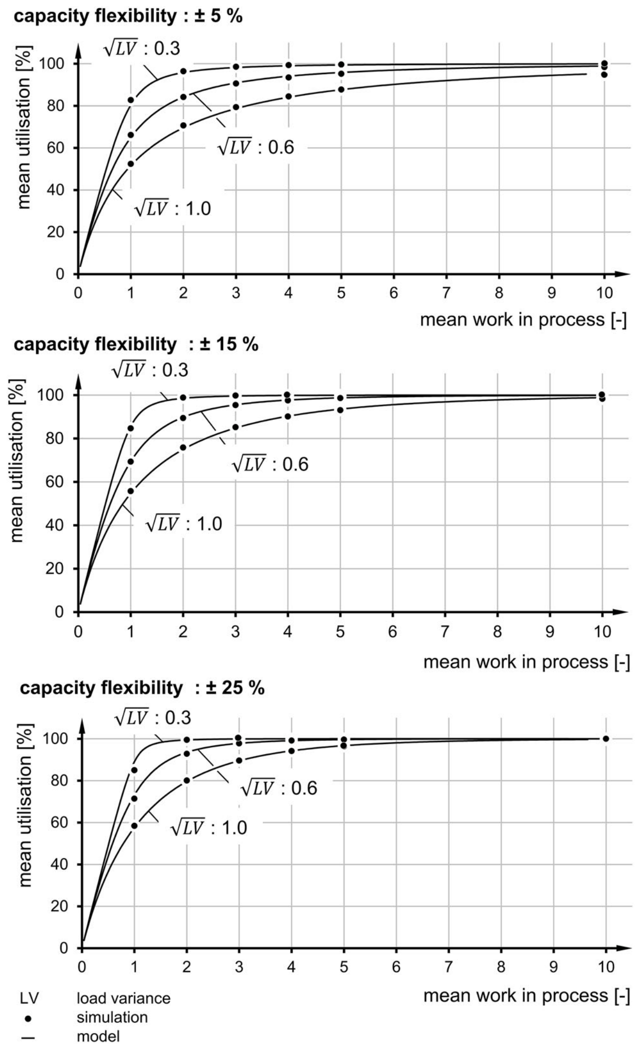

Figure 8. Logistic operating curves for different levels of capacity flexibility and load variance.

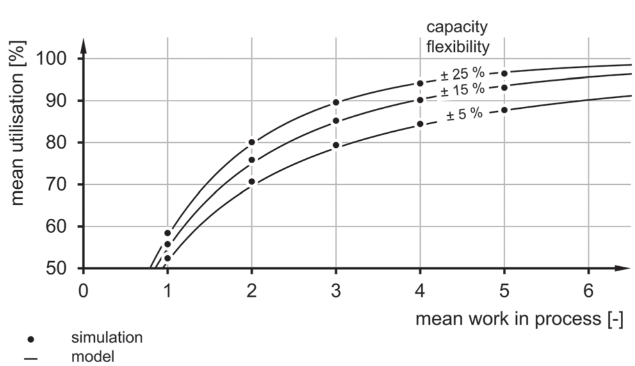

Figure 9. Logistic operating curves for different levels of capacity flexibility for a load variance coefficient of 1.0.

The results also imply that the (non-extended) Kingman equation overestimates the required WIP levels and consequently also the throughput times by far.

(2) The results confirm that the effect on the required WIP level increases with more capacity flexibility which confirms the fifth basic law of production logistics by Nyhuis and Wiendahl (2009).

(3) The results show that the effect on the required WIP level also depends on the load variance, as the relative maximum capacity is normalised by the square root of the load variance coefficient. This implies that reducing the load variance comes with a double advantage: It reduces the load variance coefficient (which is well-known from literature) and the capacity flexibility factor CF.

(4) The use of capacity flexibility is particularly efficient for high WIP levels and utilisation rates, because CF is decreasing with the WIP (Equation 14, Figure 6). This is consistent with the logistic operating curve of Nyhuis and Wiendahl (2009), which assumes full utilisation with a finite WIP.

(5) To the best of our knowledge, no other extension of the Kingman equation exists that is depicting the influence of capacity flexibility on achievable WIP levels and throughput times. At least for PPC, the developed approximation is therefore a major extension of the well-known Kingman equation.

(6) In the simulation experiments, the deviations between the simulated values from the modelled values are small, indicating that the factors capacity flexibility, load variance and WIP level are well mapped by the approximation equation. Actually, the deviations are much smaller than expected. We attribute this to the effect of capacity control reducing the influence of load variance. Moreover, the simulation ensures reproducible conditions in all experiments and therefore protects the results against the impact of disturbance variables (cf. Section 5.3 on limitations).

(7) We analyse the influence of capacity flexibility in the context of PPC. However, flexible capacities also exist in other areas: Many service providers adapt their working hours and often also the number of employees to match the demand. The potential area of application therefore extends far beyond PPC.

5.2. Practical implications

The modelling and simulation experiments show that capacity flexibility is a very effective way to reduce the required WIP and consequently throughput times compared to the rigid capacities modelled for the Kingman equation. For scheduling and capacity planning, this results in the following guidelines.

(1) *Companies should provide and use capacity flexibility to reduce required WIP levels and throughput times.*

The impact of flexible capacities on achievable WIP levels and throughput times is too big to be ignored by industry. It offers the potential to largely reduce WIP levels and throughput times and it gives ample opportunities to reduce delivery times and WIP-related costs and to increase responsiveness. The guideline points to two different directions. The first direction is to provide respectively increase capacity flexibility, e.g. by negotiating flexible working times or by supporting cross-qualification. The second direction is to implement clear instructions on how to use capacity flexibility (when should a capacity change be triggered, by what extent?).

(2) *Companies should reduce load variance.*

As pointed out above, load variance has a double impact on required WIP levels and throughput times, because it reduces the load variance coefficient *and* the capacity flexibility factor CF. Therefore, reducing load variance is an obvious and well-known approach for improving the performance of producing companies, and even more when using flexible capacities.

(3) *Companies should use the approximation model for determining appropriate settings for WIP, throughput times, utilisation, and capacity flexibility.*

The developed model approximates the achievable WIP levels, throughput times, and utilisation levels for arbitrary values of capacity flexibility. Therefore, it cannot only provide a general direction for improving performance. It also enables companies to determine specific and consistent values for the capacity flexibility, the WIP levels and the throughput times, and utilisation.

Companies should use this opportunity to answer questions as:

- What capacity flexibility is required to reduce achievable throughput times to a defined level without causing utilisation losses of more than x %?
- Which throughput times can be realised with the existing level of capacity flexibility and a utilisation of y %?

5.3. Limitations and implications for future research

The simulation experiments suggest a high accuracy of the approximation model. However, this may not obscure the fact that the modelling and the evaluation are still subject to important limitations and restrictions:

(1) *Restriction to single workstations:* In our view, the biggest limitation of the suggested logistic operating curve is its restriction to a single workstation. A similar model for parallel workstations still needs to be accomplished.
(2) *Effect of period length / excess capacities:* Neither period length nor excess capacities are incorporated in our model. The simulation experiments gave us the impression that the effect of the period length is comparatively small as long as the maximum capacity is regularly utilised. Therefore, the impact of these variables might be small for companies without excess capacities. However, there is still a lack of comprehensive analysis determining the specific capacity level at which the period length or excess capacities begin to exert a significant influence.
(3) *Effect of different capacity control methods:* All simulation experiments use the same simplified form of capacity control (fixed capacities in the period). We assume that different capacity control methods will perform differently. As we only simulated one capacity control method, we cannot make any conclusion for the performance of other methods.

From the limitations, important future research directions can be identified.

(1) *Adapting the model to parallel workstation:* In industry, many workstations are set up parallel, consisting of several machines of the same type. Similar to the Kingman equation, queueing theory also provides an approximation for parallel workstations (Whitt 1983; Hopp and Spearman 2008). We are confident that our approach also provides a good starting point also for parallel work stations:

$$WIPO_m = U_m \cdot NM + \frac{U_m^{\sqrt{2(NM+1)}}}{NM \cdot (1-U_m)} \cdot \left(\frac{WC_v^2 + TIA_v^2}{2}\right) \cdot CF \qquad (15)$$

$WIPO_m$ mean WIP in number of orders [−]

U_m mean utilisation [−]

NM number of machines WC_v coefficient of variance of the work content [−]

TIA_v coefficient of variance of the interarrival time [−]

Again, a simulation study could provide general insights on the course of the capacity flexibility factor CF.

(2) *Modelling the impact of long periods / excess capacities:* As stated above, the effect of the period length

is probably comparatively small, as long as the maximum capacity is regularly utilised. However, it is still not understood, from which point onwards a significant effect is realised. A starting point may be a simulation study, in which the period length is increased until the capacity flexibility factor CF approximates one. This implies that the capacity flexibility is (hardly) used, because the load variations (nearly) average out within a period. An analysis of the capacity flexibility and the load variance could provide insights to the dominating influencing factors.

(3) *Investigating the impact of different capacity control methods:* For order release methods, it is not uncommon to assess different order release methods by their impact on the logistic operation curves or by other performance criteria. The same should be done for different capacity control methods to provide a basis for the industry to select a suitable method. Ideally, the approximative model can be calibrated for each of these methods.

6. Conclusion

The most important result of our research is the extended Kingman equation. For the first time, it allows the determination of logistic operating curves for workstations with flexible capacities, which is based on clearly defined input data. Following the standard Kingman equation, the input data includes information about the work contents and the interarrival times of the orders. Additional information about the capacity flexibility is required to calculate the capacity flexibility factor of the extended Kingman equation. The simulative evaluation shows a high accuracy of the model. Consequently, companies should use the extended Kingman equation to consistently set the target levels for capacity flexibility, capacity utilisation, WIP levels and throughput times. Although our focus is on production planning and control, the results of the extended Kingman equation are relevant for all areas with capacity flexibility. The remaining limitations of the model offer opportunities for future research including an extension of the model to parallel workstations.

Note

1. Note that Nyhuis and Wiendahl differentiate between the capacity and the maximum possible output rate.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the German Research Foundation under grant 415040468.

Data availability statement

The data that support the findings of this study are available from the author upon request.

Notes on contributors

Hermann Lödding studied Industrial Engineering and Management at the University of Kaiserslautern. After completing a PhD in Engineering at the University of Hannover, he was working for Robert Bosch GmbH from 2004 to 2009. Since 2009, he has been working as professor at the Institute of Production Management and Technology at the Hamburg University of Technology.

Constantin Steffens studied Industrial Engineering and Management at the University of Applied Sciences Stralsund and at the Otto von Guericke University Magdeburg. From 2021 to 2022, he was working at the Salzgitter Flachstahl GmbH. Since 2022, he has been working as a research associate at the Institute of Production Management and Technology at Hamburg University of Technology, with a focus on production planning and control.

Michael Winter studied Business Administration and Engineering at RWTH Aachen University. Subsequently, he worked as a research associate at Hamburg University of Technology where he received a PhD in Engineering. Currently, he works as ERP Lead Engineer at Philips. His research focus is production planning and control.

References

Bechte, Wolfgang. 1984. *Steuerung der Durchlaufzeit durch belastungsorientierte Auftragsfreigabe bei Werkstattfertigung*. Fortschritt-Berichte der VDI-Zeitschriften. Reihe 02, Betriebstechnik 70. Düsseldorf: VDI Verlag.

Breithaupt, Jan-Wilhelm, Martin Land, and Peter Nyhuis. 2002. "The Workload Control Concept: Theory and Practical Extensions of Load Oriented Order Release." *Production Planning & Control* 13 (7): 625–638. https://doi.org/10.1080/09537280210000026230.

Busse, Tim D. 2013. "Modellbasierte Bewertung der Belastungsstreuung auf das logistische Systemverhalten." Berichte aus dem IFA 2013,6. Garbsen: PZH-Verl. Zugl. Hannover, Univ., Diss., 2013.

Cascetta, Ennio. 2009. *Transportation Systems Analysis 29*. Boston, MA: Springer US.

Deenen, P. C., J. Middelhuis, A. Akcay, and I. J. B. F Adan. 2024. "Data-Driven Aggregate Modeling of a Semiconductor

Wafer Fab to Predict WIP Levels and Cycle Time Distributions." *Flexible Services and Manufacturing Journal* 36 (2): 567–596. https://doi.org/10.1007/s10696-023-09501-1.

Erlang, Agner K. 1917. "Solution of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges", which contains his classic formulae for call loss and waiting time. https://web.archive.org/web/20110719122546/http://oldwww.com.dtu.dk/teletraffic/erlangbook/pps138-155.pdf.

Falu, I., and N. Duffie. 2014. "Adaptive Due Date Deviation Regulation Using Capacity and Order Release Time Adjustment." *Procedia CIRP* 17:398–403. https://doi.org/10.1016/j.procir.2014.01.073.

Giambene, Giovanni. 2021. *Queuing Theory and Telecommunications*. Cham: Springer International Publishing.

Gross, Donald, John F. Shortle, James M. Thompson, and Carl M. Harris. 2008. *Fundamentals of Queueing Theory*. 4th ed. Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley. http://www.loc.gov/catdir/enhancements/fy0826/2008003734-d.html.

Hopp, Wallace J., and Mark L. Spearman. 2008. *Factory Physics*. 3rd ed. The McGraw-Hill/Irwin series. Boston: McGraw-Hill/Irwin.

Khayyati, Siamak, and Barış Tan. 2022. "A Machine Learning Approach for Implementing Data-Driven Production Control Policies." *International Journal of Production Research* 60 (10): 3107–3128.

Kingman, J. F. C. 1961. "The Single Server Queue in Heavy Traffic." *Mathematical Proceedings of the Cambridge Philosophical Society* 57 (4): 902–904. https://doi.org/10.1017/S0305004100036094.

Land, Martin J., Mark Stevenson, Matthias Thürer, and Gerard J. Gaalman. 2015. "Job Shop Control: In Search of the Key to Delivery Improvements." *International Journal of Production Economics* 168:257–266. https://doi.org/10.1016/j.ijpe.2015.07.007.

Little, John D. C. 1961. "A Proof for the Queuing Formula: L = λ W." *Operations Research* 9 (3): 383–387. https://doi.org/10.1287/opre.9.3.383.

Lödding, H. 2012. "A Manufacturing Control Model." *International Journal of Production Research* 50 (22): 6311–6328. https://doi.org/10.1080/00207543.2011.631605.

Lödding, Hermann. 2013. *Handbook of Manufacturing Control: Fundamentals, Description, Configuration*. Heidelberg: Springer.

Nyhuis, Peter. 1991. "Durchlauforientierte Losgrößenbestimmung." Als Ms. gedr. IFA Nr. 225. Düsseldorf: VDI-Verl. Zugl. Hannover, Univ., Diss.

Nyhuis, P., and H. P. Wiendahl. 2009. *Fundamentals of Production Logistics*. Berlin, Heidelberg: Springer.

Petermann, D. 1995. "Modellbasierte Produktionsregelung." Dissertation. Universität Hannover.

Schönsleben, Paul. 2023. *Handbook Integral Logistics Management*. Berlin, Heidelberg: Springer.

Seok, M. G., W. Cai, and D. Park. 2021. "Hierarchical Aggregation/Disaggregation for Adaptive Abstraction-Level Conversion in Digital Twin-Based Smart Semiconductor Manufacturing." *IEEE Access* 9: 71145–71158. https://doi.org/10.1109/ACCESS.2021.3073618.

Thürer, Matthias, and Mark Stevenson. 2020. "The Use of Finite Loading to Guide Short-term Capacity Adjustments in Make-to-order Job Shops: An Assessment by Simulation." *International Journal of Production Research* 58 (12): 3554–3569. https://doi.org/10.1080/00207543.2019.1630771.

Whitt, W. 1983. "The Queueing Network Analyzer." *Bell System Technical Journal* 62 (9): 2779–2815. https://doi.org/10.1002/j.1538-7305.1983.tb03204.x.

Whitt, W. 1993. "Approximations for the GI/G/m Queue." *Production and Operations Management* 2 (2): 114–161. https://doi.org/10.1111/j.1937-5956.1993.tb00094.x.

Wiendahl, Hans-Peter, and Jan-Wilhelm Breithaupt. 1998. "Automatic Production Control." In *Beyond Manufacturing Resource Planning (MRP II)*, edited by Andreas Drexl, and Alf Kimms, 335–356. Berlin, Heidelberg: Springer.

Winter, Michael, Alexander Luttkau, and Hermann Lödding. 2021. "Produktionskennlinien mit Kapazitätsflexibilität" [Logistic Operating Curves with Capacity Flexibility." *wt* 111 (4): 190–94. https://doi.org/10.37544/1436-4980-2021-04-12.

Yan, Haoyun, Mark Stevenson, Linda C. Hendry, and Martin J. Land. 2016. "Load-Oriented Order Release (LOOR) revisited: bringing it back to the state of the art." *Production Planning & Control* 27 (13): 1078–1091. https://doi.org/10.1080/09537287.2016.1183831.

Appendices

Appendix A

Appendix A shows Equation 10 in full detail:

$$U_m = \begin{cases} -\dfrac{WIPO_m + 1}{2 \cdot (LD_{var,eff} - 1)} - \sqrt{\left(\frac{WIPO_m+1}{2\cdot(LD_{var,eff}-1)}\right)^2 + \frac{WIPO_m}{(LD_{var,eff}-1)}} & \text{for } LD_{var,eff} < 1 \\ \dfrac{WIPO_m}{WIPO_m + 1} & \text{for } LD_{var,eff} = 1 \\ -\dfrac{WIPO_m + 1}{2 \cdot (LD_{var,eff} - 1)} + \sqrt{\left(\frac{WIPO_m+1}{2\cdot(LD_{var,eff}-1)}\right)^2 + \frac{WIPO_m}{(LD_{var,eff}-1)}} & \text{for } LD_{var,eff} > 1 \end{cases}$$

U_m mean utilisation [-]

$WIPO_m$ mean WIP in number of orders [-]

$LD_{var,eff}$ effective coefficient of variance of the load [-]

Appendix B

Appendix B shows the results of the simulation experiments 1–54 for a relative capacity at maximum capacity of 1.05.

Factor levels				Utilisation		
Nr.	$CAPA_{max,rel}$	$WC_v = TIA_v$	mean WIP [−]	Simulation [%]	Modelling [%]	Relative deviation [%]
1	$CAPA_{max,rel} = 1.05$	$WC_v = TIA_v = 0.3$	1	82.76	80.43	−2.81
2			2	96.45	95.85	−0.62
3			3	98.57	98.21	−0.37
4			4	99.35	99.03	−0.32
5			5	99.62	99.42	−0.21
6			10	99.96	99.92	−0.05
7		$WC_v = TIA_v = 0.6$	1	66.16	64.98	−1.78
8			2	84.16	84.19	0.03
9			3	90.61	90.92	0.34
10			4	93.49	94.09	0.64
11			5	95.25	95.87	0.65
12			10	98.42	98.90	0.49
13		$WC_v = TIA_v = 1.0$	1	52.39	51.63	−1.46
14			2	70.66	69.61	−1.49
15			3	79.37	78.69	−0.87
16			4	84.44	84.11	−0.39
17			5	87.76	87.67	−0.10
18			10	94.77	95.34	0.61
19	$CAPA_{max,rel} = 1.15$	$WC_v = TIA_v = 0.3$	1	84.72	85.58	1.02
20			2	98.85	98.55	−0.30
21			3	99.81	99.60	−0.21
22			4	99.96	99.86	−0.10
23			5	99.98	99.94	−0.04
24			10	100.00	100.00	0.00
25		$WC_v = TIA_v = 0.6$	1	69.38	69.39	0.02
26			2	89.44	89.81	0.42
27			3	95.43	95.56	0.13
28			4	97.58	97.75	0.17
29			5	98.65	98.76	0.12
30			10	99.87	99.89	0.02
31		$WC_v = TIA_v = 1.0$	1	55.75	54.73	−1.84
32			2	75.85	74.98	−1.16
33			3	85.22	84.85	−0.43
34			4	90.15	90.31	0.17
35			5	93.10	93.56	0.49
36			10	98.34	98.85	0.51
37	$CAPA_{max,rel} = 1.25$	$WC_v = TIA_v = 0.3$	1	85.02	89.05	4.74
38			2	99.49	99.44	−0.06
39			3	99.98	99.89	−0.09
40			4	100.00	99.97	−0.03
41			5	100.00	99.99	−0.01
42			10	100.00	100.00	0.00
43		$WC_v = TIA_v = 0.6$	1	71.39	73.14	2.44
44			2	92.78	93.52	0.79
45			3	97.79	97.81	0.02
46			4	99.13	99.12	−0.02
47			5	99.62	99.61	−0.02
48			10	99.99	99.99	0.00
49		$WC_v = TIA_v = 1.0$	1	58.42	57.63	−1.35
50			2	80.08	79.63	−0.55
51			3	89.60	89.49	−0.12
52			4	94.13	94.26	0.13
53			5	96.40	96.72	0.33
54			10	99.55	99.71	0.16
					Ø	0.58

$CAPA_{max,rel}$ relative capacity at maximum capacity [−]
WC_v coefficient of variance of the work content [−]
TIA_v coefficient of variance of the inter arrival times [−]
Ø arithmetic mean [%]