



HAL
open science

Backlog Oriented Bottleneck Management – Practical Guide for Production Managers

Roman Ungern-Sternberg, Christian Fries, Hans-Hermann Wiendahl

► **To cite this version:**

Roman Ungern-Sternberg, Christian Fries, Hans-Hermann Wiendahl. Backlog Oriented Bottleneck Management – Practical Guide for Production Managers. IFIP International Conference on Advances in Production Management Systems (APMS), Aug 2020, Novi Sad, Serbia. pp.581-589, 10.1007/978-3-030-57997-5_67 . hal-03635609

HAL Id: hal-03635609

<https://inria.hal.science/hal-03635609v1>

Submitted on 8 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Backlog Oriented Bottleneck Management – Practical Guide for Production Managers

Roman Ungern-Sternberg ^[0000-0003-4750-9296], Christian Fries ^[0000-0001-9607-8897]
Hans-Hermann Wiendahl ^[0000-0002-5307-3444]

¹ Fraunhofer Institute for Manufacturing Engineering and Automation IPA, 70569 Stuttgart,
Germany
roman.ungern-sternberg@ipa.fraunhofer.de

Abstract. Today’s production systems become more and more complex, comprising a multitude of different resources interacting with each other. The Theory of Constraints (TOC) describes the dilemma that the overall output of such a system is constrained by one or more resources called “bottlenecks”. Identifying those bottlenecks has been subject of many research papers. Classical approaches focus typically solely on the bottleneck utilization and neglect accumulated deviations against the planned utilization. This so-called backlog is critical as it has a direct effect on the on-time-delivery. This paper therefore provides a practical approach that can be used by the operative management in order to identify, prioritize and manage bottlenecks in a backlog situation efficiently and effectively.

Keywords: Bottleneck, Management Guide, Production logistics

1 Introduction

A bottleneck is a resource whose capacity requirement exceeds the available capacity in a period [2, 3]. If all production orders have to pass through the bottleneck, the maximum production output corresponds to that of the bottleneck [6: p. 115], resulting to the fact that an hour lost at a bottleneck is an hour lost for the overall system [3]. Classical approaches focus solely on the bottleneck utilization and do rarely provide guidelines for their management. Furthermore, accumulated deviations, which are represented in backlog, are typically neglected. Backlog however, is critical as it has a direct effect on the on-time-delivery. This paper therefore develops a practical two-step approach for analyzing and improving operations. The intended users are production managers in need of a systematic but practical approach to select critical bottlenecks in manufacturing to improve on-time-delivery.

2 Literature Review

This section gives an overview of existing identification methods and defines the terms bottleneck and backlog.

2.1 Bottlenecks

Definition. Within this paper, the capacity bottleneck is defined as the resource with the highest load factor (load divided by maximum capacity) during a defined period. A selection of methods for detecting and managing bottlenecks is summarized in the following section.

Management. Adams uses a graph based iteration process to detect bottlenecks based on throughput time and lateness. The Shifting Bottleneck method (SB) solves bottlenecks from iteration to iteration for all machines. In each iteration step, the machine with the largest maximum delay is selected. Then the bottleneck is solved on this machine and an optimal processing sequence of the orders is determined [1].

Based on the duration a machine is active without interruption Roser develops the Active Period Method [8]. Bottlenecks are detected based on the idea that the longer a machine is working without interruption, the more likely it is constraining the overall systems performance [9].

H.-P. Wiendahl describes the Bottleneck Oriented Logistic Analysis (BOLA) as a method for identifying capacitive bottlenecks in the material flow [16, 17]. This method allows to derive problem specific measures based upon analyzing the logistics behavior using the concept of the logistics operating curves [6]. The BOLA is thereby a permanent control's task and thus a continuous process for improving logistics [5: p. 170].

The existing approaches consider backlog only indirect by observing lateness as a relevant parameter. As backlog is a critical parameter in production control, it will be directly considered in the presented method.

2.2 Backlog in Order Based Production Systems

Definition. The backlog corresponds to the amount of work that a resource should have completed by a certain date but has not yet completed. Thus, the backlog is calculated as the difference between the planned and actual output [7: p. 65f, 14: p.148, 18: p.80].

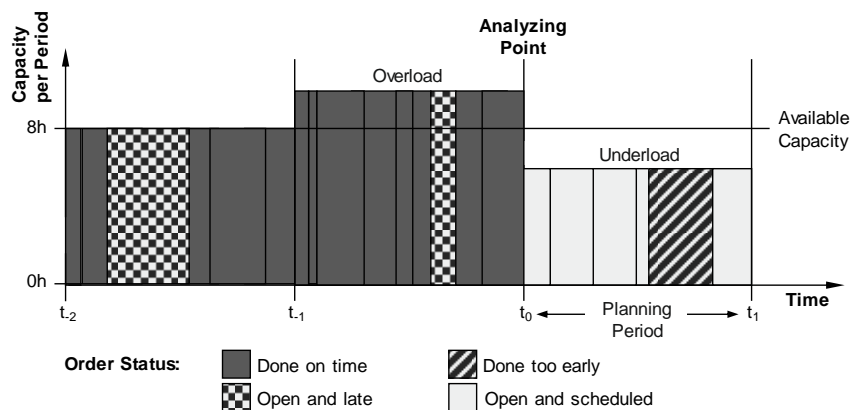


Fig. 1. Resource order schedule (assigned by due date, showing a backlog situation)

The backlog range is the time needed to reduce the backlog to zero (assuming that no new orders are planned) and calculated as the backlog divided by the planned capacity per period of the relevant resource [14: p.149]. Figure 1 shows an overall backlog situation with more late orders (too late done) than advanced orders (too early done) for one resource.

Management. Unmanaged backlog generally leads to longer throughput-times, which in turn leads to sequence deviations; both typically cause low on-time-delivery rates. For backlog management there are three general approaches sorted by effectiveness:

1. Allocation of **additional capacity** (e.g. overtime, extra shifts) to reduce backlog; given that sufficient work-in-progress is available at the backlog resource [4: p.551].
2. Planning of **less work content** in future periods to use the free capacity for backlog reduction, e.g. by external processing or declining orders in the respective periods.
3. **Rescheduling/postponement** of existing production orders to reduce the calculated backlog is generally not solving the problem but shifting it. If directly linked customer orders cannot be postponed (e.g. as customers refuse later delivery dates) this leads to low on-time-delivery rates.

3 Proposed Heuristic Approach

This paper develops a two-step approach for the operative bottleneck management: (1) Critical bottlenecks are prioritized based on their backlog situation and general management guidelines are derived. (2) The contributions of different causes to the overall backlog are described, quantified and countermeasures are suggested.

3.1 Prioritization / Resource Selection & General Management Guidelines

Real world production systems have many resources. Therefore, a practical prioritization approach is required in order to wisely use the limited management capacity.

Prioritization / Analysis. The task of factory planning is to evaluate and design static, long-term bottlenecks. For a first structured and qualitative identification of requirements for the PPC design and operation, the turbulence profile can be used [12]. For the evaluation and prioritization of actions, short-term bottlenecks with already fixed production orders (and due dates) additionally the backlog situation should be considered for taking appropriate actions while keeping high on-time-delivery rates. To facilitate the analysis the so-called Backlog-Load-Portfolio is introduced (see figure 2): On the y-axis the current resource situation is expressed by backlog range. For benchmarks or heterogeneous backlog tolerances, a relative backlog range is recommended [15]. On the x-axis the potential to reduce backlog by capacity increase expressed by the load factor (or alternatively the utilization) is depicted. The bubble size reflects the operational cost of the resource, expressed in the machine-hour rate. The depicted limits are experience based and should be adjusted case-by-case.

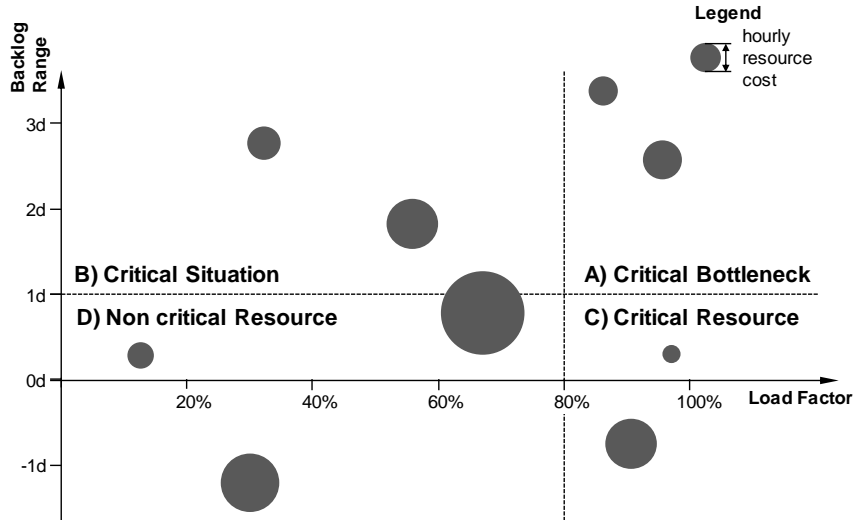


Fig. 2. Backlog-Load-Portfolio, based on [13: 208f]

General Management Sequence & Approaches. Based on the relative backlog and demand situation four general strategies are recommended in the following sequence:

- a. **Critical bottleneck:** The resource has high backlog and capacity demand. Alternative modes of flexibility (e.g. alternative or external resources) and/or the re-negotiation of due dates with customers should be considered.
- b. **Critical situation:** The resource has high backlog but low demand. The high short-term capacity flexibility allows capacity adjustments for reducing the backlog.
- c. **Critical resource:** The resource has low backlog but high demand. The resource should be monitored closely in order to avoid an additional backlog situation, e.g. by avoiding starvation or breakdowns.
- d. **Non-critical resource:** The resource has low backlog and demand. Because of its high short-term capacity flexibility any potential backlog situation can be counteracted. The resource should be planned efficiently, yet flexible if capacity is required.

3.2 Quantification & Countermeasures of Backlog for Selected Resources

Generally, backlog is caused by due-date deviation, throughput-time deviation or start-date deviation. Figure 3 depicts a more detailed hierarchy of backlog causes. This section describes definitions of the respective causes, approaches for quantification and potential countermeasures.

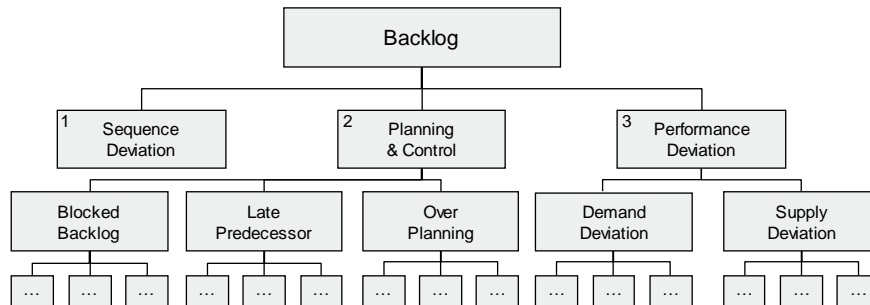


Fig. 3. Reasons for Backlog, based on cause-effect relationships of [4]

Sequence Deviation. Sequence deviation is the deviation of the planned and the actual rank of an order [4: p.85]. Within this paper the sequence deviation is measured roughly, i.e. all orders within one planning period have the same rank, so sequence deviations within a planning period do not affect the backlog situation.

Sequence deviations lead to symmetrical backlog and advance situations. Thus, their effect on the overall backlog is the minimum of too late and too early orders. E.g. a resource processed 8 h of work too early and is late for 20 h, thus 8 h of the backlog is caused by sequence deviation (the overall backlog equals 12 h).

Countermeasures are disciplinary measures or technical restrictions such as blocking of sequence deviations by IT or physical setup on the shop-floor e.g. by FiFo-rail.

Planning & Control. These backlog causes can be split into three groups.

Blocked Backlog. Orders that are too late and available at the resource, but not processable cause blocked backlog. An order is not processable if it cannot be started, independently from the capacity situation at the resource [5: p. 218].

The backlog effect of blocked orders (or blocked WIP) is the sum of orders in backlog which are not (yet) processable. Delayed processability regularly leads to sequence deviations [4: p.103].

For effective countermeasures, further analysis is required. Main reasons for not processable orders are missing closures due to technical quality defects, materials (in assembly processes) or manufacturing tools [5: p.218]. Depending on the order release strategy, limiting the work-in-progress by postponement of order release is helpful to keep stable throughput times [4: p. 102].

Late Predecessor. Orders that already late, but didn't yet (physically) arrive at the resource for processing cause backlog due to late predecessor, e.g. by order creation with past due dates or late order processing, both leading to start deviation for the resource.

The backlog effect of late preceding processes is the sum of all orders in backlog at a resource, but are not yet available.

The problem should be analyzed and solved going upstream starting at the last resource, including the planning activities such as production and sales order creation and

order release. To eliminate the root causes an analysis of the turbulence germs is recommended [14: p.209f].

Over planning. A resource is over-planned if the planned capacity exceeds the available capacity in a period [10: p. 677]. Planning more work content on a resource than capacity is available in that period leads to backlog in the magnitude of over planning. This is typical for the MRP approach, if the planning is assuming no capacity competition of orders (so called “planning against unlimited capacity”). The effect is calculated per period (and not per order) as the difference between the available capacity and the sum of work content for the respective period. Limited over-planning should be compensated with short-term capacity flexibility. A countermeasure could be the use of APS (advanced planning and scheduling) with limited capacity planning approaches.

Performance Deviation. This difference between planned and actual performance within a period can be caused by supply and/or demand deviations.

Underperformance (because of lower capacity than promised or higher demand, i.e. standard times, then planned) leads to backlog in the magnitude of the deviation between planned and actual performance. Two cases of performance deviation can be separated within an analysis period:

1. Cases with a reduced performance compared to the planned performance within an analytical period. Here the causes are ambiguous and cannot be assigned clearly to supply or demand deviations. Thus, the guideline is to increase the sampling frequency of the time series data to reach the second case (see figure 4).
2. Cases with non-performance within an analytical period. These cases generally can be classified unambiguously as supply deviations with planned (e.g. maintenance) and unplanned non-performance (e.g. breakdowns) periods, where the latter contributes to the backlog.

The higher sampling frequency required for the second case are commonly achieved by data recording at resource level as production feedback data usually do not have such a high resolution. Thus a first action is tracking the resource status with high sampling frequency, e.g. by machine data acquisition or cyber-physical-systems. Remaining periods with less than expected performance (but not non-performance) can be caused by supply or demand deviations and need further analysis.

Countermeasures for backlog due to supply deviation (e.g. lower than planned resource performance) are generally programs for higher performance, such as preventive maintenance. Countermeasures for demand deviations (e.g. wrong standard times) are adapted demand data, e.g. by advanced calculation methods for standard times.

The distinction between supply and demand deviations in complex systems such as semi-automated assemblies with changing operator performance (supply deviations) but also learning curve effects (demand deviation) is ambiguous. The definition of distinctive criteria between the two categories as well as the required statistical analysis of these multi-causal issues is beyond the scope of this article.

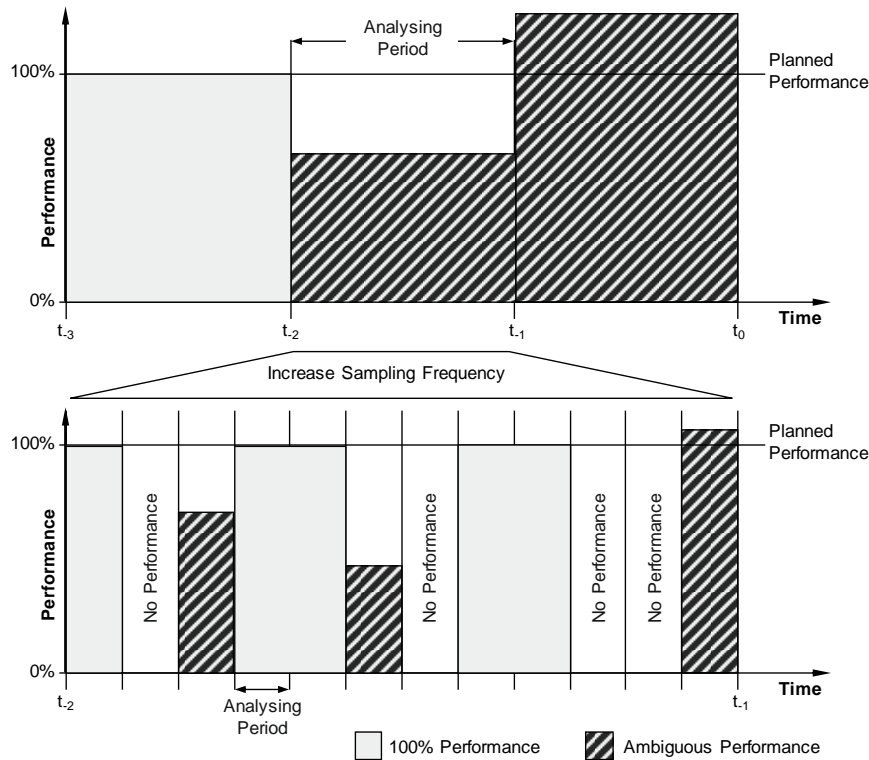


Fig. 4. Increase of sampling frequency to identify performance deviations unambiguously.

3.3 Limitations

The article's approach is on backlog management of single resources. However backlog can be inherited from other resources, e.g. due to wrong transition times or late finalization of preceding operations. As shown in above not all effects can be exactly quantified or even assigned to one exact cause. The analysis focuses on the current situation and does not cover past backlog dynamics over time or the proactive detection of bottlenecks by prediction. Sufficient data quality is a prerequisite for precise and trustable analytics [11, 12: p. 8ff, 222ff].

The order creation and calculation of the due dates is not part of the approach, e.g. the scheduling procedure applied or the handling of customer requests with shorter than standard lead times (so called express orders). In addition, an effective application of the presented results requires profound logistical competence of the managers.

4 Conclusion & Outlook

The article follows a two-step approach for backlog management in system operations with general management recommendations and a quantification of backlog reasons for

selected resources. Recommendations are given, when the applications of cyber physical systems could support a better understanding for backlog causes.

This approach supports the production management in focusing on relevant resources and to take correct actions to mitigate backlog situations.

Further research should be conducted on the required data quality, the automation of the analysis and suggestion of potential countermeasures as well as the automated detection and correction of supply and demand master data.

References

1. Adams, Joseph; Balas, Egon; Zawack, Daniel (1988): The Shifting Bottleneck Procedure for Job Shop Scheduling. In *Management Science* 34 (3), pp. 391–401.
2. Cox, James F.III; Blackstone, John H. (1998): *Apics Dictionary*. 9th ed. Amer Production & Inventory
3. Goldratt, Eliyahu M.; Cox, Jeff (2014): *The goal. A process of ongoing improvement*. Fourth revised edition, 30th anniversary edition. Great Barrington, MA: North River Press.
4. Lödding, H. (2012): *Handbook of Manufacturing Control*. New York, USA: Springer Publishing.
5. Nyhuis, Peter; Wiendahl, Hans-Peter (2009): *Fundamentals of production logistics. Theory, tools and applications*. Berlin: Springer.
6. Nyhuis, Peter; Wiendahl, Hans-Peter (2012): *Logistische Kennlinien. Grundlagen, Werkzeuge und Anwendungen*. 3. Aufl. Berlin: Springer Vieweg (VDI-Buch).
7. Petermann, Dirk: *Modellbasierte Produktionsregelung*. Diss. Universität Hannover 1995, published in: *Fortschr.-Berichte VDI, Reihe 20, Nr. 193*, VDI Verlag, Düsseldorf (1996)
8. Roser, Christoph; Nakano, Masaru; Tanaka, Minoru (2001): A practical bottleneck detection method. *Winter Simulation Conference Proceedings*, 2, pp. 949-953.
9. Roser, Christoph; Nakano, Masaru; Tanaka, Minoru (2002): Detecting Shifting Bottlenecks (*International Symposium on Scheduling*. Hamamatsu, Japan), pp. 59–62.
10. Schönsleben, Paul (2011): *Integrales Logistikmanagement*. 7. Aufl. Berlin: Springer.
11. Wang, Richard Y.; Strong, Diane M. (1996): Beyond Accuracy: What Data Quality Means to Data Consumers. In *Journal of Management Information Systems* 12 (4), pp. 5–33.
12. Wiendahl, Hans-Hermann (2007): Turbulence Germs and their Impact on Planning and Control – Root Causes and Solutions for PPC Design. In *CIRP Annals* 56 (1), pp. 443–446.
13. Wiendahl, Hans-Hermann (2002): *Situative Konfiguration des Auftragsmanagements im turbulenten Umfeld*. Heimsheim: Jost-Jetter, 2002. (IPA-IAO – Forschung und Praxis, Nr. 358). Zugl. Stuttgart, Univ., Diss., 2002
14. Wiendahl, Hans-Hermann (2011): *Auftragsmanagement der industriellen Produktion. Grundlagen, Konfiguration, Einführung*. Berlin, Heidelberg: Springer (VDI-Buch).
15. Wiendahl, Hans-Hermann; Steinberg, F. (2016): Montageversorgungsanalyse in der Einzel- fertigung. Ist-Situation, Potentiale und Ursachen. *WT Werkstattstechnik*, 105(9), 597-603.
16. Wiendahl, Hans-Peter (1995): *Load-Oriented Manufacturing Control*. Berlin, Heidelberg: Springer Berlin Heidelberg.
17. Wiendahl, Hans-Peter; Nyhuis, Peter (1998): *Engpassorientierte Logistikanalyse. Methoden zur kurzfristigen Leistungssteigerung in Produktionsprozessen*. München: TCW-Verl.
18. Yu, K.-W. (2001): *Terminkennlinie. Eine Beschreibungsmethodik für die Terminabweichung im Produktionsbereich*. Diss. Universität Hannover 2001, published in: *Fortschritt-Berichte VDI, Reihe 2, Nr. 576*, VDI-Verlag, Düsseldorf (2001)