TECHNICAL PAPER

Bottleneck Oriented Load Planning in Heat Treatment – Optimizing the Production Flow Saves on Time and Resources
Bottleneck Oriented Load Planning in Heat Treatment – Optimizing the Production Flow Saves on Time and Resources

Karl-Michael Winter
Process-Electronic GmbH, Dümnauer Weg 30, 73092 Heiningen, Germany,
km.winter@process-electronic.com

Abstract
Making money in heat treating is getting more and more demanding. There is obviously still the economical crisis that has caused the biggest impact on all of us, nevertheless knowing the true dynamic capacity of our - still partially turned down - equipment and being able to squeeze out even the smallest capability will optimize the cost to earnings ratio.

Everyone is using computer systems in order to plan their work-flow; however, often a disadvantage of such systems is their missing knowledge of what has to be done with the parts when it comes to breaking down the shop orders into real loads.

This article is giving an overview on the existing planning layers and what different methods are used and explains how a consequent planning around bottlenecks will result in an optimized throughput while at the same time trying to minimize the total number of equipment used – hence saving resources.

IT-Systems in Heat Treatment
Having a look on the various computer aided systems used in a modern company with a heat treating department we can group those systems according to their tasks on hand. Based on the number and the positioning within the company the result can be presented as a pyramid (Figure 1).

![IT-Layers within a Production Company](image.png)

Figure 1 - IT-Layers used in production companies
On the top we have the Enterprise Resource Planning (ERP) System handling all commercial activities in the company, like quoting, order processing, purchasing or invoicing and accounting.

In addition, as the name already suggests, this is the level where the planning of all resources takes place. In other words all existing equipment, the personnel and the money is planned in a manner to run all upcoming business processes as efficiently as possible.

Below the ERP is the production site control level where so-called Manufacturing Execution Systems (MES) are managing and (fine) planning the production processes. At this level all quality related data will be collected as well.

Below the MES is the control layer where Supervisory Control and Data Acquisition Systems (SCADA) finally execute and survey the production, interfacing with the control and process layer.

The boundaries between the various layers are fuzzy and, depending on the language used such a set of interconnected systems might also be described in a different manner.

On the top we will then find an inventory control system communicating with one or more production control systems. The production control systems are exchanging data with control centers that are sending the parameters and commands necessary to operate the tasks on hand to the connected PLC's and controllers.

In this setup the boundaries between the different layers are as well fuzzy. Depending on the complexity of the production chain respectively depending on the production depth and the number of equipment used layers might be omitted, in which case their tasks will be distributed between the neighboring layers.

The roles of the different levels in the production planning differ depending on the abstraction level, i.e. the distance to the executing machine. Therefore from the ERP system down to the control centers and PLC’s a more and more detailed refinement is taking place.

A production planning on the enterprise level for that reason is not able to consider the real manufacturing situation at every single machine. The planning is based on assumptions where theoretical production capacities are balanced with calculated production times for the orders on hand. The goal is to reach an optimum in terms of usage of the machinery, through-put time and adherence to delivery dates. At the same all of this should require a minimum stock of material and parts in production in order to have as little as possible money locked up.

The resulting list of shop orders, roughly planned, will then be sent to the MES. The MES will now break down the shop orders into manageable production lots using its knowledge about the actual situation on shop floor derived by the machinery and shop order data acquisition and the operator schedules. These production lots will be scheduled onto the available equipment in such a way that the given production times will not be exceeded. The goal on this level is to come up with as short as possible throughput times while at the same time trying to optimize the machinery usage. To cut it short: whatever the ERP feeds into the MES has to go through the production in order to fulfill the set targets.

But the MES still is having a more abstract view onto the real production as it has to deal with a huge variety of machines and processes available in a company.

Therefore it might be necessary to do a further refinement of the planned shop order operations, especially if the equipment that is assigned to execute such an operation is not a single machine but a production cell consisting of several machines. Very often the MES has a view onto cells as being black boxes, not knowing about the machines within the cell. An operation in such a case will be something like „Heat Treating“. The scheduling of the included sub-operation like „Cleaning“, „Case Hardening“ and „Tempering“ onto the machines within the cell will then be done by the cells control centre. The scheduling might be done either manually or partially automated.
Typical Planning Methods
The different methods used in the production planning can be distinguished into the one that are using the push principle and the others using the pull principle.

By using the push principle new orders will be pushed into the actual planning and manufacturing situation and planned onto the available resources by running a continuous scheduling based on the given start and delivery dates. Using the production flows (work plans) attached to the parts to be produced the orders will be split into operations and every operation will be scheduled onto an according capacity. A capacity in this context is the sum of the machinery hours needed to execute a given operation, for example „gaseous nitrocarburizing“. Having done that, a capacity leveling function is checking for possible collisions. Resource overloads can be cleared by postponing non-critical orders or by scheduling overtimes. The planning result is a detailed list of operations defining what material respectively what order will have to be started on what machine or group of machines at a given date and time and when this operation has to be finished.

By using the pull principle on the other hand a new order will be planned at the end of the production chain. Again the order first has to be split into operations and the last operation is used as the starting operation for the further planning. The next-previous operation will get the order to deliver the material in time. The scheduling is therefore done backwards, pulling the orders through the production. This is reducing the complexity and also the time needed for the planning as now each workplace will manufacture the parts by taking them out of a buffer in front of the machine into a buffer at the end of the machine that is treated as the buffer in front of the next workplace according to the production flow. Each workplace can easily launch a re-order at the workplace in front of itself when seeing a low level in its front buffer. The goal of this principle is to have as little as possible stock between the operations and provide optimum delivery reliability. It is best suited to deal with serial production and high quantities, keywords: KANBAN, Just-in-Time.

Highly sophisticated systems are using algorithms applying mathematical models onto the complex planning problems in order to optimize the planning result. More practical oriented systems are aiming for getting a local optimum in a short time by running a series of trials and errors. The latter systems are acting similar to a human being and therefore their results will better match the expectations of a human production planner.

The requirements for the planning result are partially contradicting. Ideally it should reach:

- Minimum throughput times
- High delivery reliability
- Minimum stock
- Maximum capacity usage
- Minimum costs

Reducing the throughput times will come to lower parts on stock but on the other hand the usage of the machinery might be low. In an order oriented manufacturing the focus will be more on the delivery reliability and short throughput times. But on the other hand we don’t want to spend a lot of money on providing over capacities.

To solve this dilemma it is possible to pick a system that favors different requirements. We differentiate between inventory oriented planning and resource oriented planning.

An inventory oriented planning is trying to reach an optimum product flow at the end of the production chain by adjusting the stock of parts between the various operations. These buffers are exactly that big that the delivery reliability can be reached. The inventory oriented planning can be done load-oriented (push) or according to the KANBAN principle (pull).

Resource oriented planning systems try to avoid bottlenecks within the production chain. The Optimized Production Technology (OPT) approach optimizes the product flow through the operation having the smallest capacity. A bottleneck will always arise if the available capacity is smaller or maximum equal to the capacity needed for the orders to process. The bottleneck oriented disposition (EOD, German: Engpassorientierte Disposition) is focusing on material bottlenecks.
In short, OPT optimizes the capacity bottlenecks by varying the production lots whereas EOD adjusts the send-ahead quantities, i.e. the number of parts that will be delivered to the next operation in one shot; the optimization of the product flow at the capacity bottleneck will be solved implicitly.

**Fine Planning of Loads in Heat Treatment**

But how is this related to heat treatment?

The planning activities within the heat treatment differ in type and complexity.

In a captive heat treatment treating parts out of a serial production there will typically be an ERP system that might or might not be connected to a MES system, depending on the size of the company. The planning has to adjust to a steady flow of parts in order to grant optimum delivery reliability.

In a commercial heat treatment we will find completely different conditions and requirements. Here, bigger companies are as well using the ERP-MES approach in their IT environment whereas smaller companies which do not have to network a lot of production sites will rather use a highly specialized inventory control system, connected to production management and control computers.

Accordingly, the resulting task list (operations) delivered to the machines respectively to the production management and control computers will vary (Figure 2). This might be based on shop orders with production lots optimized to the manufacturing of parts but also already planned loads given to an automated heat treatment cell. In addition, the quality of the planning result also depends on the in-build knowledge of the planning software about the special requirements in heat treating.

![Figure 2 - Breaking down a customer order into consecutive operations on a production batch](image)

It is kind of simple to plan operations around machines and workplaces that operate with a direct reference to an order. In this case the workplace is only able to execute one order after another. If a shop order is registered on the machine all parts being produced by the machine will belong to this shop order and can be reported within the machinery and order data acquisition system to the planning system. This type of workplace does not have an internal storage.

This changes immediately if we are dealing with industrial furnaces. Conveyor belt furnaces and pusher furnaces are treating parts that might belong to many shop orders. The registered shop order will be at the front of the furnace whereas parts coming out of the furnace will belong to a shop order that has been de-registered.
several hours ago. It is getting even more complex if we are dealing with furnaces treating parts packed in baskets on trays. A load for a pusher or a batch furnace is able to hold parts of various shop orders (Figure 3). To make it even more demanding the size of the loads for the different machines in a heat treating shop can be different, resulting in the need of re-packing the parts between the operations.

![Figure 3 - Assigning parts of different orders to one load](image)

This is a situation that cannot be covered correctly by classical planning systems, this requires a human being. It needs years of experience in order to arrange heat treating parts in such a manner that it is possible to reach an as big as possible throughput while still all parts will be treated according to their specification.

**Sequencing and Load Preparation**

In the case that readily packed loads are entered into a fully automated heat treating cell the controlling computer system has to schedule the loads efficiently in order to give as short as possible through-put times.

An adequate method for this task will be to optimize the usage of the capacity bottlenecks.

According to the OPT philosophy:

- The usage of a non-bottleneck does not depend on it's own capacity but is given by other constraints restrictions within the system
- Activation and utilization of a resource are not synonymous
- An hour lost on a bottleneck is an hour lost of the total system
- An hour saved on a non-bottleneck is just a mirage
- Bottlenecks govern both and inventories

Imagine a cell consisting of several batch/iq furnaces, some drawers but only one washer, in this setup the washer will turn into a bottleneck the moment only loads with shallow cases have to be treated. The sequencing must assure that there will be no idle time on the washer. An additional constraint would be not to have too long waiting times for a load between case hardening, cleaning and tempering to avoid any hydrogen embrittlement.

If the cell consist of several pit type furnaces working into a common oil bath there will be a totally different requirement. In many installations moving the load into the quench has to be done manually or at least requires an operator to take care that there will be no fire due to a system failure. With the carburizing furnaces working 24x7 but the operators only available in two shifts during the working days the loads have to be planned in such
a way that one after another will be ready for quenching without too much delay. If a load would have to stay within the furnace for a too long time the parts could end up outside the specification. And obviously, having the quench waiting during operator times will slow down the overall performance.

The planning tool cyclically simulates all loads within the cell and as well all loads that are planned to enter the cell and creates a projection of the found local optimum displaying the load and operation sequence with start and finish times. Planned outages like maintenance etc will also be part of the planning (Figure 4).

If we now consider a department or even the whole heat treating shop as an automated cell - with a manual load transport - this planning approach can be applied with some extensions.

Chained machines or fully automated cells will be treated as single machines within the production planning and can be planned as either recursive cells within the department acting as the „main-cell“ or separated into single machines within the department.

Similar to the big brother ERP the planning tool combines machines into groups according to their capabilities. This enables the control computer to sequence both, a series of operations – case hardening, cleaning, tempering – and to resolve a complex operation – case hardening complete – in such a way that even a possible back-up manufacturing can be considered automatically.

Assigning priorities to the machines enables to schedule loads onto the equipment until a set maximum capacity is reached before requesting the next machine of the same group for the further planning. Assigning priorities to shop orders enables an optimized sequencing able to handle even fast-track orders that have to be done immediately. In order not to have the lower prioritized orders moved to the very end of the planning list the assigned priority will change dynamically depending on the total delay time.

Another problem that has to be solved is the differing sizes of the production lots of the machines we have to plan; the tray of a central washer need not hold the same number of parts like the tray of a batch/iq. In addition
we might have to re-pack the trays if two shop orders share one operation but the next operation needs to be done on different machines or requires different process parameters.

**Data Interface to the Superior Planning System**
When applying this type of fine planning the control computer also has to provide the machinery and order data and has to report all information like produced quantity, scrap quantity, outages, register and de-register time stamps to the superior computer system in the expected format. In this way the data can be used in the same way as usual and easily integrate into the planning tool on enterprise level.

**Summary**
Putting the focus onto bottlenecks while planning the complex material flows within the heat treatment, offers an elegant tool to optimize the throughput but at the same time the machinery usage. If this task is performed on the specialized control layer the result will much better adopt to the actual manufacturing situation on shop floor as it would be possible with the high level planning software on the ERP or MES with their more abstract view.