SYSTEMA
Art of Automation

Advanced Manufacturing Planning & Execution

A Guide for the Semiconductor Industry

FOLLOW OUR EXPERTS · SYSTEMA.COM/BLOG ·
Blog

Table of Contents

Introduction

Arranging work in a production process is a challenge in every factory. Typically, the path towards better tool utilization and higher throughput starts with going paperless. Once material and tool states are digitized, considerable gains can be achieved by working out dispatch rules to automatically select the next work item for a given tool in a more optimal way. This process usually starts around bottleneck tools and is often adopted for all tools soon after. These so-called dispatch rules are based on expert knowledge and intuition. Establishing an automated rule-based dispatcher allows factories to formalize common practices on the shopfloor.

An automated dispatching system provides not only considerable gains in productivity, it also enables transparency of shopfloor operation. Very often, there are measurable performance differences between shifts that can be leveled by establishing well-defined dispatching rules to control priorities when assigning material. A rule-based dispatching system only complements human expertise. By carefully tracking operator compliance with the dispatcher recommendations, the system can be iteratively refined to democratize *best practices* of shopfloor operations.

After paper has been banned, expert know-how has been democratized, and a high level of dispatch compliance has been achieved, productivity is usually greatly improved. However, since shopfloor management has recently partaken of *the grail of optimization*, the question typically arises if one could do even better? So, let's try to shed some light on methods that allow optimizing production beyond intuition and rules.

Rule-based dispatching relies on ad-hoc procedures to allocate material and capacity. Although being simple in preparation, this approach often turns out to be sub-optimal: maintenance, staffing resources, and complementary processes of production are projected separately and are often not considered sufficiently. This often leads to missed production targets, intra-organizational tension, and, most importantly, lost business and dissatisfied customers.

Here's the good news: there is a way to unlock more capacity once a rule-based dispatching system has been established.

The answer lies in advanced planning and scheduling (APS). By aligning efforts and data between operations, logistics, planning, and management, it is possible to dramatically increase throughput and maximize on-time delivery. Let's shed some light on how to fulfill every planner's dream to fully utilize capacity by planning iteratively with factory mechanics and the current factory state in mind. Unlike previous systems for planning in linear production lines, APS iteratively plans production based on a holistic data model of the factory – known as the digital factory twin – representing available material, personnel, plant resources, and capacity. APS is applicable in various situations such as the following:

- Make-to-order (versus make-to-stock) manufacturing
- Non-linear lines with cyclical or repetitive product routes
- High-mix / low-volume settings with every low-volume order competing for resources
- Complex manufacturing flows with hundreds of steps and / or components
- Highly dynamic environments with complex / fragile physics, mechanics, or chemistries in tools (resulting in unpredictable downtime for setup / retooling) or material (resulting in scrap)

Successful manufacturing often comes down to the manufacturer's ability to fully maximize every aspect of production. With complex products, process steps numbering upwards of a thousand, reentrant processes, and little margin for error, optimizations provide significant competitive advantages.

READ THE HIGH-AUTOMATION WHITEPAPER

In such settings, it is likely these companies have made substantial investments in capital and operations but have failed to make adequate investments in methodology, hindering their ability to effectively plan and schedule.

What is scheduling?

It's important to recognize the subtle, yet important, distinction between scheduling and dispatching. Automated dispatching systems are designed to respond to events on the production floor by evaluating events and resources in real-time in order to optimize production flow. Event-driven dispatching keeps equipment operators informed of the best lot to run, at the current moment, on the equipment they are operating. These systems synthesize and evaluate real-time production dynamics in order to provide nearly immediate updates to operator dispatch lists. This means that production experiences virtually no delay in receiving help to prevent excess tool wait times. For best results, the dispatching rules should be the same for every tool in the factory.

Scheduling is the process of arranging, controlling, and optimizing work and workloads in a production process or manufacturing process.

In contrast to automated dispatching, which creates an ad-hoc schedule while being executed, a scheduler works out an optimal plan **ahead** of its execution. By doing so, it allows for more efficient allocation of plant resources such as machines and personnel. Its purpose is to minimize the production time and costs by telling a production facility when to make what, with which staff, and on which equipment. Manufacturing scheduling is an algorithmic planning method to ensure production processes and purchase materials are aligned with orders and production targets.

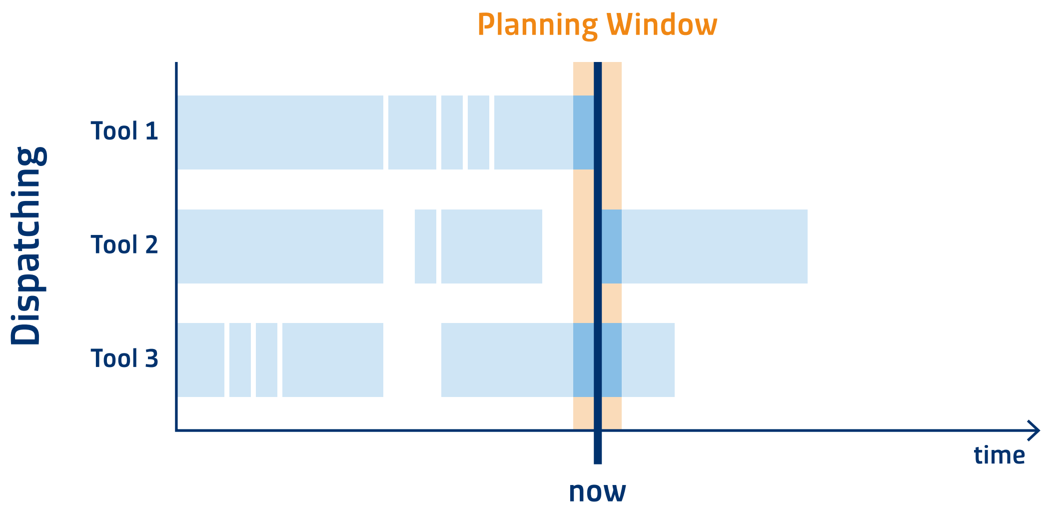

*Figure 1: Dispatching creates an **ad-hoc** schedule while being executed.*

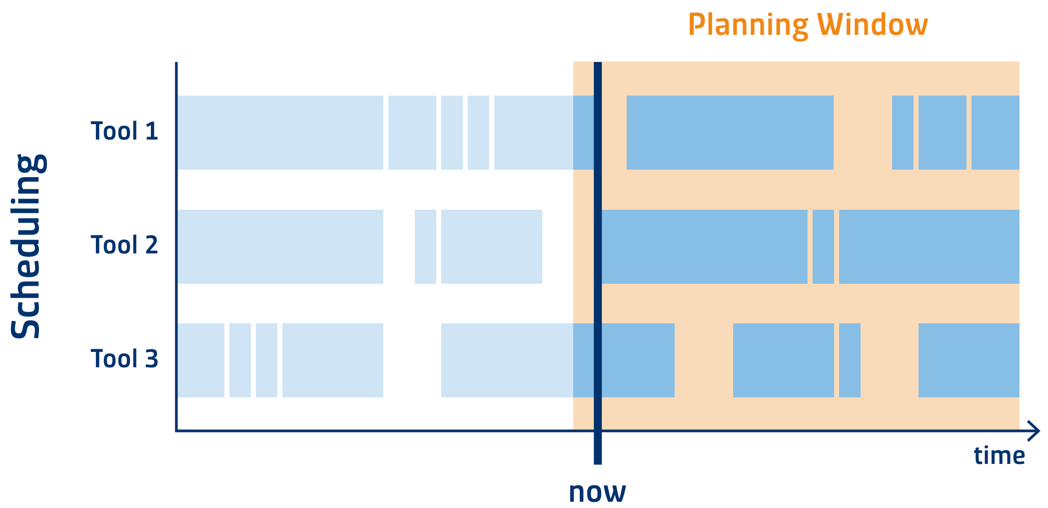

*Figure 2: A scheduler works out an optimal plan **ahead** of execution.*

Scheduling provides the following benefits:

- Optimizes production execution by relating materials with methods on the right machines driven by a qualified operator
- Meet or even exceed production targets
- Increases production efficiency
- Reduces time and cost related to process change-over or retooling.
- Balances and reduces WIP (Work-in-Progress)
- Aids in personnel load leveling and shift planning
- Predicts accurate delivery dates

Scheduling is a hard problem

In factories without existing automatic execution planning, rule-based methods should be implemented first to provide a significant boost to factory performance. However, such rules and heuristics usually do not provide an optimal solution to fully unravel a factory's capacity. That's because rule-based approaches typically create an ad-hoc schedule without arranging all known work in time. Instead, ahead-of-execution planning is needed to squeeze all of the juice from a factory: job arrangement and selection become more complicated quickly, often not following human intuition. For example, when two or more processes share resources, it may be non-trivial to find the best schedule.

Conceptually, scheduling is an ahead-of-execution planning method, where goals shall be optimized given limited resources under multiple constraints (figure 3). There is a wide variety of methods to calculate production schedules. The bad news: it can take a significant amount of computing power if there are a large number of tasks, tools, steps, or competing orders.

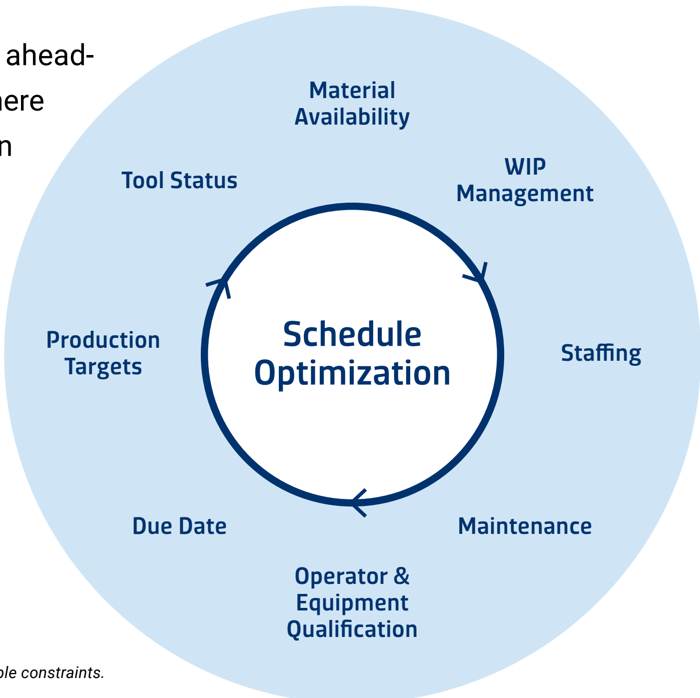

Figure 3: Scheduling optimization must account for multiple constraints.

Scheduling use cases in manufacturing are in most cases NP-complete / NP-hard, so it is easy to verify a given solution to a problem in a reasonable time. But it may take an unreasonable amount of time to find an optimal solution. There is no magic shortcut or silver bullet here: Brute force computing is often not an option, and quick algorithms that rely on heuristics tend to not provide an optimal solution in many situations.

Practitioners may need to adopt problem-specific simplifications in order to get applicable solutions without eliminating critical components of their scheduling models. By analyzing the pain points of a production site, bottlenecks can be revealed. Non-critical entities such as work-centers that do not relate to capacity gain or loss can be abstracted away, leading to a reduced problem complexity.

When writing down the scheduling problem for a particular production site, planning engineers need to define the required resources and processes. In particular, they need to work out hard constraints that define what a feasible schedule must look like, and soft constraints that describe what makes a good schedule. A hard constraint **must not** be broken. For example: To produce a part, step A must be executed before step B. Soft constraints **should not** be broken, if possible. For example, an order should be produced by Friday. Using a constraint programming engine, decisions on variables and values are made and, after each decision, a set of logical inferences to reduce the available options for the remaining variables' domains can be performed.

Another approach is agent-based modeling, which describes the processes and machines as actors and constructs a feasible schedule under various constraints using simulations. More recently, AI methods have been proposed as a solution to overcome some of the challenges specific to computational time and problem formulation. However, it remains to be shown if / how such methods can be converted to a real-time environment. By combining different methods such as mixed-integer programming, simulation-based optimization with constraint-programming, or AI, hybrid approaches may be preferable in pursuit of a balance between the solution efficiency and the schedule's performance.

Example: Semiconductor Manufacturing

Semiconductor manufacturing is known to be one of the most complex and expensive manufacturing processes. Unlike linear assembly lines, semiconductor manufacturing machines and processes are replicated along each product route. Each product and / or production process will have its own rules and logic. This, naturally, imposes multiple constraints when characterizing the schedule – both at the existing equipment as well as the entire shop floor. For example, given a 10-step process, a disruption at step seven is going to impact not only the material at step seven but also the material currently at step three with a future customer-commit date. Conversely, a disruption at step three will likely impact the scheduling at future steps of this material as well as future material. As a result, semiconductor manufacturing often experiences trouble recovering from unforeseen events such as unscheduled equipment downtime, unexpected changes in shift plans, or facility evacuations. Still, there is a constant need to improve planning capabilities to maximize utilization and minimize bottleneck resources.

In classical scheduling theory, a semiconductor factory would be modeled as *Job Shop Problem*. It's an instance of a so-called non-timed production planning problem. Multiple jobs are processed on several machines. Each job consists of a sequence of tasks, which must be performed in a given order, and each task must be processed on a specific type of machine. For example, jobs could be wafer cleaning, chemical wafer decomposition, or etching.

Different constraints characterize a job shop problem:

- Precedence constraints: Tasks need to be executed in an order defined by the product route definitions. That is, no task for a job can be started until the previous task for that job is completed.
- No overlap constraints: Machines can often only work on one task at a time.
- Batch constraints: some machines such as furnaces may allow / require pooling (a.k.a. batching) of production units into groups for processing
- Tasks, once started, must run to completion
- Machines can have duplicates or belong to groups of identical machines
- Machines can have sequence-dependent setups
- Machines can require a certain gap between jobs or no idle-time
- Processing times per job could be deterministic or follow some statistical distribution

Scheduling attempts to fulfill various optimization objectives on each machine. Common objectives of the job shop problem include:

- minimizing cycle time
- ensuring high-priority lots are preferred
- ensuring due-date requirements are met
- maximizing equipment utilization

Deficiencies with existing planning practices

There's no reason to change anything if nothing is broken. So, the first core principle in scheduling is to acknowledge what doesn't work (or is inefficient) and what scheduling will resolve. Among those deficiencies might be the following:

1. The invisibility of shop floor variability

The root causes for why planning and execution diverge so often is the constantly evolving factory state. *"Stuff happens"*, such as:

- a tool may go to a *down* state
- a bottleneck tool engineer may need a longer coffee break than usual
- unscheduled maintenance tasks occur
- staffing is short due to illness or other reasons
- execution / misprocessing errors occur
- manufacturing priorities change
- supply chain disruptions impacting availability of raw materials for production

READ THE ROOT CAUSE ANALYSIS CASE STUDY

While scheduling methods can incorporate forecasting of the manufacturing process, the future is uncertain. Even the smallest disruptions in resource availability may result in adverse cascading effects causing the most optimal production schedule to be obsolete and unusable. To account for unpredictable changes, the factory schedule needs to be updated frequently, regardless of how the plan is established, recorded, or implemented and a reschedule calculation should be done as fast as possible.

2. Lack of (or missing components within) the digitization of the manufacturing model

A schedule can help to streamline production only if it is grounded in accurate and current data. Production schedules rely on accurate setup and runtime data. Estimated or assumed processing times will not always reflect actuals, which means a schedule can't ever be in sync with shop floor reality because of incorrect assumptions in the first place.

> By implementing an integrative holistic view on human, machine, material and methods, algorithms can account for dynamics in production when building a schedule.

To realize advanced planning and scheduling, it is necessary to establish visibility into processes, materials, personnel, and machines in the form of a digital factory twin. Without a holistic view, planners and algorithms lack the data to identify changing shopfloor conditions. Fortunately, equipment automation and IIoT are powerful tools used to digitize the factory quickly and efficiently. Digitization in production will enable real-time visibility to account for dynamics in production when building a schedule.

READ THE INDUSTRIAL INTERNET OF THINGS CASE STUDY

3. Suboptimal job-start policies

Due dates represent a commitment to the customer. Adhering to due dates is one of the highest priorities for planning and management. But how can due dates be used most efficiently to govern production at scale?

One common strategy is to start orders immediately after they arrive. This is motivated by the idea (maybe misconception) that the earlier work starts, the earlier work is done. However, the early bird does not always catch the worm. In fact, with too much WIP flooding the shop floor, material movement and waiting times will increase dramatically – similar to how we often get stuck in traffic during rush hour. Furthermore, having too many jobs in WIP confuses true priorities leading to incorrect data used for dispatching decisions.

By tightly controlling WIP, based on a factory scheduling solution, order cycle times can be minimized significantly. An optimized order / lot release policy ensures that a factory is neither running dry nor is drowning in WIP.

4. The risks with over-relying on AI

Data-driven decision-making and schedule optimization can dramatically improve efficiency and throughput. However, scheduling algorithms are still maturing; this requires a level of skepticism, constant re-evaluation, and confirmation that all aspects of the system are working as intended. Following the scheduling software's results without factoring in conditions that may not be able to be addressed within the model (due to complexity or time) may result in sub-optimal plans. After the initial excitement wears off, planners often do not see the expected / promised gains in production from scheduling applications. In some cases, the decision will be made to roll back to former ad-hoc rules (FIFO, CCD, Critical Ratio) for WIP control and prioritization.

5. The ubiquity of the classical manufacturing model

Classical production planning is often designed to answer the question, *"When will we have the capacity to process an order?"* To do so, planners often rely on using an ERP system for planning for the next quarter ahead. This capacity plan is then worked out into a more specific spreadsheet schedule that this finally executed on the shopfloor using dispatch list rules and shift planning. Such an approach relies on the assumption that a capacity-loading analysis is sufficient to arrange the orders in the correct order for executing in production. However, this is not necessarily the case. Due to the dynamic nature of resource availability, a plan might be outdated before even reaching the shop floor. Attempts to execute production using a schedule prepared weeks or even months ago will always, inevitably, fail. Plans must be in sync with reality to realize the true potential of a production site.

So even with all the love and effort that goes into plan and schedule preparation, the resulting plans are sometimes not viable and may, in fact, impede production.

Core principles of manufacturing scheduling

To promote advanced planning with data-driven scheduling, the following core principles apply:

1. Establish a digital factory twin

To enable a schedule that accounts for production dynamics and variability, the current state of the shop floor must be digitized.

This will start with adopting paperless material tracking, but must finally culminate in a digital factory twin that provides a rich holistic picture about what's going on at any given moment. What is the status of the tools? Where is all the WIP and in what state of completion? What personnel are available today? What urgent maintenance actions are scheduled in the afternoon?

Also, a mature master data model is required. This data model may include (but is not limited to) the following:

- Steps
- Process Flows
- Alternative Flows
- Products
- Recipes
- Equipment
- Materials
- Cycle Times
- Due Dates
- Reticle Management
- Run-to-Run Data

2. Control your WIP

A mountain of WIP is a common impediment to production efficiency. As demonstrated by Little's Law – a central dogma in queuing theory – sending more work into WIP will slow down shop floor operations. Too much WIP confuses true priorities and increases the chance that a *wrong* job could be selected.

In contrast, a factory-state aware lot release allows for controlling how much new material is sent into WIP. It should not be realized based on assumptions but should be baked into the scheduler model. More specifically, a sensible WIP balance should be modeled as a soft constraint when optimizing the schedule. This will allow factory management to prioritize work that achieves defined production objectives such as due dates while maintaining a workable WIP.

3. Prioritize on-time-delivery rather than due-date

In many factories, due-date is the default method for work prioritization. In some cases, scheduling is often driven by due dates only. The sooner the date, the more urgent a lot / workpiece is considered. However, when doing so, planning does not account for if a job is in danger of actually missing its due date.

Let's assume two competing jobs exit with due dates in four weeks and two weeks respectively. The first job may be in more jeopardy to get done in time because some downstream process step may compete for a different set of orders on a specific tool. So job complexity and resource capacity along remaining product routes define the truly optimal job order. So, it may be beneficial to process jobs that are due later down the road first.

To assess such effects, it's critical that a planner or a scheduling program has visibility into what's happening once a job is in process. If not, a schedule may be based on false intuition or incorrect expectations, leading to suboptimal fab performance.

A path towards better planning and data-driven scheduling

Every production site is a unicorn with different requirements. That's why off-the-shelf solutions often fail to deliver or lack the necessary integration interfaces with existing shopfloor systems such as equipment integration or dispatching. What's recommended in most cases, is to build on a solid established factory scheduler model designed for a matching type of production (job shop, product mix, objectives, constraints) which can be customized to the specific needs, rules and – always present – oddities

Digital twin refers to a virtual model – a comprehensive physical and functional representation – of components, products and process within a production environment.

of a factory. Do not reinvent the wheel. To minimize rollout and integration efforts while maximizing a positive impact on the production, choose an advanced production scheduling solution that is tailored to the needs model of your manufacturing business. In particular, it is important to choose a system that can be integrated with live data from the shop floor.

Underpinning almost all scheduling solutions is a formalized description of the production processes and required resources. Such a process model often does not exist in the first place and needs to be worked out first.

Combining data from a digital factory twin with the factory model allows setting up a generative simulation model of the factory.

This model sets the foundation for optimization methods described above, which will – if executed correctly at a sensible level of abstraction – provide a scheduling solution to guide shopfloor operation until it has run in sync with reality.

Apart from data-driven advanced scheduling, there are other benefits in establishing a digital model of a production site. It provides means to explore so-called what-if scenarios. E.g. it allows projecting how changes in machinery or staffing would impact production. Similarly, such models enable a feature often demanded by customers, which is calculating order completion dates including confidence intervals.

To enable optimization using data-driven scheduling the following steps need to be implemented.

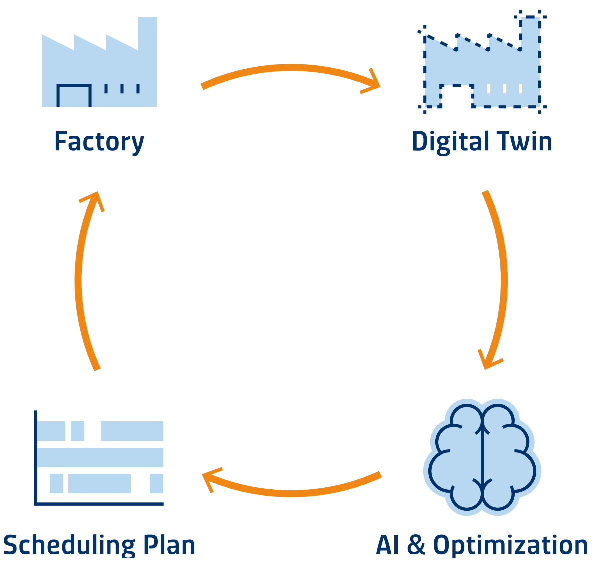

Figure 4: Combining data from a digital factory twin with the factory model allows projections for determining how changes in processes, machinery, or staffing would impact production.

1. Measure

…what is happening now at the shopfloor and establish a digital factory twin. Often this relies on additional digitalization to establish a holistic 4M digital twin of the line, work-area, or factory. IIoT can be a great tool here, giving planning tools and staff a more complete picture of resource state and capacity. Even without any scheduler being in place, such visibility into processes will help to reveal areas where production is hitting capacity issues or bottlenecks are emerging. A more digitized production environment will open up many opportunities for optimization and cost-reduction.

2. Analyze

…to identify inefficient spots in the production chain. Analyze order date fulfillment for trends and variability. Unravel the inherent mechanics / rules of your production.

3. Define

…the most mission-critical business objective to be optimized by the scheduling solution. That sounds easier than it is in practice where often many competing goals coexist initially in any larger organization.

4. Slice and dice

…the problem to minimize the scheduling problem to be solved. Combinatorial explosion is unavoidable if too many variables are considered for optimization. Break processes down into manageable pieces, and choose a correct model of abstraction that is based on the defined business objective.

5. Model

…your manufacturing process as a constraint optimization scheduling problem.

6. Implement

…a simple, FIFO-based approach and evaluate results. This can be done manually by only processing the *oldest* material at a given operation / step, or, better yet, can be implemented with a dispatching solution, such as SYSTEMA's Event-Driven Dispatcher (EDDi). By default, EDDi enforces a FIFO dispatching approach. Many benefits will be realized by this initial implementation, including:

- Validation of MES context information
- Results from an initial *controlled* approach to scheduling through dispatching
- Better results specific to known goals and customer requirements

With a basic dispatching scenario in place, more complex scheduling can be applied, such as due dates, cycle-times, waypoints, Kanban, SPC-based *golden path* processing, dynamic lot prioritization, enforced compliance of material choice by floor operators… and more as mentioned in the next point below. Really, the sky is the limit!

7. Play

…what-if scenarios. Compare scheduling solutions / heuristics against the status quo as the current baseline to identify the most suitable scheduling approach. Use the model to understand the relation between WIP, line balancing, and capacity. Simulate disruptions such as unscheduled tool-down, reduction in personnel, or high-priority orders. This will help to understand inherent factory mechanics and will pre-assess planning alternatives in the case of actual breakdowns or bottlenecks. Assess how more or fewer tools would affect production outcome (or not).

8. Operationalize

…with the factory dynamics in mind. Synchronize the scheduling model with the digital-factory-twin as often as possible. Always strive to keep your plan up-to-date with plant reality to avoid an unfeasible schedule. Be prepared to have a backup execution policy in place, in case the planning solution runs out of sync and needs to be readjusted first.

9. Execute

…the plan to fulfill the promise. A dispatcher gives a practical shape to a production plan by presenting the specific action items at the right time to shop floor personnel. Dispatching is concerned with giving a practical shape to an overall production plan.

Summary

Deriving a schedule for semiconductor manufacturing is known to be very hard. In SYSTEMA's experience, only a few top-runners in the market have been able to operationalize scheduling in semiconductor manufacturing to optimize production beyond ad-hoc dispatching prioritization.

There is no shortcut to establishing planning practices that optimally utilize capacity. Rather, optimization occurs in stages. The first step towards better planning is shop-floor visibility. Gaining visibility into your shop floor will provide the necessary data to feed any scheduling program. Secondly, a rule-based dispatching system allows to formalize human expertise and usually provides significant improvements for major key metrics such as throughput and tool utilization. Keeping the dispatching rules as simple as possible will simplify maintenance and enhancements.

Once the program is populated, business objectives need to be defined and the manufacturing process must be formalized into a scheduling model. Finally, the model needs to be parameterized with live data from a digital factory twin to account for the current factory state. Plans are known to be valid for a limited time, so a plan requires constant and regular updates to remain viable for shopfloor execution.

While setting up a factory schedule may seem expensive and tedious at first, it often pays off quickly in savings in machinery acquisition, more efficient tool utilization, increased throughput, and more orders delivered on time.

The results are happy customers, happy management, and more fun at work. Production scheduling is key to elevating manufacturing operations.

At SYSTEMA we strive to provide planning solutions that are designed with scalability, modularity, and operational stability in mind. With our long-standing expertise in digital twin solutions and equipment integration, we are a strong partner across many industries providing data and methods to elevate shopfloor efficiency to the next level. Our portfolio includes scalable data collection tools to capture the current factory state from shopfloor IT, MES, and IIoT. We provide data consulting to reveal bottlenecks and unused capacity. We can help optimize production with scheduling and AI to drive efficiency while reducing rework and waste.

REQUEST A CONSULTATION

Questions about smart manufacturing?
Please feel free to contact us.

SYSTEMA Art of Automation

SYSTEMA Systementwicklung
Dipl.-Inf. Manfred Austen GmbH