It's Not Rocket Science

A Primer on Planning and Forecasting

By Jeff Robinson

Planning is not particularly difficult to understand. Unfortunately, many people introduce a lot of unnecessary complications and special cases into the topic of planning. Forecasting, creation of projected ship schedules, planning lot starts in a factory, and scheduling of piece parts or bill-of-material items seems an incredibly complex problem

However, if you look at the basic elements of planning and identify the deliverables involved, it really is rather simple after all.

Planning is basically ***a set of reports***. (They might be associated with a manufacturing or production lot tracking systems or other activities). The planning system, however, does not track lots; it does not ship lots. All a planning system does is generate detailed reports about what has happened or reports that contain predictions about what will happen.

There are three basic elements of planning

- Demand
- Standards (yields and cycle-times)
- WIP

Demands

Demand is the list of time phased requirements; a list of customer orders, if you will. Basically demands are what parts are required by what customers in what quantities on what days.

Without demands a production or manufacturing tracking system cannot have a full planning system. All it can have is activity reporting and forecasting. If you do not have defined demands to meets, then you really do not have a plan. Indeed, any set of actions or lot starts will be equally acceptable. Forecasting will tell you when those lots starts will be complete and all start schedules will generate some form of output, but all forecast will be equally accurate.

WIP

WIP, or Work-in-Process, is simply an accurate accounting of lots currently within the factory. The WIP tracking system need to identify all lots that exist with their quantities, current locations and the positions on their respective processes. The planning element of WIP consist not only of current lot information but of lot history as well.

Standards

Standards are expectations. They are the values of cycle times and yields for each process increment (or stage). Summing up the cycle times for all of the stages in a process results in the cumulative expected cycle time. Multiplying all of the yields for each stage results in the expected cumulative yield for a process. These expectations can be used to make predictions about where a lot will be and how it will yield. It can be used to project movements and be used to compare with current processing activity to detect changes that occur and to identify deviations from plans.

Application of Basic Planning Elements

With WIP history alone one can generate	Actual Cycle Time and Yield reports [W]
With Actual CT and Yield Reports and Standards one can get variance reports [W,S]	Actual versus Standards
With WIP and Standards one can generate	Equivalent Finished Goods (EFG) or WIP flush Projected Ship Schedules, Individual Lot schedules, etc. [W,S]
With EFG and Demand one can perform and generate	Lot Pegging Allocated and Unallocated WIP Projected Delinquency Reports [W, S, D]
With Demands and Standards on can generate	Time Phased Gross Material Requirements [S,D]
From Gross Material Requirements and EFG one can generate Excess WIP reports	Time Phased Net Material Requirements and [W, S, D]
From Net Material requirement and Standards one can generate	Start Schedules [W, S, D]
From Lot Pegging and EFG and WIP one can generate	Dispatch Reports [W, S, D]

Definitions

Actual Cycle Time and Yield reports

Generated from WIP histories, ***actual cycle times and yield reports*** are essential. These values, if consistent and accurate, should be used to establish values for standards. However, there are some environments where actuals are not available. In some cases, (like R&D labs) few lots go through the same processing and therefore histories of only transient value.

Actual versus Standards variance reports

Generated from Actual CT and Yield Reports and Standards, ***Actual versus Standards reports*** are the primary tools used to detect when reality begins to differ from previous expectations. Like SPC charts, these reports detect changes as they occur. If not in place, that standards used to generate all other reports may become invalid and no one may notice.

Equivalent Finished Goods (EFG), WIP flush, Projected Ship Reports

Generated from WIP and Standards, ***EFG reports*** are the basic predictions of what lots or parts will be shipped in what quantities on what days. This is essential information

Individual Lot schedules

Also generated from WIP and Standards, ***Individual Lot schedules*** are reports that show where specific lots will be as they complete their processing. ILS reports show what stages will be transitioned on what dates and with what quantities. This is the first and most detailed forecasting of lot activity.

Lot Pegging

Lot Pegging is the matching of individual lots to specific orders. By projecting when lots will be completed (EFG) and comparing it to Demands determine what lots need to satisfy which specific orders. This is an essential activity in order to determine what the required date for individual lots should be. If some lots scrap out, other lots may need to be accelerated and to be assigned new required dates in order to satisfy previous demands.

Allocated and Unallocated WIP

These reports are based upon the pegging process**.** ***Allocated WIP*** are lots that are pegged to orders. ***Unallocated WIP*** are lots that are not pegged to individual orders.

Projected Delinquency Reports

If a lot is pegged to an order but the lot is predicted to be completed late, then one can identify which specific orders are projected to be delinquent. It is a valuable report that allows planner to identify delinquencies BEFORE they occur. When delinquencies are thus forecast, changes to demand can sometimes be made or customers can be notified in order to minimize the damage that might result.

Time Phased Gross Material Requirements

Demands and Standards can be used to generate ***Time phased gross material requirements***. This is often called a “Bill of Material Explosion”. It is a report that shows what parts would have to be started in what quantities on what days in order to meet the demands that have been defined on the system.

Time Phased Net Material Requirements, Excess WIP Reports

If you take the list of materials that you will need (Gross Requirements) and subtract from them the materials that we are projected to have on particular date (EFG), then you now have a list of ***Time Phase Net material Requirements***. This is basically a report that shows you what excesses and what shortages you will have of what parts in what inventories on what days.

Start Schedules

From this list of net shortages, you can automatically determine what parts will be needed to be started in what quantities on what days (in order to meet current demands).

Suddenly all of the pieces begin to come together. Now the system can not only determine what shortages might occur over time, the system can actually determine what lots should be started in order to meet demands.

Levels of Planning

Strategic

Product development, the coordination of marketing and engineering development activities; in concert…

Timing is critical...what parts will you make...what parts will you sell. How many, to what customers. What factories should b build what overall capacities will be needed.

Tactical

What are the current backlog of parts; what is the current demand; what lots should I start over the next few weeks. What problem am I projected of the next few months; what does my forecast look like; am I going to be in trouble

Operational

What problems are there today; where are my bottlenecks, where is there too much WIP
Where is product stuck; where are cycle times too high or yields too low
Where is the line unbalanced; where are we beginning to deviate from plan
What lot do I track in next?

Dispatching

Ultimately all the planning in the world is just a set of reports that really don’t change anything. On the lowest level, an operator is faced with the task of tracking lots through a set of equipment. Their principle question with all of the lots in front of them is "Which lot do I track in next?" If they track the wrong lots, then some lots can fall further and further behind schedule, schedules will slip and customers' products will be delivered late.

The report that recommends what lot should be tracked next into a particular piece of equipment is called a "Dispatcher". With the use of a dispatcher, the operator can be told what lot to track in next in order to correct deviations from plan. In order for a dispatcher to work, all of the aforementioned elements have to be in place.

Without a dispatcher, however, all the planning systems in the world are useless reports that are taped to walls in an upstairs conference room. *Without a dispatcher all the reports in the world do not substantively change the reality of what occurs on the manufacturing floor. They are a waste of ink and do not result in any discernable changes.*

The dispatchers is the essential piece that "closes the loop" with the rest of the system. If a dispatcher is developed and deployed, then changes in demand or WIP will dynamically change the recommendations of what lots should be tracked next.

The incorporation of dispatching tools with these other elements results in "Closed Loop Regenerative MRP", a dynamically adaptive material Requirements Planning System.

Unfortunately, there are a variety of different dispatching algorithms that can be applied to dispatching reports. Distressingly, many experts argue about which algorithm is the best and the controversy about which should be used often compromises the ability to implement real-time dispatching. The reality that results is that different locations within a factory often employ different queuing priorities that work against one another and dispatching is left unaddressed or ineffective.

FIFO – First In First Out
Shortest time First –
Highest Priority First
Critical Ratio
The furthest behind schedule first

Each has different advantages and disadvantages. Each different dispatching algorithm results in different behaviors on the floor. Most unbalance the line or increase deviation from original plan.

FIFO for instance promotes chaos over time. Lots that get ahead of schedule stay ahead of schedule. Lots that fall behind stay behind. Variation from plan increases with each step in the process.

Highest priority first unbalances the line. High priorities are assigned to lots that are behind, and when they are so flagged, they race ahead. However, once they catch up, they continue to jump in front of other lots and force other lots to fall behind.

Critical ratio is a dynamic lot prioritization scheme that determines a lot's priority by the ratio of how log the lot has left to complete divided by the time remaining before it is requited. Lots behind schedule have critical ratios of greater than one. Lots that are ahead of schedule have critical ratios less than one. The problem is that this unbalances the line. Lots that start on time typically slow down until they are so far behind schedule that they start moving and then they race at the end.

The best line balancing algorithm is pure schedule lag. Lots that are the most behind schedule have the highest priority. Lots ahead of schedule slow down; lots behind schedule speed up until they catch up.

However, there is great controversy here because different people have different experiences and different preferences.

Logistics

The problem with all of these reports is that things are constantly changing. Some lots get further ahead, some lots slow down and fall behind. Pieces of equipment go down in unexpected ways. Other sources of variation cause things to change randomly.

Unfortunately we cannot tolerate systems that are too sensitive. We cannot change start schedules hourly. We cannot develop systems in which "the tail wags the dog". Therefore we need to "set things in concrete" periodically and establish sets of standards and start schedules that are stable for a time before we allow them to change.

This task of establishing a set of planning factors, (Demands, Standards and WIP), is called ***Master Schedule Preparation (MSP)*****.** When a snapshot of each of these is prepared, and all of the reports are generated; when lots are pegged to orders; then a Master Schedule is set. (It may only last for a week before it is changed or altered by realities of scraps and delays in the factory or changes to demands from customers, but in that interim everyone has a picture of the factory as it compares to customer demands and manufacturing projections.

The frequency with which the MSP is re-run and changes are allowed to occur is a reflection of the responsiveness and flexibility of the organization.

However, the greatest problem faced by planning managers are not the technical aspects of lot tracking databases and reporting systems, it is the administrative and logistic support necessary to put all of these elements into place.

1. The implementation of planning necessitates doing things differently than they way things have been done before. This means that, if you used to expedite lots in the past, that you must stop doing so and manage different problems in different ways.
2. Turing on planning also requires that you perform tasks that were never performed before. Everyone in the organization will be able to look you in the eye and tell you that these tasks “are not their job” and they will be correct. Yet somehow, someone will have to do these things. Someone will have to generate all of these new reports; others will have to read them. Some people will have to enter standards and find ways to effectively maintain them. As time progresses, the system will have to be maintained and new problems will be identified. People will have to be assigned to these tasks and new procedures will have to be developed to ensure that they are done.

These are the greatest challenges for a planning manager; training, maintenance, flexibility, adaptability, discipline.

Planning Principles

Before further discussion, it is appropriate to outline some basic principles about planning. Because of the many complicated situations that planning is called in to handle and because of the complications that arise when predictions are translated into possible courses of action, it is best to clarify objective and goals. In doing so, we find that basic approaches and strategies are defined and clarified.

1 - **Standards are expectations** (of stage yield and cycle time) that are used to predict what will occur in the future and therefore should reflect reality.

> This means that standards, if not derived from actuals, should accurately reflect what is or has previously been achieved.
>
> Whenever possible standards should be derived from historical actual cycle times and yields. However, there are environments, such as R&D fabs, which may have not actuals for individual parts and flows upon which standards can be developed.
>
> This is a very important principle. It is why it is listed first. If people put in place standards that do not reflect reality, virtually all other aspects of planning are compromised. Many managers install standards that are not based upon reality. They enter standards that do not reflect what is REALLY happening on the floor. They enter the cycle time and value of what they WISH were occurring. This way, reports show nice outcomes and lots are predicted to always come out on time. However, these predictions will be wrong (since they do not reflect current realities) and every report, prediction and plan will incorporate these errors. Things will look good initially, but the planning system will flounder and fail in the long run.

2 - **Reality changes**

> There are many factors that can change actual cycle times and yields; most significantly changes in equipment sets or WIP inventory levels. As conditions change, actual cycle times will change, thus standards need to be rigorously maintained in order for estimates and expectations to remain accurate.

3 - **Variation** in standards (particularly cycle-times) **must be minimized**.

> The amount of variance or expected errors associated with a standard stage cycle time is important to the accuracy of any forecasting. The greater the variance the lower the ability to make accurate predictions. In order to make planning accurate, variance must be minimized.
>
> It should be noted that any lot prioritization algorithm except performance to plan dispatching increases the cycle time variance of lots at each tracking step. The only

way to reduce cycle time variation is to base the selection of lots upon time based performance to plan algorithms.

If you select (or move) the wrong lot...you make thing worse!

4 - **Dispatching is** the most **important** aspect of implementing planning

The purpose of planning is to affect change that will improve performance or the on-time delivery of material. Unless the planning process provides a means to actually change the selection of lots waiting for processing on the factory floor, the entire planning process become ineffective. Somehow, changes in plan and deviations from expectations need to be reflected in different recommendations to shop floor personnel or ultimately nothing will change and no improvement will result from the planning efforts.

The use of performance to plan dispatching reduces cycle time variance and optimizes factory performance in achieving on-time delivery goals. Lots that get ahead of schedule will tend to slow down and allow lots that are behind schedule to catch up. Performance to plan dispatching provides a 95% efficient line balancing algorithm and repairs the unbalancing effects of unplanned stochastic events.

Unless something changes on the floor...nothing changes.

5 - **Performance to plan metrics** are based upon the difference between forecast end date (based on standards) and the lot's required date. (i.e. DEMANDS)

Whether this delta is expressed in terms of slack time or critical ratio is not important. However, the required date of a lot should depend upon what customer order a lot is pegged to. This means that for accurate on-time delivery, facilities need accurate demands against which they can generate a formal master schedule. Without demands on cannot assess realistically whether a lot is being required out on the correct date or in the correct quantity.

6 - **In order to achieve 100% of on-time delivery, all parts must be completed early!**

If lots are not completed early, then they are late. This also means that if demands are not received before lots are started, then there is no guarantee that they will be available.

Because there is variation in actual cycle time achievements, the lot required date must be moved earlier than the customer commit date, i.e.

Required date = MSD - variance

This changes the required date to be earlier than the MSD by the amount of observed or expected variation.

7 - **The goals of optimizing on-time delivery and high manufacturing productivity are mutually exclusive**.

Optimizing on-time delivery is different from achieving high levels of equipment productivity. The striving for the first goal results in the selection of the lots that are the furthest behind schedule thus minimizing cycle time variance. Efforts to achieve the second goal results in the selection of lots that reduce equipment setup and switchover times, thus keeping the equipment as busy as possible.

The theory of constraints emphasizes that bottlenecks need to be treated differently from non-bottleneck pieces of equipment. In order to optimize overall productivity, bottleneck equipment must remain busy and non-bottleneck equipment must at times be idle (since non-bottleneck equipment sets by definition have excess capacity). Thus two different dispatching algorithms are likely needed for these two types of equipment sets.

Note: Bottleneck equipment is generally limited to about 5% of overall equipment. Thus performance-to-plan dispatching can be applied about 95% of the time and bottleneck dispatching by setup should be applied the rest of the time. These tradeoffs reasonably reflect the ratios of real bottlenecks in typical manufacturing facilities. Applying either scheme (perf-to-plan or selection by setup) all of the time negates the other goal.

8 - **Bottlenecks** (low productivity points) **are where the WIP is.**

A bottleneck may be instantaneous or historical.

If there is no WIP at a bottleneck, then

1. the location is NOT a bottleneck (worry about real bottlenecks instead), or
2. there is insufficient WIP in the factory to keep even bottlenecks busy (increase starts to build WIP), or
3. the line is substantially out-of-balance (use perf-to-plan dispatching to fix line balance)

If you have to push things to get them to the bottlenecks...they are not bottlenecks or you simply don’t have enough WIP in the line!

9 - **Knowing Theoretical cycle times is nice; but not essential**

Knowing theoretical CT is useful to identify when certain required dates are possible or impossible. If new reports were written, it would be nice to know how actuals compare to the theoretical times (x2 or x3 or x4) But the determination of theoretical cycle times is a time consuming and arduous task that is beneficial more for metrics reporting than planning or dispatching per se.

If volume is significant enough, theoretical CTs may be inferred by calculating the 5 percentile mark for history at individual stages. This is faster than using traditional stopwatch measurements for each recipe in the fab.

10 - **To change when a lot will be delivered, change the required date, not the lot's priority.**

On time delivery requires lot selection by required date....not priority

If different standards are maintained for different priorities, adverse results will occur if a performance to plan dispatcher is utilized. If a priority 3 lot is a week behind schedule and the lot's priority is changed to priority one, and if different standards are maintained for the two priorities, then the lot that was a week behind schedule may suddenly appear to be a week ahead of schedule. The result of the different standards would be that increasing the priority causes the lot to stop moving until it is behind schedule again. This is not the likely behavior that was desired. One should change the required date to get a lot out earlier. Changing the priority will not get a lot out earlier. It will simply make is move faster when it is behind schedule.

11 - To understand changes in plan**, one must understand changes to demands**

Whenever possible one should retain the demands from the previous master schedule. Deltas and changes in orders can then be detected. Ideally individual orders should be numbered uniquely so that changes in dates of orders can be understood.

12 - **Whenever possible forecast demands should be in the future**

To optimize on-time delivery metrics, many facilities do not enter demands until after lots have been started or scheduled. This compromises the value of having demands at all (which is to drive a start schedule). Demands entered after a lot start is not a true demand.

13 - **Idle equipment is normal**

Often, people panic whenever a piece of equipment has a low percent utilization. This is normal in facilities that are not perfectly balanced (i.e. where the throughput capacity of all the pieces of equipment in the line are not exactly equal. Indeed there are four different reasons why a piece of equipment will become idle at any time. These include:

1 – Equipment capacity imbalance.

The exact percentage of the time that a location will be idle is given by relationship

$$100*(1-1/N)$$

where N is the multiple of capacity over the lowest equip in the process flow.

In a steady state situation, with no equipment variability, there is only piece of equipment that will never go idle is the location with the lowest throughput. ALL other pieces of equipment, MUST become idle for some period of time.

2 – Product Mix.

If you have a specific piece of equipment that is used only for a specific product or technology and if no product it currently in the line on that flow, then that piece of equipment will go idle 100% of the time. Again, this is normal.

3 – Line Imbalance

Another reason that pieces of equipment go idle is when normal stochastic variability is introduced into the line. Some lots get ahead of schedule and other lots fall behind. The amount of unplanned time that equipment becomes idle is a function of the amount of line imbalance that occurs. Some causes for line-imbalance include: equipment downtime, insufficient chemicals, lack of operators, improper lot selection (dispatching)

4 – Insufficient WIP

A final reason that equipment might be idle is when there is simply and insufficient numbers of lots in the line. Imagine an entire line of equipment with a single lot traversing the line on its process. Virtually none of the equipment will be utilized at any single time.

A review of appropriate actions to be taken for each of these situations might be:

Condition	Action
1 – Equipment imbalance	None
2 – Product mix	None
3 – Line imbalance	Rebalance line (dispatching)
4 – Insufficient WIP	Increase starts

Advanced planning

Time phased material constraints
Equipment Capacity Planning
Advanced reporting
Multiple case analysis

Implications/Environments

There are four basic environments that can implement planning.

1 - Low Volume/Low Mix

This is a PROMIS training class

2 - Low Volume/High Mix - Typical of R&D fabs

Lot processing in such environments may be on an exceptional basis. No two lots may go through the same processing. The development of actual cycle time data may be impossible in such circumstances. The development of actuals may have to be based upon theoretical processing times and the formulation of stage cycle times may best be based upon these aggregated recipe values filtered through lead-time conversion tables.

This environment is the classical “custom workshop”. Each item manufactured is a one of a kind item. Only a few items of that type may ever be made. Each one is manufactured to different specifications and be unique. (Indeed in Research and Development environments, a single lot may be split many different ways and each wafer may undergo different processing and be virtually unique.)

3 - High Volume/Low Mix - Dedicated production lines like MOS-10

These environments are ideally suited to PROMIS planning. The high volumes allow standards to be derived from actual lot histories. The low product mix allows the facility to build standards for individual parts or process flows.

This type of production environment models what is known as “repetitive” manufacturing. In such environments there is let emphasis on demand or customer orders since the same few product lines are being continually started. The opportunities in this environment are higher efficiencies because there is less equipment setup and switchover as lots are processed. Reporting is greatly simplified because of the few product lines involved. (This is an almost ideally manufacturing environment.)

4 - High Volume/High Mix - Large diverse manufacturing facilities or ASIC fabs, like MOS-6

These environments have sufficient lot transactions to permit standards to be derived from actual lot histories, but the high mix of different products precludes the maintenance of standards by part or process. The use of reference procedures by technology family will be the likely method of organizing a manageable set of processing standards for planning.

This is the most difficult environment. It is the traditional “job shop”. Every lot in the line may be different from every other lot in the line. The close association between an individual lot and a demand or order becomes critical. Equipment setups and switchovers are frequent, and efficiencies are correspondingly lower. At the same time, the high product mix can complicate the very task of reporting across the multitude of lots and part variations. Exception reporting become essential since manufacturing reports cannot be summarized the same way as in the High Volume/Low Mix environment. This is the realm where planning is needed the most. It is also the place where it is hardest to deploy.

The high number of different parts and processes will likely require additional tools to facilitate the reporting and maintenance of standards so that the setting of planning flags by part, process and lot does not become prohibitively labor intensive.