Drum Buffer Rope and Buffer Management in a Make-to-Stock Environment

by Eli Schragenheim and Rudi Burkhard

Introduction

During the 80ies Drum Buffer Rope assumed a make-to-order (MTO) environment. Since many implementations of DBR had at least some make-to-stock (MTS) production we simply transformed a stock order into an MTO by giving it a due date. We probably did not give sufficient thought to the question, 'what due date should we give to a MTS order?' The commonly used procedure is to generate due dates based on some sort of forecast of expected demand for the stocked item or to simply use average production lead-time to determine the due-date. Whatever was done, every MTS order in a DBR environment had a 'dummy' due date. Almost certainly the MTS production was not all needed on the chosen due-date. Some of the MTS was probably needed much earlier and quite a lot not needed until quite some time after the original due-date. The result of dummy due-dates is a very significant impact on schedules, lead-times and WIP.

The Distribution and VMI Viable Vision templates rely on stock – so a good MTS solution is critical to their implementation. Since the way of handling MTS does have a significant impact, it became necessary to develop new and better methods to dramatically reduce this.

We need to recognize the fact that in a MTS environment we cannot really have a bottleneck. Such a situation would soon result in too little stock and our Distribution or VMI solutions will quickly fail or fall into disrepute. We need sufficient protective capacity for any Viable Vision MTS solution – there can be short excursions when the CCR is a bottleneck, but these cannot (must not) last.

This document will present:

- Why we do have an absolute need for Make-to-Stock for some of the Viable Vision models.
- The problems of batching and forecasting.
- Using buffer management in a MTS and in mixed MTS & MTO environments.
- The simple rules we should 'play by'.
- Using MTS to build component stocks.
- Managing the transition to MTS (or to MTO).

Why Make to Stock (MTS)?

Make-to-Order (MTO) should be the normal and desired way to produce. It is the least wasteful since our target is to make only what the customer has ordered (and really wants). In this day and age of increasing consumer demand for variety MTO makes even more sense. The increasing pace of new product introductions and product customization, the pressures to reduce inventories and cost all scream for MTO, and yet MTS is often still absolutely necessary. The non-TOC community uses MTS coupled with forecasts to gain production efficiencies and to smoothen the load on production. Both TOC and Lean communities claim this is a flawed reason for making to stock. In the past the problem of insufficient computer power and data storage forced the manufacturing environments to produce to stock because there was no practical way to link customer and low level orders. With today's computing and telecommunication power etc. this is no longer valid. All the data and information you could ever want are available almost instantaneously.

There is, of course, a very good reason why we need to produce stock. Whenever our production and delivery lead-times are longer than the customer's willingness to wait (the client's tolerance time) – we must produce to stock. There are special cases, such as demand caused by special holidays like Christmas or promotions that the company initiates. These special cases are characterized by very high demand in a very short period. In this article we assume that demand fluctuates (including seasonality) within 'reasonable' limits without huge short peaks.

The Problem with MTS (from the TOC Perspective):

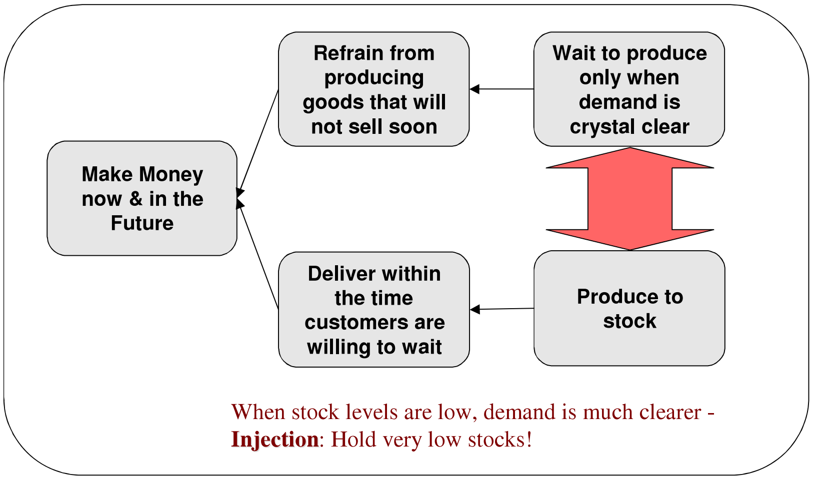
We cannot know whether we will ever sell all of what we produce to stock. In an MTS environment we face the following risks:

- We have to invest in significant amounts of materials that we may not sell for a long time so that our stock deteriorates and we can no longer sell it, or the product simply becomes obsolete. We certainly risk the loss of those materials we employed to produce the product.
- In many cases we dump the over-stock by selling it at reduced prices. In this case we may well lose a substantial part of demand for our newest products, which we would have sold for the full price.
- In many cases we consume precious CCR capacity that could have been used for a product truly needed by a customer now – instead of producing a product we don't need right now.

Unavoidably we will end up with too much of some products and shortages in others. Making to stock is actually a compromise between lots of inventory and no inventory. It looks like we need to push our solution as far as possible towards having no inventory – without risking our customers' patience because of a shortage. Our conflict looks like this:

Drum Buffer Rope and Buffer Management in a Make-to-Stock Environment

by Eli Schragenheim and Rudi Burkhard



<u>The Direction of the Solution</u>:

In an MTS environment **hold very LOW stocks** so that almost all of our stock will be sold soon; and

Replenish stocks very quickly and frequently – there should be <u>no batching</u> unless absolutely necessary. Use (S)-DBR to plan production and Buffer Management as the guide for priorities in execution. Our ability to replenish very quickly and frequently is of course the prerequisite for very low stocks. Our focus is on the various parts of replenishment time and how it can be reduced to the minimum possible.

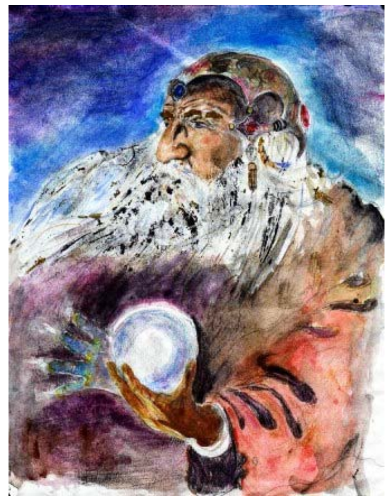
<u>The Batching Problem</u>:

Production likes (big) batches because they help solve an apparent problem – the continual heavy pressure they suffer to improve costs and efficiencies. Bigger batches apparently result in better efficiency and lower costs, this is the common belief, by minimizing the impact of (many) product changeovers. On top of that larger batches and fewer changeovers make life easier for the people on the shop floor – changeovers are hectic and stressful.

High inventory standards and forecasting far into the future encourage production planners to schedule large batches. 'Knowing' what will be needed (from the forecast), having the space for large inventories and the pressure for efficiency and lower costs lead to larger batches. To make matters worse, the min-max algorithm for inventory management mandates a minimum batch - the difference between max and min defines the batch. Big batches are almost built into the system as a policy.

Unfortunately big batches bring with them some dramatically undesirable effects. While we are producing a big batch, a large portion of the products produced is certainly not needed now, or even in the near future. Big batches are an excellent way to block our capacity from production of those things we **do need right now**! Big batches increase production lead-times and lengthen cycle times (the time it takes from producing a product to the next time that product is going to be produced). At the same time big batches lower due date performance and our capability to flexibly meet customers' (urgent) needs. Customer service, unavoidably, suffers. And we lose opportunities for better service that will lead to higher prices and/or greater sales volumes.

The Role of Forecasting in MTS:



Forecasting's job is to answer this question: "How many units of our product will be sold during the next period?" We have already seen that once we have the answer, it tends to help justify large batches through the perceived certainty given by the forecast. We use forecasts even though they are notoriously inaccurate – we use them because, 'what else is there to plan the future with?' Forecasts are seen as an essential tool (or maybe a crutch?) to plan and schedule production.

Actually forecasts do not pretend to be accurate. Forecasts were developed to provide partial information in order to improve the quality of uncertain decisions. The mathematical logic behind forecasting acknowledges the uncertain nature of the forecast information, and supplies an expected **average result** and an estimation of the **forecast error**. Our problem is most decision makers do not understand the nature of forecast information leading to decisions that many times are poor.

What do we need from our forecasts? Knowing what will be sold **on average** is not enough – we know very well that sometimes there will be fewer customers and sometimes a lot more. We need to know how much we **might sell** in the next period, and we need to decide how much of that amount we are ready to provide to potential customers. If we need, or want, to maintain high level of service, we must invest in more than just the average forecast amount.

The forecasting period isn't much help either. Typical forecasting periods are weeks or months. But, what is the horizon of demand do we need to make our decisions? Not necessarily what might be sold next week or even next month. The real horizon we should consider is our **replenishment time** – how long it will take to replace what was just been sold. We need to translate the normal forecast period to the period our decision depends on. Most decision makers do not know how to do that. What we **might sell** in 15 days is not necessarily half of what we **might sell** in a month. When we consider shorter periods of time the relative statistical fluctuations are usually larger. we also must take into account that the distribution of sales throughout a month is not necessarily flat. Some weeks in the month may well show higher sales than in other weeks. Since the concept of a (accurate) forecast is such an attractive solution, huge efforts have gone into improving forecasting techniques and methods – but have had little impact on the quality of actual planning and scheduling decisions. Products are scheduled for production only once in a relatively long period – because of the (false) impression of accuracy given by the forecast – while not recognizing that **forecast accuracy deteriorates rapidly the further into the future we look**.

This practice of producing once in a relatively long time makes the evaluation of how much safety stock is needed very difficult (safety stocks indicate that people do recognize the fallibility of forecasts). The longer the horizon, the greater the uncertainty and the higher safety stock the business needs. Management pressure for lower inventory levels causes many to have insufficient safety causing customer service to deteriorate. Instead of using inventory to protect sales, management policy seeks lower inventories because someone has said lower inventories are good. The cause-effects between batches, forecasts, inventory levels and customer service seem often to be ignored.

Attempting to solve the forecast accuracy problem by frequent re-forecasting introduces a lot of noise and nervousness and it becomes impossible to assess the forecast's validity.

Many businesses describe themselves as having something like 6 weeks of stock – without really understanding what that number means and the wrong impressions it can give. Some products in this "6 weeks of stock" will be sold out within a week, while others will cause us to wonder if we will ever sell them. The nice comfortable number of 6 weeks inventory hides shortages and surpluses. Articles are sold individually and not from an aggregated 6 weeks stock! (The devil is in the detail of each articles own demand). On top of that, our expected aggregate demand can be way off – so the 6 weeks we thought we had could be quite a different number in reality.

If forecasts are such a problem for us, why do we even bother to use them? It is pretty straightforward to use aggregated forecast to help us determine when we might want to invest for added capacity because such an aggregate forecast is relatively accurate (the forecasting error is small enough). But why would we ever want to use something as problematic as a detailed forecast by article and location – over a fairly long horizon? Maybe we had better look for an alternative approach for decisions regarding required stock levels of an article in a specific location.

<u>Isn't our Stock ≡ (Identical to) our Forecast?</u>

The reason for stock is to have enough availability for all the sales that are likely to occur between now and when we expect replenishment. Since we don't like to miss any sale our inventory is supposed to be enough to do just that. The amount of our inventory is an expression of the maximum we expect to sell until the next reliable replenishment.

Of course if our inventories are screwed up they won't reflect what we expect to sell. Only when inventory levels are correct do they reflect expectations of demand. In fact our inventory levels are not necessarily any more accurate than any other forecast, but, in the interests of avoiding confusion why don't we stick to the one forecast we must have anyway and try to manage it better? Our inventories may not be any more accurate, but they do include a (intuitive) factor for uncertainty – the stock level protects the maximum you expect to sell.

Stock (Inventory) and Time:

The concept of **Time Buffers** is well entrenched in Theory of Constraints thinking. As we have seen above stock can be expressed as the expected time it will take to sell it (we have 6 weeks of stock). Stock at the CCR (Capacity Constrained Resource) is expressed as the amount of processing time it will take to pass it. Expressing stock in terms of time relates better to the questions production or supply chain managers have in their minds. Having 100 units on hand might have an importance for Finance (these units have a value), but an operations manager needs to know – 'is it enough to last until I can get (or make) more?'

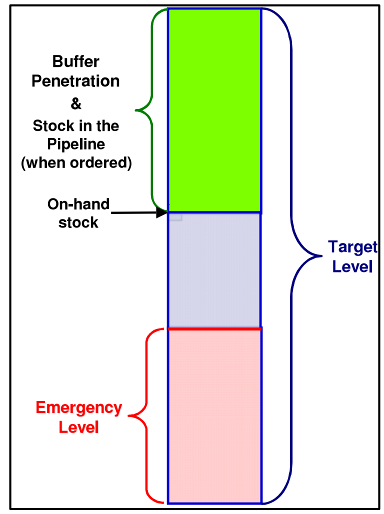
As we know, translating stock into time has its pitfalls. The numbers are based on averages and on past performance – they are forecasts. Since demand is changing all the time frequent translation into time can lead to confusion. 400 units in stock and sales of 100/week, represent a stock of 4 weeks of sales. Then we get a poor week and sell only 80. Now the same 400 units represent 5 weeks of sales – what is correct? Did inventories go up? What number do we intend to use? What do we do with these numbers?

We need to decide when and how we will use stock and time representations of our inventories.

There are 2 types of decisions for managing inventories:

- **Planning decisions**: Determining what are the appropriate stock levels:
  - When we start the inventory management process for a, say, new product, we need a starting point. Someone must make an estimate (yes, a forecast!) of how much we might sell until the product's inventory is replenished. The estimate must be factored for possible delays in replenishment. Obviously this estimate is a time related estimate and it will be quite inaccurate. We need a way to adjust inventory according to actual performance (actual sales).
  - As we move forward in time actual sales performance is used to adjust our (initial) inventory level up or down. The algorithm must ask the questions:
    - Did actual stock levels go down too much – did we lose or did we risk losing sales?
    - Did actual stock stay too high for a long period of time – are we holding way too much stock?

  These decisions are based on actual physical stock levels.
- **Execution decisions**: During production there is a need to decide on production priorities in order to meet the entire schedule. These decisions are based on actual physical stock levels.



The Finished Goods Stock Buffer – Basic Structure:

The diagram to the right describes the finished goods buffer structure. Target level (t) is the amount of stock we replenish to. Since we always replenish to the Target level the stock in the pipeline will be equal to Buffer Penetration at the time a replenishment order is placed.

The quantity on hand (On-hand stock (oh)) is what we actually have at the consumption point (in the warehouse).

The emergency level is an amount that gives us enough time to expedite replenishment should buffer penetration go that deeply into our stock. The emergency level is the equivalent of the red line in production management.

Buffer Status: Percentage penetration into the Buffer

$$BufferStatus = \frac{(t - oh) * 100}{t}$$

The main planning decision is to determine the Target level (t). The target level is the equivalent of a 'Shipping Buffer'. It protects against both demand peaks and delays from the supply system. It determines the maximum amount of stock in the system (from the supplying unit to the warehouse) if a replenishment is ordered immediately, the target level is equal to the total stock in the system. Delays in re-ordering will reduce total stock from the maximum – delays such as batching to collect a minimum batch size before placing a work-order.

As soon as our total stock goes below the target level a request for a replenishment quantity is issued. It is easy to see that responding immediately whenever total stock is below the target will reduce the need to forecast. We are simply replacing what has just been sold or consumed. The quicker or more frequently production can reliably respond to inventory consumption the less inventory we need, the shorter our (forecast) horizon and the better our inventory will reflect (near-term) demand.

Sizing the Target Level (t):

To find the appropriate size of the initial target level the following 5 parameters must be considered:

1. The average rate of product sales or consumption.
2. The variability of sales or consumption. We need to know how much demand we might have to fulfill.
3. The average replenishment time.

4. The variability of replenishment time – how reliable is our source of supply? We need to be reasonably certain that our replenishment will arrive before we have a stock-out and miss sales.
5. The service level to customers we want to (or must) achieve. The service level we want to achieve will condition what we do about demand and replenishment time variability. If we need nearly 100% certainty of delivering on time, then we need to cover much more of our variability than for clients where reliability is not much of an issue. Clients' tolerance time is usually longer for items that are slow movers – so we need less cover.

TOC solutions assume a very high level of service in order to enable the various offers like VMI or Distribution. Certainly maintaining a high service level is desirable, but can we always afford to have such a high service? The TOC emphasis is on shortening replenishment time and once we have achieved that, then the level of inventory is low enough to be affordable.

The generic rule:

The Target Level is the 'maximum' forecasted consumption within the average replenishment time, factored for the unreliability in replenishment time.

Think of 'maximum' as the amount of demand for which we are ready to invest in inventory in order to ensure no shortages even if sales are very high during a period.

Determining the Emergency Level / Red Level:

The emergency level, usually referred to as the **red level**, has two different roles. On one hand it serves for both a signal for **expediting action** and our last protection against missing a sale. It needs to give us a signal to expedite allowing enough time to complete expediting before the stock-out. On the other hand it serves as a feedback on **whether our target level is appropriate**, in other words, to check whether we should increase the level of our target. We should certainly increase our target levels of a particular product if it constantly penetrates into the red level. We don't want a huge number of expedite actions, as that would simply cause chaos in production. To prevent too many expedite actions our target stock must be high enough to get a low enough number of penetrations into the emergency/red zone. At the same time the target level must be low enough that we do get, from time to time, an expedite situation. **Never having to expedite is a sign we have too much stock on hand**.

The current practice in most TOC implementations is to have the default red level set at one third of the target level. This default proves to be good enough for both roles of the emergency level – provide a timely expediting signal and to provide feedback on the adequacy of the target level. There is no real need to optimize the 1/3 ratio, as changing it will not bring much added value.

Buffer Management – Feedback for Planning Decisions:

The TOC buffers are divided into three, usually equal, zones. The last third of the buffer is the red zone (the emergency zone). The middle zone is called the yellow zone and the upper zone is green. As the target level, the buffer in MTS, consists of both the on-hand stock and the stock in the pipeline, the regular expectations are that the on-hand stock will be in the Yellow Zone, not too full and not too empty.

When a product penetrates regularly into the emergency level it signal's more than just the need to expedite replenishment from production. It is very important that we recognize what is causing the signals and why.

The reason for frequent or too much penetration into the emergency/ red zone could be significant demand growth for one or more products. It could be some other problem that has caused replenishment times to increase. Or, we could be seeing increasing variability of demand or replenishment time. Whatever the reason we need to track these and use Pareto to select the most important reasons for emergencies and then find solutions to the underlying reasons. (This is a nice example of using TOC to provide information for a 6-sigma or Lean improvement project).

There can also be too few (or no) incursions into the red zone – or on-hand inventory never goes (far) below the green zone. The green zone is the highest 1/3 part of the target level. When on hand stock is above 66% of the target level it means the stock is in the green zone and the buffer status is below 33%. Being in the green zone for a relatively long period of time is almost as important as too many incursions into the red. Too many incursions indicates we are at risk of hurting delivery performance to our customers, too long a time in the green indicates we have too much protection – too much stock that we don't really need. Again we should discover why a product has too much inventory – a competitor has launched a new and better product etc. The analysis should lead to the necessary corrective action – which could be to just lower the target so that we do get a few incursions into the red (emergency) zone.

Eli Goldratt's Target Level Management Rules for Distribution Environments:

Since we already have a set of rules for distribution systems, why not use them for our MTS solution?

1. When a penetration into the 'red zone' occurs monitor how deep the penetration is. If the penetration is *too deep* or it persists for *too long* then the target level should be increased. What is "too deep" or "too long" depends on the required service level for your business and the amount of risk for a stock-out you are willing to run. Certainly, if the penetration consumes the entire (or more) red zone the target needs to be increased. Dr. Goldratt's recommendation is to accumulate the daily penetration into the 'red zone' during the average replenishment time. If the accumulation is equal or above the size of the 'red zone' then increase the target level by 33% (and by that action also the size of the emergency level).
2. The recommendation for decreasing the target is also by 33% - after actual inventory has remained in the green for a whole replenishment time period.

The recommendation of a 33% adjustment ensures a quick response to changing (increasing) demand.

3. Every time you increase the target level WAIT for a replenishment cycle before starting to check again. Only after a second replenishment cycle should you make another change up or down.
4. Every time you decrease the target level, then at first the on hand stock is ABOVE the new lower target level (assuming you decreased the buffer by 33% and your green zone is also 33%). Wait for the inventory to get down below the new green level and only then start to check again for the conditions to decrease the target level further.

In production (on the shop floor) the previous rules should be used as a general recommendation, but with a lot of care. A 33% increase in the target level can cause the release of a relatively large batch that can significantly delay other production and cause a cascade of increases in target levels and many large batches. We certainly don't want to create chaos in production. **If the load on the CCR is near to its capacity limit, then increasing the stock buffers (the target levels) can cause long delays**, causing deep penetrations into the red zone of other products. Increasing target levels will then be counter-productive– they would dramatically aggravate the situation. When many products go into the red you also have a dangerous situation that you can very easily make much worse by increasing all target levels. A good principle to follow is to be very careful and use your head whenever you get signals to increase targets. Further, when you do increase the target level of a product, it is advisable to split this increase into two or more batches in order to not cause too long delays to other orders in production.

Releasing Production Orders for Stock Replenishment:

When the total quantity of the items of a certain product, both in the warehouse plus the open production orders for that product are less than the target level, a request to produce the difference is issued. However, capacity considerations may prevent the immediate release of all requests. Thus we need a procedure to schedule daily requests based on their relative priority.

Production orders are released into production based on the requests relative buffer status. The formula for buffer status of the request is:

$$BufferStatus = \frac{(t - oh - po) * 100}{t}$$

t = target level

oh = on hand stock

po = older (in progress) open orders for the specific product

The buffer status for every production order is calculated based on the quantity downstream to the order itself. Thus both the on hand stock and the quantity of older in progress orders are considered relative to the size of the target level. The requests with the highest %age number have the greatest (relative) penetration and therefore get the highest priority in production. All requests are ranked with the highest Buffer Status at the top. The top of the list has the highest priority for production.

It makes no sense to release too much work into production – this will only lead to confusion and in any case the CCR can only do so much work within a period of time. The total planned load[1] of released orders, should be limited by a "CCR Buffer Time". This "CCR Buffer Time" is the equivalent of the CCR time-buffer in an MTO environment. The "CCR Buffer Time" should be about the same as the required replenishment time.

We said there is no point in releasing stock replenishment orders today if the current WIP from release to the CCR (the planned load) is equal to or greater than the "CCR Buffer Time". By holding back these orders and waiting a day, their priority (and the quantity to be released) will change – increase or decrease in relative terms. Items that have penetrated into the emergency zone should be released immediately. Their high priority, given by their relative Buffer Status will make sure they overtake older orders with a lower Buffer Status.

It is important to recognize that in a MTS environment **a CCR must not (cannot) be a bottleneck for long!** A long-term bottleneck in an MTS environment will result in stock levels going toward zero and all that good work to achieve near perfect availability of product at customers will be wasted. A bottleneck, if it looks like it will persist, must be dealt with. Thus, monitoring the planned load on the CCR, and ensuring it won't exceed the limit of "CCR Buffer Time" is necessary to smoothen the load in the shop floor, while being responsive to the replenishment requests. In addition it signals when the load starts to be too high and a capacity increase is absolutely necessary.

Capacity Management

An emerging bottleneck, especially in an MTS environment with commitment to maintaining availability, might disrupt the performance of the whole shop floor. In an MTO environment we know how to promise delivery dates based on the planned load plus 1/2 of the buffer. In an MTS environment it is impossible to control the demand in this way because not all the firm production requirements, which should be part of the planned load, have been included in the planned load. In MTS we normally include in the planned load only production orders already released to production. We need to recognize the entire capacity requirements, so that management can take the necessary actions to either increase capacity or to find a way to restrain market demand.

As suggested above the flow of released orders is monitored and controlled by the planned load to not exceed the CCR-Buffer-Time. However, as a general capacity management tool this is not enough. It does not include the "should-be" released orders that are for the moment choked from release. Our suggestion is to maintain a planned load that includes ALL requests (or suggestions) for release to production. In other

[1] The term Planned Load is defined and explained in a previous white paper by Eli Schragenheim: Using SDBR in Rapid Response Projects. The name of the file is SDBR in RR Projects.

words, for every stocked product include a dummy order with the total required replenishment quantity. This "total planned load" represents the real load on the CCR. When the total-planned-load gets longer and longer, implying a longer replenishment time than we can tolerate, management must either add capacity or restrain demand.

How can we know ahead of time when to take effective actions? We watch the pace at which planned load increases. The rule is to play safe. When it appears that planned load is approaching about 80% of the longest tolerable replenishment time actions should be taken.

Peak-Time Behaviour:

Since demand will always fluctuate and if demand peaks are high enough our CCR will, from time to time, become a temporary bottleneck. Sometimes we can anticipate a peak – since usually we know when promotions take place, or we know the businesses seasonality etc. If we know, then target levels should be increased in advance of the peak and restored to their original levels towards the end of the peak. During a peak we expect stock levels to go down – whether or not we anticipated it.

During a peak actual production batch sizes will increase because we are selling more during the time it takes to release an order and the time-to-release itself gets longer at a peak period.

Off Peak-Time Behaviour:

Most of the time demand will be off peak and there is no problem of capacity at the CCR – it can (easily) accept all production requests. Batch sizes will be relatively small because they will be made up of only a few days of sales. Load on the CCR should be monitored to make sure sufficient capacity remains despite the additional set-ups that will occur.

More on batching in MTS the TOC way:

We have already described the evils of batching. However, in too many cases batching cannot be eliminated. When setups exist they have an impact on capacity. We have already seen that in peak periods we naturally get longer batches. But, sometimes we need to batch even in off-peak periods. Sometimes we are forced to implement certain minimum batches to face real capacity limitations.

Our question is: should the minimum batch be part of the target level, or does it come on top. To illustrate this question, let's use an example.

Suppose a product's target level is 100 units and we are forced to produce in minimum batches of 30 units. If the total inventory in the system is 99 – should we create a production order of 30 units, or should we wait until the total inventory has decreased to 70, and then release the minimum batch production order?

Note that in the latter case our actual protection is only 70 units. Since we usually define the replenishment level as our buffer we must replenish it whenever our total inventory is below 100. The unavoidable result is many times we'll have more than 100 units in our system (on hand plus the production pipeline). This is the price we must pay for having significant setups that significantly impact our capacity.

The rule is: Target level is defined as the net buffer. Minimum batches, when required, <u>come on top of it</u>, resulting in more inventory than necessary to protect the market.

<u>Managing Priorities in a Mixed MTS & MTO Environment</u>:

MTO and MTS orders, both use 'Buffer Status' to provide prioritization. The MTO Buffer Status is based on penetration into the **time buffer** to produce the product while the MTS Buffer Status is based on the penetration into the **stock buffer**. In an MTO environment the Buffer Status gives us information about the likelihood of meeting the due date. In an MTS environment Buffer Status tells us something about the likelihood of product availability in the near future (we need to recognize that the Buffer Status in a MTS environment can change suddenly). Lets look at an example of a mixed MTS & MTO environment.

> ***<u>Example</u>****: On March 1st we have 3 orders at our CCR. The question is – what should be the priority for these 3 orders?*
>
> - ***<u>Order 1</u>:*** *A large MTO order for product P3 due on March 18th. Our Shipping Buffer is 2 weeks or 14 days.*
> - ***<u>Order 2</u>:*** *A MTS order for product P1. 1000 units is the target level for this product. The replenishment order is for 500 units and we have 430 units on hand.*
> - ***<u>Order 3</u>:*** *A MTS order for product P2. 300 units is the target level for this product. The replenishment order is for 100 units and we have 106 units on hand.*
>
> *For both MTS and MTO products the emergency zone is 30% of the buffer or at 70% and more penetration into the buffer.*
>
> ***<u>Question</u>:*** *Which of the 3 orders should the CCR produce first?*

$$BufferStatus = \frac{(t - oh) * 100}{t}$$

The equation to the right is for the MTS environment. The one for MTO is analogous.

So, the status of the large MTO order is very comfortable since we have 2.5 weeks to the due date and the Buffer is 2 weeks long. ***The Buffer Status is negative*** *and there is no urgency for this order whatsoever.*

The Buffer Status of P1 is: $\frac{(1000-430)*100}{570} = 57\% penetration$

The Buffer Status of P2 is: $\frac{(300-106)*100}{300} = 65\% penetration$

So P2 is the most urgent, P1 next and the MTO order is not urgent at all.

NOW SUPPOSE *that the same 3 orders are still at the CCR on March 10, 8-days before the due date of the MTO order. What is the priority situation now?*

- *The MTO order is now due in 8 days (out a 14 day shipping buffer).*
- *There are now 200 units of P1 on hand.*
- *There are 102 units of P2 on hand*

Clearly priorities have shifted and P1 is now clearly the one with the highest priority. P1 has even crossed the red line and consumed some of the emergency zone. It should be expedited as soon as possible.

What Products should be made to Stock?

Having stock available in your warehouse is a big advantage. You can satisfy a customer's need (or want) without delay – lead-time is made up of order handling and transportation. As long as you can guarantee product availability it follows that clients will favour your store (business) over others that have a lower level of availability. Your sales, turnover and profits should go up.

Having product in stock makes sense for those products that are standard and fast movers. Fluctuations of demand are relatively lower compared to slow movers so the risk of holding too much stock is also lower, as is the risk of obsolescence. Operation's job is to make sure availability of the chosen standard and fast moving MTS products is better than that of competition – good enough that the company can guarantee availability if it so chooses.

Slow moving products are another story. By their very nature demand is much more uncertain making the risk of surplus and obsolete stock significantly higher. It is much better to manage slow movers as make-to-order products – as long as they can be produced within a short enough lead-time. If the tolerance time of clients (the time they are willing to wait for a product) is longer than the delivery lead-time the company can make these products to order. Using S-DBR and Buffer Management will often make a short enough lead time possible – sometimes with the help of an intermediate, more aggregated, semi-finished stock that can supply several to many end-products.

Mixing MTO and MTS Policies for the Same Product:

Should we (sometimes) maintain a safety stock for products that we make to order? We might want to hold some stock of MTO products because we know we sometimes get very urgent requests AND we can get very nice premiums if we can meet the timing of these requests. Sometimes we might have extreme pressure to quote very short lead-times (the client has a lot of power and/or a competitor is offering such lead-times). Such pressure may be forcing us away from an MTO mode towards MTS. During such times holding stock for MTO products is a temporary measure until the MTS process is properly in place.

Whenever we hold stock for MTO products we are effectively dividing our shipping buffer between both **STOCK** and **TIME** buffers. This will make it much more difficult for us to determine whether we are really in the red and by how much. (An order might be in the red of the time buffer, and not be in the overall red at all because there is plenty of stock available in inventory.) We still need to know how much of available stock is covering earlier orders before we can decide whether or not our production order is truly in the red and needs expediting!

Consider also that by splitting between TIME and STOCK buffers, the red zone of the time buffer will not provide enough time to expedite an order through production. It can be a very confusing situation when we keep both time and stock buffers.

- We don't really know how to define the 'right' red zone.
- It is not clear or easy to determine what the buffer status of the combined buffer is.
- So, we recommend the combination only for the transition time when moving from an MTO to an MTS mode of operation.

Firm Future MTS Orders for a Stocked Product:

It will happen that some customers place firm orders with due dates sometime well beyond our horizon for operations. These are FIRM orders with firm due dates – we can count on them. Should we ignore this known firm demand and simply ship it from our normal MTS stock when the time comes? Or should we allocate (e.g. reserve) stock for this future shipment and reduce it from our current stock? If we do allocate (reserve) stock for such an order, when should the reservation take place? Would you expedite production of a product when stock is reserved for an order due for delivery in two months? OR, should we generate a special MTO order and would we expedite such an order when we have enough finished product already sitting in stock?

Clearly we want to think twice before producing for an order when we have plenty of stock in our warehouse. If we start this sort of practice we may well end up blocking something else – much more urgent and/or we may end up with more inventory than we really need.

After reviewing all options and reminding ourselves of the policy of applying only SIMPLE solutions, we recommend the following:

- If the firm future order represents a relatively small quantity (less than 10% relative to our buffer size), then such orders consume from regular stock without any special treatment – no reservations of stock.
- If the firm future order is not that small, then it should be fulfilled as an MTO as a “separate SKU” – to avoid the confusion of 2 buffers.

Stocked Components of End Products:

Many times it makes sense to stock components that are common to several (or many) end products. This situation is likely to occur in “V” or “T” production environments. The advantages of stocking components are:

- Significant reduction in the production lead-time of many end products. This advantage can be achieved if all components for assembly or all the relatively long lead-time components are stocked. Of course, the amount by which the product lead-time is shortened will have major impact on the decision to manage the components for availability (by making them stocked items).
- If some operations for component production have long set-up times and therefore these resources have insufficient capacity to make-to-order – then MTS is a good way to deal with the problem, because making them to stock allows using larger batches without compromising the lead time of the end items.

If a component is a stocked item, then we assume availability of that component for any production order of the end item. Making this assumption requires components to be very nearly 100% available for any possible end-product production order. The material release schedule of an end product has to include release dates for these MTS components.

Which Intermediate Item (Component) or Product should be stocked?

There should be a clear gain to your business for transitioning it to MTS. Since moving to an MTS environment will most likely increase both Investment (I) and Operating Expenses (OE) it seems clear that deciding for MTS must help increase sales volumes and/or allow us to charge higher prices (or possibly protect our existing business from competitors). You can achieve value with MTS for customers that value short and reliable lead-times. Your MTS operation must make a real (significant) cut in lead-time and be very reliable for an effective impact on customers. **All** components of such a finished product must either be in stock or be able to be produced from raw materials in a (very) short time.

An important part of the job is to identify the right customers and ensure that they recognize and honour the value of short reliable lead-times by more purchases, higher prices and greater loyalty to your business.

The investment in an MTS operation must be financially appropriate – added inventories should not be too large. In other words, the variability of demand must not be so large that we end up with huge inventories with too high a probability of no sale.

Sometimes the benefit from stocking components will be largely internal – by freeing the capacity of a resource that would otherwise be a bottleneck. Typically very long set-up times and dependent set-up situations are candidates for component MTS. Many times you will find such cases at (or near) the bottom of a V plant. Then it often makes sense to hold stock one or two layers up from the bottom – considering the required lead-time and the competitive edge you can get and the amount of inventory you need to meet the necessary lead-times.

Managing Transitions:

Before transitioning from one environment to another (MTO to MTS for instance) you need a carefully developed transition plan. All the predicted effects during and after a transition need to be clearly verbalized in order to prepare your plan. You need to evaluate the planned transition time period – does it make sense or will it lead to problems. Basically you are looking for the risks in the transition and how, when and where to put in place a plan that limits risks and gets the transition done within a reasonable amount of time (a transition period is never a comfortable time). The plan should also have checkpoints to monitor the effects of the transition; and to check that the expected performance is actually being achieved.

Managing the Transition from MTO to MTS:

Just remember that to do this move you must build up MTS inventory while, at the same time, continuing to deliver the regular demand on time and in full. Sufficient excess capacity in the production system is a necessary condition to be able to build the required MTS stock within a reasonable amount of time.

Before you have the necessary stock for MTS in place DO NOT make the change! If you do, the target stock level you define will cause an immediate production-order release of the target level quantity. In the majority of cases it is recommended that you approach the target level slowly (using excess capacity to expedite should that become necessary). For similar reasons only a few items should be transitioned during the same time period. You need to avoid causing a sudden huge over-load on your CCR.

Managing the Transition from MTS to MTO:

As you move away from MTS to MTO, deliveries will of course continue to be made from stock until this is drained from your warehouse. The transition to zero inventories is critical. You are moving from a situation where it is relatively easy to achieve and maintain a good due date performance – because you have unassigned stock you can use for any order. When stock levels have reached zero you no longer have this cushion – you are relying on your (shipping) buffer times to make sure you continue to deliver on time. During the transition, as you reach zero stock, you need to monitor your time-buffers and their sufficiency carefully.

(When moving from MRP to TOC you have a similar situation. MRP environments will often have safety stocks in their system that should be drained. As you do that you have a similar situation to the above.)

Dealing with Seasonality

Seasonality means regularly repeated changes to market demand depending on a period in time. The most common seasonality is within a year, more demand at a certain season and less in the other seasons. We might also find seasonality within one month (like in the end of the month) and within a week (more sales during the weekend).

Annual seasonality is certainly relevant to MTS in SDBR as it means an expected increase or decrease of the target level. The dilemma is whether to include a change in the target level based on our anticipation for the seasonal effect, or let buffer management identify the change in the trend and only then change the target. The advantage of buffer management is that it relies on a real change in the demand, as any anticipation is based on a forecast and thus might be wrong.

We have two problems in relying on buffer management to identify the seasonal effect: One is that the signal could be too late and, depending on the net impact of the seasonal effect, we might lose sales. The other is that many products might be impacted at the same time by a seasonal effect. Then capacity limitations could cause very serious shortages. We had better prepare for seasonal peaks early enough to have enough time to build the new target levels.

It is the latter effect that forces us to prefer being active and take an anticipated seasonal peak seriously – meaning change the target level early enough to be able to deal with capacity limitations. It calls for a detailed production/capacity planning where the targets are slowly (could be in several steps) increased until they reach the target level for the season. At such a time we should not allow dynamic buffer management to reduce target levels, (they are higher than what they should be at that time, in anticipation of the seasonal peak).

Within a season one should also anticipate the end of a peak, and to reduce target levels, even though buffer management does not yet show any signal of reduced demand. Again we recommend doing it in steps aiming to reach an appropriate low target level by the end of the season.

If replenishment is very fast, monthly and weekly seasonality must also be addressed as well. That means finding ways to implement repeated changes in target levels based on seasonality within relatively short time periods.

The Special Cases: Huge , but Short Peaks of Demand

Certain holidays, or promotions, cause peaks of high demand followed by periods of very low demand. When the peak itself is longer than the replenishment time, the previous paragraph describes the appropriate way to handle it.

But, if the peak is huge, and lasts for very only a short period of time: equal or shorter than the replenishment time, then the only option is to start with enough inventory to cover the whole peak. Here there is no real contribution of buffer management. The only option is to forecast the peak demand and make an intelligent decision. When we say forecast, we had better note that “average demand” is NOT what we look for in intelligent decision-making. What we have is a forecast of a reasonable range: The minimum and maximum demand within the peak. Then, the decision should be made by product deciding whether to risk shortages or surpluses of inventory. This is a typical decision on uncertainty where one needs to decide which damage is more important to be protected from.