Process flow industry—scheduling and control using theory of constraints

E. SCHRAGENHEIM, J. COX† and B. RONEN

This paper discusses the applicability of the Theory of Constraints (TOC) in the process flow industry and compares TOC to the current practice. The Drum-Buffer-Rope methodology is altered to meet the needs of the process flow environment. Guidelines for the strategic location of and determination of the reduced WIP inventory levels are provided.

1. Introduction

The process flow industry differs from other manufacturing types. Their scheduling practices, as modelled by Taylor and Bolander (1991), are certainly distinct from scheduling methods used in other types of manufacturing. The process flow industry represents a substantial part of all industry; however, very little research has been conducted on its scheduling and production problems. New scheduling methodologies, such as kanban and drum-buffer-rope, have penetrated the industrial world in the last 10–15 years but have not had much impact on the process flow industry. The process flow industry is certainly a very capital intensive industry and production scheduling does impact plant design, production practices and plant performance financially.

This paper argues that the basic scheduling of the Drum-Buffer-Rope (DBR) methodology (Goldratt and Fox 1986 and Schragenheim and Ronen 1990) can and should be applied to the process flow industry. The model of Taylor and Bolander (1991) will serve as an illustration of the current scheduling methods used in the process flow industry. The concept of DBR will then be deduced and compared to the current scheduling methods using the global performance measurements of throughput, inventory and operating expenses. The use of DBR reduces inventory levels and improves responsiveness while maintaining potential sales and without incurring any additional operating expenses.

2. The Taylor and Bolander model

Prior to describing the underlying principles of the Taylor and Bolander model for scheduling in the process flow industry, a few definitions related to process structures are required. A process unit (P) performs a basic manufacturing step or operation. A stage is defined as a series of process units (for example $P1_0$ through $P1_{k0}$). Stages are identified by inventories being located both upstream (I1) and downstream (I2) of the

Revision received May 1993.

† To whom correspondence should be addressed.

Department of Management, University of Georgia, Brooks Hall 419, Athens, GA 30602-6256, USA.

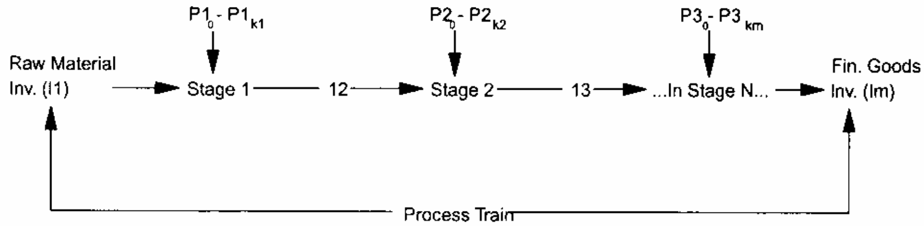

Figure 1. The relationship of processes ($P1_0 - P1_{k1}$) to stages to process train.

stage. Stages are organized into process trains. A process train is a fixed, sequential series of stages in which a family of products is produced. Figure 1 illustrates the relationship of these terms to each other.

Two scheduling terms are used by Taylor and Bolander. Material dominated scheduling (MDS) is defined as a scheduling technique that first examines the quantity and the sequence of the parts needed from the current levels of the surrounding inventories for a stage and then checks the feasibility of producing that schedule with the stage capabilities. Processor-dominated scheduling (PDS) starts with an optimized schedule for best utilization of the stage's resources and then checks the feasibility of both downstream and upstream inventories for supporting that schedule.

According to the Taylor and Bolander model a finite loaded schedule is produced for each stage of a process train using either PDS or MDS. Scheduling a stage is equivalent to scheduling a single resource. Taylor and Bolander (1991) state 'Separating different stages with inventory allows these stages to be managed somewhat independently'. The major difference between the Taylor and Bolander model and known shop floor scheduling techniques is that a stage is independently scheduled over the entire time horizon without simultaneously scheduling or considering the upstream and downstream stages. Once a schedule for a stage has been fully constructed, the surrounding inventory levels are checked, considering now the impact of the previously scheduled stages. If either the minimum or maximum inventory levels are violated—the schedules are corrected. The underlying assumption of the Taylor and Bolander model is that such violations are infrequent, otherwise the scheduling process will be slow and cumbersome.

3. Role of inventory in the process industry

In the Taylor and Bolander model for the process flow industry, WIP inventory is used in order to simplify the scheduling efforts. The high levels of inventory which are needed to decouple adjacent stages are contrary to the notion of 'zero inventory'. The evils of high levels of WIP inventory are widely accepted today. Both JIT and TOC have shown that excess inventory affects both operating expenses and sales (Goldratt and Fox 1986). Some level of WIP is, however, certainly necessary to ensure meeting all the potential market. The effective placement of WIP in discrete manufacturing is discussed by Finch and Cox (1989).

When and how much WIP is required to support production in a process flow industry? A different approach to scheduling might produce better results than traditional approaches. In comparing different scheduling approaches the global measurements of Throughput, Inventory and Operating Expenses will be used. Throughput (T) is defined, by Goldratt and Fox (1986), as the rate at which the system generates money. It is interpreted as sales minus the raw material cost (which is the money that just passes through the system rather than being generated by it). Inventory (I) is defined as all the money the system invests in purchasing things the system intends to sell. This means that only the actual purchasing expenses are considered—without any added value. Operating Expenses (OE) are all the money the system spends in turning Inventory into Throughput. Goldratt and Fox (1986) have shown that the three measurements are equivalent the net profit and return on investment measurements. In comparing scheduling approaches, a superior approach to the Taylor and Bolander model is one that produces, at least, the same throughput with the same operating expenses, but uses less inventory.

3.1. *Finished goods inventory*

The need for finished goods inventory exists when the customer tolerance time (CTT), defined as the longest delivery time the customer will accept, is less that the production lead time (PLT). If CTT is greater that PLT it is better to use a 'make for order' approach. We suggest that in other situations the level of the finished goods inventory should be based on three factors:

(1) The average consumption for the time frame that equals the difference between PLT and CTT. The assumption here is that PLT is greater than CTT otherwise finished good inventory should be zero.
(2) Fluctuations in the market demand exist. When the market demand increases, the organization might lose throughput. Safety stock should be provided to protect against losing throughput due to a temporary increase in the market demand.
(3) The fluctuations in the PLT. Production lead time is, in the vast majority of the cases, highly variable because of a variety of reasons like: downtime, bad quality, absenteeism and fluctuations in the supplier lead times. Either safety stock or consideration of more than average PLT (safety time) should be employed here to protect against the possible loss of sales due to production lead time fluctuations.

In our view it is preferable to treat the combination of the three factors together rather than separately. The main reason is that setting a specific inventory level to monitor the combination is simpler—hence more practical—than monitoring each factor separately. Another reason is that the three factors may not be independent so treating each one of them separately may result in a less than satisfactory overall level of finished goods.

Let's define the minimum level as the total level of finished goods inventory that is supposed to protect the sales while a new batch of the specific product is in process (demand during lead time). We include in this term all the necessary safety stocks. It is a strategic decision, based on marketing and production considerations to decide upon the level of protection to be set as the minimum level.

3.2. *Raw material inventory*

Raw material inventory levels should be based on similar factors. The frequency of deliveries (or the vendor lead tome when no regular delivery times exist), and the fluctuations of both market demand and vendor reliability are useful in determining the minimum level of raw materials inventory. The cost of a raw material may also impact the decision.

3.3 *Work in process inventory*

Work in process inventory (WIP), according to Taylor and Bolander, is planned between stages. Suppose the planned inventory between two stages is cancelled—would the throughput level of the system be affected? When stage N and stage $N+1$ merge by cancelling the planned WIP inventory between them, stage $N+1$ is no longer independent. The actual sequence of its operations is dependent on stage Ns sequence. It is not necessary that the sequences at $N+1$ be identical to that of stage N. There may still be actual WIP between stage N and $N+1$. The existence of the WIP stems from the actual operation times and sequences of the two stages. The word 'actual' means that in addition to any schedule time estimates there are statistical fluctuations. In this case one should consider letting stage $N+1$ operate according to the availability of inventory between the stages rather than to operate according to a schedule (plan). The following discussion supports the idea of eliminating most, if not all, of the planned inventories between stages.

4. DBR concept in a process flow environment

The Drum-Buffer-Rope (DBR) methodology (Goldratt 1990 and Schragenheim and Ronen 1990) strives to let the market demand determine the production requirements. The basic vehicle, according to DBR, by which the market demand may impact actual production requirements is by tying the first production stage directly to the market demand needs. This is the rope segment of DBR.

The idea, illustrated for the two stages in Fig. 2, is to translate the market demand requirements, in quantity and time for every batch, into raw material requirements. The translation takes into account the current WIP that may reside between and in the stages. Thus the net requirements for the raw material release are dictated by the market demand. The succeeding stages will have to confine their sequencing to the available material only—according to the actual needs of the market. The main 'pull' algorithm ties the market demand to the raw material stockroom release.

Do we lose throughput using this idea? Throughput may be lost if stage $N+1$ is not utilized properly because of its dependence on stage N. However, when the resources at stage $N+1$ have excess capacity, as in most cases, no throughput will be lost because $N+1$'s excess capacity is enabling its resources to overcome any delay caused by unavailability of material or extra setups. In other words stage $N+1$ may not be fully active or 'efficient' but still supply all the current market demand.

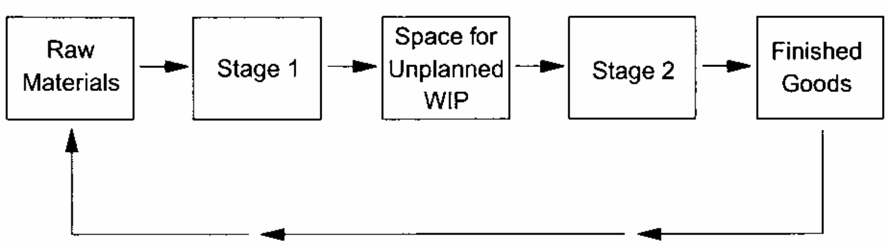

Figure 2. Market demand is translated into raw material requirements.

4.1. *Capacity considerations*

When capacity is expensive, management is tempted to treat it like a scarce resource. However, expensive machinery is often very fast, so it may not be the crucial limiting factor. The first step is to identify the real factor(s) that limits the performance of the system, the constraints of the system (Goldratt 1990).

It is implied by Taylor and Bolander (1991) that no resource sharing is allowed between different stages and different trains. Each train is independent of the other trains. Hence the weakest link in the train is the one that limits the throughput of this particular family of products. The weakest link might be the market demand (meaning the train can produce more output) or a capacity constraint resource (CCR or a bottleneck). In any case, the vast majority of the resources do have some excess capacity.

Suppose the constraint lies in stage N which is neither the raw materials availability nor the market demand. This means that a certain resource in stage N lacks capacity relative to the market demand and hence constrains the production process. Such a resource cannot support WIP inventories requirements downstream including the finished good inventory. From this description, the following conclusion is made. When we do have enough capacity to maintain inventories it means that no bottleneck or capacity constraint exists in any stage. Therefore, whenever planned finished goods inventories are maintained—the market demand is the only constraint. This description is correct for the vast majority of actual companies. Even when the market demand is the constraint we may have some capacity considerations in scheduling.

4.2. *Impact of setup times*

Suppose that finished goods inventories have to be maintained (CTT < PLT) to support market demand. The distribution of the lead time directly impacts the required level of the finished goods inventories. In order to minimize the production lead time the lot sizes should be reduced and the number of setups will be proportionally increased until a constraint develops. This constraint is the resource located within a certain stage, for which its excess capacity has been exhausted by the added setup time. This resource determines the minimum lot size and thus can be defined as an inventory constraint (it limits the further reduction of the inventory levels). The minimum lot size can be implemented by fixing a maximum finished goods inventory level so:

$$\text{Max} = \text{Min} + \text{minimum\_lot\_size}$$

The Min has been defined above as the market protection level for the three safety factors previously defined. The Max now completes the Max-Min policy of managing the finished goods inventory. Defined in this way, any time the finished good level for a specific product drops below the Min level (which serves as the order level) a production request is generated for an amount that is greater than on equal to the minimum lot size.

4.3. *Scheduling considerations*

The inventory constraint, the most loaded resource, can still gain additional capacity by scheduling larger lots than the minimum lot size. It is rather difficult to reduce the lots exactly to fit a 100% load on the inventory constraint. Larger lots should be used to gain some excess capacity (by converting set-up time to processing time) for the inventory constraint. However, the actual utilization of the inventory constraint should be carefully monitored. It does not have enough excess capacity, given the reduced lot sizes, to bear too many delays without causing damage to throughput. When the delay is relatively small, the schedulers should increase the lot sizes, temporarily, in order to reduce the delay/lost capacity. Longer delays may cause the finished goods inventory to be exhausted resulting in lost throughput. The inventory constraint should be protected against long delays such as a lack of material. The inventory constraint should be treated as a capacity constraint in the sense of the DBR technique (Goldratt 1990, Schragenheim and Ronen 1990). The DBR methodology avoids wasting the constraint's capacity by preparing a detailed schedule for the relevant stage (which contains the constraint) based on the market demand and then provides a time buffer, called the constraint buffer, to protect the constraint schedule from disruption. The concept of the time buffer means to schedule the release of raw material enough time ahead of every scheduled instruction start time for the constrained stage. In the DBR methodology the term 'enough time ahead' refers to the amount of time that ensures that in 95–99% of the cases the material will reach the constraining resource before the constraint needs the material. In other words, it is the average time for material to reach the constraint plus 2 to 3 SDs. In the particular environment at hand, where the inventory constraint has some excess capacity, we suggest that a protection of 70% is enough.

In Fig. 3, the example demonstrates the DBR methodology in a process flow environment. Analysis of the market demand for the production line and the constraint capabilities for producing these items is used to determine the lot size requirements given the Min level. The determination of the Min level may involve a forecast (or possibly the customer orders at hand) to determine examining the demand during lead time. In the DBR terminology the schedule for the constraint is called the Drum. The resulting raw material release schedule is called the Rope. In the example above, the Drum consist of a sequence of jobs to be run at stage N which contains the constraint. The timing for material release is based on the projected start times for stage N determined by maintaining a Gantt chart of the constraint. Since material is released earlier than the average time to reach stage N, on average the material will wait at the constraint's stage thus creating inventory in front of stage N.

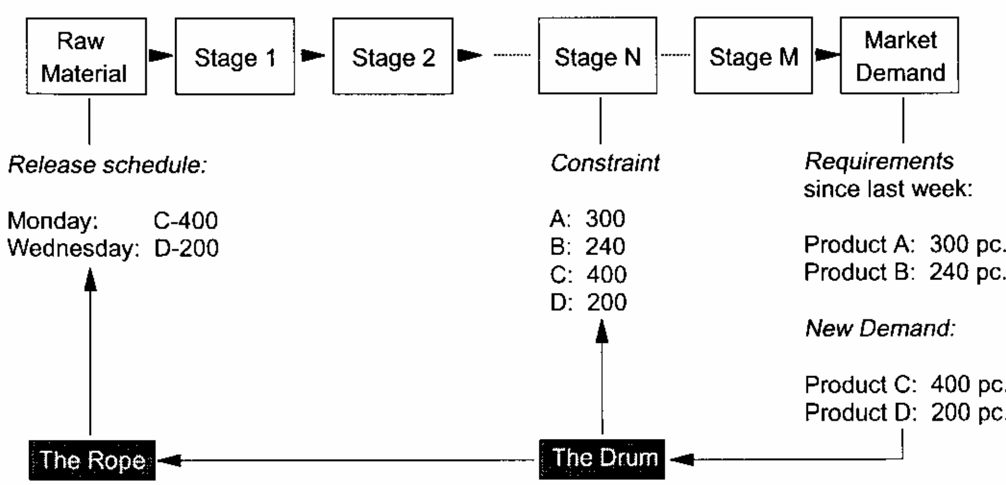

Figure 3. Relationship of Drum, Rope, Market demand and Raw material.

Goldratt (1990, chapters 31–33) develops a detailed algorithm for creating the Drum schedule in accordance with a set of due dates. The instructions for the Drum are ideally scheduled 'shipping buffer' times prior to the due dates. The shipping buffer is defined as the worst-case-lead-time from the constraint to the finished goods area. This definition suits the make-to-order environment. In contrast, in the make-to-stock case, the shipping buffer is represented by the Min finished goods inventory level. The Drum can be scheduled any time prior to the exhaustion of the constraint buffer. The actual sequence of products across the constraint depends on the characteristics of the inventory constraint. This means that a finite forward loading algorithm, for the inventory constraint only, is better suited to a make-to-stock system than Goldratt's algorithm.

What about the rest of the production process or process train? The Rope, the raw material schedule, establishes a practical schedule for these non-constraints. The advantage of DBR is that when delays and downtime occur, the foremen of the non-constraints are not bound by a definite schedule. If product A, for instance, is late but product B (done on another unit of the previous resource) is somehow ready, the foreman should start on B before A. What is the impact? The re-sequencing will work fine unless significant dependent setup considerations arise. This case will be discussed later. The following guidelines summarize a simple DBR scheduling methodology for the process flow industry, assuming a make-to-stock environment. Prerequisite: identify the most loaded resource; the one that dictates the minimal lot size. Determine the minimum lot size so that this inventory constraint resource will be fully loaded except for provisions for downtime and inaccuracies in the data. Fix the Max and Min levels so that the Min will provide adequate protection for the combination of the three safety factors. The Min level should be changed when the forecast points to a possible significant change. The Max should be set to the Min plus the minimum lot size.

The DBR loop: determine the production requirements for the next period. The production requirements are comprised of those products for which the finished goods level plus the previous unfinished requirements, is lower than the Min level. The new requirements are the replenishment amount to the Max level.

Schedule the capacity available at the inventory constraint resource. Use a finite forward scheduling algorithm.

Develop the raw material release schedule based on the projected start times for the constraint resource minus the estimated constraint buffer time. Goldratt suggests a modified algorithm for computerized systems (Goldratt 1990, chapter 35).

Implement the raw material schedule. Do not release material earlier than the release schedule even when the first operation is idle. Notice that an appropriate constraint protection has already been considered. Material maintains its greatest flexibility in raw material form. Also, releasing material too early creates the impression of fully loaded operations.

5. Several complications

The simplicity of the DBR methodology can withstand complications imposed by the environment. Changing the buffer sizes, time-buffer for the inventory constraint and the Min finished goods inventory level, provide adequate protection for most complexities. However, management should investigate changing the design of the production process through JIT techniques in order to meet the simple concepts of DBR. This approach will reduce the inventory levels and may indirectly increase the Throughput level as well.

5.1. *Dependent setup considerations*

Dependent setups exist in many process flow industries. This means that the setup time for a certain product is dependent on the previous product processed. In most cases, an overall preferred sequence can be applied. The detailed schedule of the constraint will have to conform to the preferred sequence. The Rope (raw material release) will reflect the preferred sequence. Non-constraints that are very sensitive to the sequence will have to maintain this preferred sequence even if fluctuations have changed the natural sequence. This will impact the size of the time buffers used.

Cases may exist where an overall preferred schedule is not practical because several resources have quite different preferred schedule sequences. A problem might arise only if a non-constraint would turn into a bottleneck when the overall sequence, which differs from its own preferred sequence, will be forced on it. The non-constraint should then have guidelines on which sequence to follow. This means that the non-constraint might have to wait for the 'right' material to show up, even when material for other jobs are available to it. Such problematic non-constraint resources will have a considerable impact on the WIP inventory level. In the extreme case, the lead time to the constraint area, and from the constraint to the finished goods, will have to take into account long delays because of the different setup considerations at non-constraint resources. The extra WIP, created by the increased time buffers and finished goods inventory levels, will ensure that those resources will not waste too much time waiting for the 'right' type of material. Even in this extreme case the WIP inventory level will be lower, when the DBR methodology is used, than when planned inventories are used to decouple stages. In order to decouple all stages the inventories have to be big enough so that any sequence at any stage is possible.

5.2. *Occurrences of interactive constraints*

When capacities of the various stages are close, then some variations in the setup times and/or variations in the mix may create interactive constraints. The existence of interactive constraints is dependent on the buffer sizes. When the constraint buffer and/or finished goods inventory are very large, resources with little excess capacity will still be non-constraints, because of the amount of time provided to complete the jobs in the preferred sequence.

When the need for shorter buffers is evident and no easy way exists to gain additional capacity, then the drum should include a detailed schedule for each constraint. Constraints should be decoupled by providing enough time between the constraints to execute the specific schedule. Goldratt (1990, chapter 39) outlines the method.

6. The concept of finished goods WIP (FG-WIP) inventory

Suppose we have a family of four product types. At the end of the process, each of the four basic types splits into ten different products.

Let us call the resource department where the last split takes place Z. In this case it is wise to hold WIP inventory in front of Z, the diverging point, and to reduce the levels of the finished goods inventory. This WIP inventory task is to protect the global fluctuations on a whole family of products, e.g. the A product line. Until Z all the operations for any end product, Ai, are the same. The actual determination of Ai into a specific product is made at Z. So, all the three factors (average consumption, fluctuations in market demand and flucutuations in product lead time) that determine the finished goods inventory level for the whole family of product A can be applied in

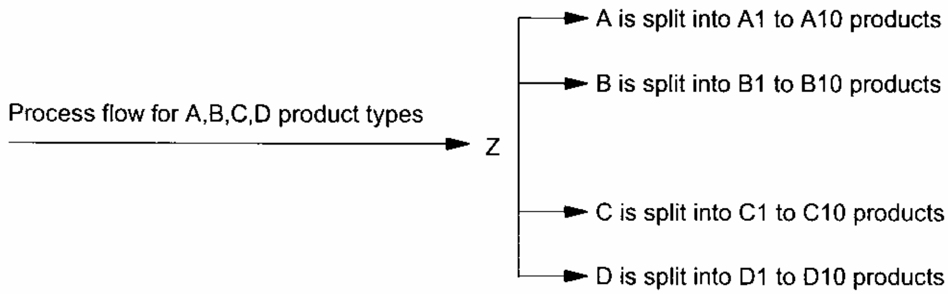

Figure 4. Strategic placement of finished goods—WIP to reduce total inventory.

front of Z. Cox and Finch (1989) explain the advantages of maintaining WIP inventory at diverging points in V-structured industry. The process industry is certainly a logical V structure.

The planned WIP inventory in front of a diverging point is used to protect the market. The overall inventory in the finished goods plus the WIP inventory in front of the diverging point should be lower than the inventory level that would give the same level of protection at the finished goods area only. The characteristics of the diverging point WIP inventory are similar to the finished goods inventory, hence let us call it—finished goods WIP inventory (FG-WIP). When FG-WIP is used in place of a finished goods inventory, the shop floor has been actually divided into two: the main process flow stage and the diverging point (represented here by Z) stage. Z will be scheduled separately and also will be rescheduled more frequently than the rest of the train. The task of the FG-WIP inventory, in this case, is to protect the individual markets within the product family. It is best not to prepare a schedule for Z, but to control it continuously from the sales department based upon actual sales of specific products. A major organization objective is to use JIT techniques to reduce the PLT between the the divergent point and the market to less than CTT. If the response of Z is faster than the CTT there is no need for any 'real' finished goods inventory. This separation may be appropriate for some of Taylor and Bolander's examples such as Coors and Scott's paper (Bolander and Taylor 1990, Taylor and Bolander 1991).

7. Controlling the execution of the plan

The idea of Buffer Management in a DBR implementation was presented by Goldratt and Fox (1986) and further analysed and tested in Schragenheim and Ronen (1991) and in Goldratt (1990, chapters 20–23). The control of the flow between the raw materials area and the inventory constraint for the process flow industry is similar to that used in discrete manufacturing industries as described in Schragenheim and Ronen (1991).

The use of Buffer Management between the inventory constraint and the finished goods inventory needs some modifications. Instead of the DBR's shipping buffer, which is a time buffer to protect due dates, we have actual finished goods inventory to protect from both fluctuations in the market and those of the shop floor. Let's define an 'expedite line' level for the finished goods inventory. When the actual finished goods level drops below this 'expedite line' level, as shown in Fig. 5, it means that any additional fluctuation may exhaust the finished good inventory before a new lot is completed. This is a direct threat to the throughput. What can be done when the

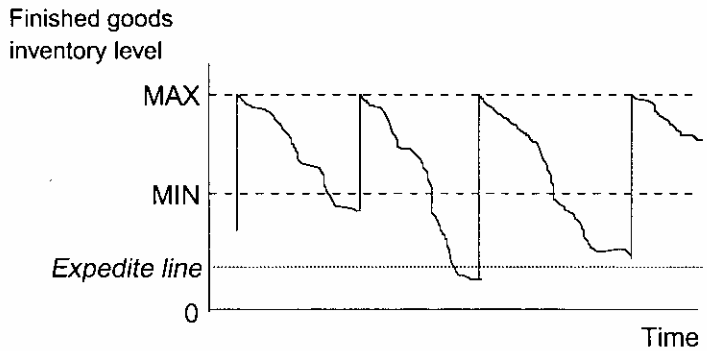

Figure 5. Expedite line to trigger expediting activities.

'expedite line' has been breached? The logical action is to expedite the lot which should be somewhere along its routeing. The level of the 'expedite line' should relate to the lead time under expediting conditions. As a first approximation for the expedite-line level one third of the Min level (which includes all the safety stocks) is recommended. When the lot has not been processed by the constraint, earlier scheduling of the lot on the constraint should be considered.

Whenever expediting occurs, it is highly recommended to collect the data on the actual resource that created the problem and to analyse the data periodically. If a certain product crosses the expedite-line frequently, this means that its Min level is too small. Monitoring the frequency of crossing the expedite-line may indicate a shift in the market demand before sales averages indicate a shift.

8. Conclusions

Huge inventories have a negative financial effect on process flow industries. In addition to a negative effect on the operating expenses, it also has a negative effect of sales due to poor quality and being slow to introduce product changes. Traditionally, in the process flow industry, which is capital intensive, inventories are used to give the impression of good control and high equipment utilization.

DBR achieves simplicity by enabling management to utilize only the capacity needed to meet the market demand. It focuses on scheduling a very few processes that dominate the balance between capacity available and market demand. DBR is also easy to control using buffer management. DBR uses constraints and finished goods inventory only when it serves to ensure the throughput potential.

DBR implementation exposes the hidden capacity associated with maintaining high inventories levels. In a dynamic market, shifts in taste are not recognized and responded to by using traditional practices. In traditional practice, as modelled by Taylor and Bolander, the most loaded resource is frequently busy processing a product that is not really needed now. Imagine, for example, a 5-stage system—how fast (more precisely: how slow) a change in the market can be identified and responded to when inventories are maintained between stages. Compare this to the responsiveness of a DBR system that establishes a minimum level for the finished goods inventories and a constraint buffer based on the production capabilities at the constraint. The six WIP inventory levels needed to decouple the five stages and the market are large relative to strategically placing WIP inventory at control points to ensure the validity of the DBR scheduling methodology.

DBR uses the shortest time buffers, which directly control the WIP inventory levels, and the minimum finished goods inventories, which are needed to maintain an adequate protection of the throughput (or the sales). Using inventory for the protection of the market makes the DBR approach a superior methodology. The exploitation of the capacity of the most loaded resource increases the system throughput and uses the excess capacity of non-constraints to shorten the production lead time. The overall inventory level is much lower than in traditional practice. The reduction in the inventory level with the possible reduction in overtime due to a better control system suggests a lower level of the operating expenses as well.

References

BOLANDER, S. F., and TAYLOR, S. G., 1990, Process flow scheduling: mixed-flow cases, *Production and Inventory Management Journal*, **31** (4), 1–6.

COX, J. F., and FINCH, B. J., 1989, Effective placement of manufacturing inventories. *APICS 32nd International Conference Proceedings*, Orlando, FL, pp. 658–661.

FINCH, B. J., and COX, J. F., 1989, Strategic use of WIP inventory: the impact of bill of material shape and plant type. *Production and Inventory Management Journal*, **30** (1), 63–67.

GOLDRATT, E. M., 1990, *The Haystack Syndrome* (New York: North River).

GOLDRATT, E. M., and FOX, R. E., 1986, *The Race* (New York: North River).

SCHRAGENHEIM, E., M., and RONEN, B., 1990, The Drum-Buffer-Rope shop floor control. *Production and Inventory Management Journal*, **31** (3), 18–23.

SCHRAGENHEIM, E. M., and RONEN, B., 1991, Buffer management: a diagnostic tool for production control. *Production and Inventory Management Journal*, **32** (2), 74–79.

TAYLOR, S. G., and BOLANDER, S. F., 1990, Process flow scheduling: basic cases. *Production and Inventory Management Journal*, **31** (3), 1–4.

TAYLOR, S. G., and BOLANDER, S. F., 1991, Process flow scheduling principles. *Production and Inventory Management Journal*, **32** (1), 67–71.