Fundamental Concepts of Theory of Constraints: An Emerging Philosophy
Ajay Gupta, Arvind Bhardwaj and Arun Kanda

Abstract—Dr Elyahu Goldratt has done the pioneering work in the development of Theory of Constraints. Since then, many more researchers around the globe are working to enhance this body of knowledge. In this paper, an attempt has been made to compile the salient features of this theory from the work done by Goldratt and other researchers. This paper will provide a good starting point to the potential researchers interested to work in Theory of Constraints. The paper will also help the practicing managers by clarifying their concepts on the theory and will facilitate its successful implementation in their working areas.

Keywords—Drum-Buffer-Rope, Goldratt, Production Scheduling, Theory of Constraints.

I. INTRODUCTION

The concept of Theory of Constraints has been emerging as an important philosophy to manage different functional area of any type of business organization. It was started in the year 1980 when Dr. Goldratt developed a production scheduling software called OPT (Optimized Production time-table) to help his neighbour. The software was a great success. Later on, the concept behind OPT software became popular as nine rules of OPT. Since then, Goldratt has written a number of books [1-7] like “The Goal”, “The Goal 2”, “Production The TOC way”, “Theory of Constraints”, “Critical Chain”, “The Haystack Syndrome” “The Race” etc. These books explain many different aspects of TOC. Many of these books are written in the form of novels. So, a person studying them has to filter out the conceptual points by himself. Many researchers around the globe are working to develop and enhance the concepts of TOC and to explore the possibility of its application in different areas. Since the concepts of TOC are not available at a single place, in this paper, an attempt has been made to compile and present different concepts of TOC at a single place. This paper can act as a starting point for the researchers who want to work in TOC. The paper briefly describes Nine OPT rules, five step focusing process of TOC (POOGI), thinking process tools, DBR scheduling technique and Buffer management etc.

Ajay Gupta is working as Associate Professor in the Department of Industrial and Production Engineering at Dr. B. R.Ambedkar National institute of Technology, Jalandhar, India- 144011 phone: 9195010-30373; fax: 91-181-2690320; e-mail: guptaaj@nitj.ac.in.
Dr. Arvind Bhardwaj is working as Professor & Dean (students) in the Department of Industrial and Production Engineering at Dr. B. R.Ambedkar National institute of Technology, Jalandhar, India- 144011

Dr. Arun Kanda is working as Professor in the Department of Mechanical Engineering at Indian Institute of Technology, Delhi, India- 11.
performance of the system, the constraint link has to be improved. After improving the constraint, some new constraint will emerge and we will have to improve that new constraint. So, the process of improvement has to be an ongoing process and not a one time effort. TOC suggests five step focusing process for management and improvement of an operations system based upon this philosophy.

Goldratt [1] identified a host of common problems often found in typical production environment such as late customer order shipment, excessive expediting, constantly changing production plans, high finished goods inventory and long production lead times. The cumulative effect of these problems is that the management is in a constant state of crisis management. Generally the approach of the management is to solve these problems in isolation by taking them up one by one. However, it is important to realize that these problems (called undesirable effects in the language of TOC) are generally caused by one or two core problems. Once this core problem is identified and addressed. As much as 70% of undesirable effects are eliminated. The following sections explain some fundamental techniques of TOC

II. FIVE STEP FOCUSING PROCESS OF ON GOING IMPROVEMENTS (POOGI) -

Any system must have at least one constraint/bottleneck. The five-step process of on-going improvements is used for identifying the bottleneck and managing the production system with respect to this bottleneck. Efforts are expended to relieve this limitation on the system. When a bottleneck is relieved, the firm moves to a higher level of goal attainment and one or more new bottlenecks are encountered. The cycle of managing the firm with respect to new bottleneck is repeated, leading to successive improvements in the firm’s operations and performance. This cycle can be represented in the form of five focusing steps for ongoing improvement [1,4]. These steps are:

[1] Identify the system constraint(s).
[2] Decide how to exploit the constraint(s).
[3] Subordinate all other decisions to step 2.
[5] If in any of the previous steps a constraint is broken, go back to step 1 but do not let inertia become the system constraint.

Identify the system constraint(s): - These may be physical (e.g. materials, machines, people etc), market (insufficient demand) or managerial (e.g. erroneous policies). Generally the organizations have very few physical constraints but many managerial constraints in the form of policies, procedures, rules or methods. It is important to identify these constraints and prioritize them according to their impact on the goal(s) of the organization [10].

Decide how to exploit the constraint(s): - If the constraint is physical, the objective is to use the constraint as effectively as possible. A managerial constraint is not to be exploited but it has to be eliminated and replaced with a new policy, which will support increased throughput. [10]

Subordinate all other decisions to step 2: - This means that every component of the system (non-constraint) must be adjusted to support the maximum effectiveness of the constraint. Because a constraint dictates a firm’s throughput, resource synchronization with the constraint provides the most effective means of constraint utilization. The non-constraints contain productive capacity and idle capacity. If we use the non-constraint beyond productive capacity, it will not increase the throughput but inventory only. [10]

Elevate the constraint: - If the existing constraint is still the most critical resource in the system, rigorous improvement efforts on these constraints will improve their performance. As the performance of the constraint improves, the potential of non-constraint resource will be better realized, leading to overall improvement in the performance of the system. [10]

If in any of the previous steps a constraint is broken, go back to step 1 but do not let inertia become the system constraint: - The first part of this step makes TOC a continuous process. The second part is a reminder that no policy or solution is appropriate for all the times and in all the situations. It is critical for the organization to recognize that as the business environment changes, business policy has to be refined to take account of those changes. Failure to implement step 5 may lead an organization to disaster. [10]

It is generally agreed that any process can be improved upon given enough time, efforts and resources. The important question is from where to start the improvement activity. The process of ongoing improvement discussed above helps us in prioritizing the improvement efforts by stating that we should improve the bottleneck first. Similarly, we should rectify the causes of quality problems at the post bottleneck operations. There may be some defectives in the components processed by upstream operations. But these should be inspected and removed from the batch of components to be processed at the bottleneck.

Following are the two prerequisites to implement this five-step method of on going improvements

[1] Clear definition of the system under investigation and it’s purpose
[2] An appropriate measurement system which align the system to its purpose

Although the 5 step process of improvements TOC can be applied to any process, at any level of management or in any type of organization, the best results come from:

a. Understanding the interdependencies between and across the processes that are used to deliver a product or service.
b. Understanding the impact of these interdependencies and normal process variability on the combined overall performance.
c. Appropriately buffering for interdependencies and normal variability so that the performance can be predictably and consistently high.
The 5 steps of TOC enable an organization to create a stable and reliable value delivery system. In such a system, the management can quickly respond to any market opportunity. Oftentimes, we can proceed through these five focusing steps without difficulty acting at an operational level, but other times the constraints may be an organizational policy—sometimes explicit & sometimes implicit, unstated and intangible and as such all that is perceived is a seemingly unrelated tangle of symptoms or problems. In such cases, the use of TOC logic trees is appropriate and can facilitate diagnosis of nature of illness (core problem), lead to prescription of appropriate remedies and to the institution of a treatment program. [9]

III. TOC THINKING PROCESS

The thinking processes are a set of tools and techniques which allow an individual or a group to solve a problem and/or develop an integrated strategy using the rigor and logic of cause and effect, beginning with the symptoms and ending with a detailed action plan that co-ordinates the activities of all those involved in implementing the solution. It provides a theoretical framework and tools for continuous identification and removal of system constraints [11]. These tools are

Current reality tree: - Current reality tree is used to identify the core problem in a system by listing and linking all the undesirable effects together. It is found that the various undesirable effects can be linked with one another through successive layers of cause and effect relationships [2] and ultimately one core problem can be found in most of the situations. Once this core problem is solved, most of the undesirable effects disappear. The effectiveness of the current reality tree depends on the experience and intuition of the individuals involved in preparing it.

Evaporating cloud (Conflict resolution diagram): - It is used to find the solution of the core problem identified with CRT (Current reality tree). It is not always easy to solve the core problem because it is probably the problem that has existed for a long time. Most probably, everyone in the organization knows about that problem. But they do not know that it is the cause of most of their headaches. So, why this problem has not been solved? Generally, the reason is conflict. Within the organization, there are interests that would be jeopardized by the solution of the core problem. Thus the problem persists. The organizations learn to live with these problems rather than attempting to solve them. Behind most of the conflicts are certain assumptions, if some of these assumptions are found to be invalid or can be made invalid by our actions; the conflict evaporates like a cloud, i.e. how the name of this technique is evaporating cloud.

While trying to resolve a conflict, people generally develop compromise-based solutions. Compromise based solutions do not eliminate the problem. They force both the effected parties to make some compromise and live with the problem. In other words, a compromise based solution ensures the permanent existence of the problem. Let us first examine the meaning of a problem. A problem is defined as something that prevents us from achieving an objective. So, to solve the problem by evaporating cloud method, first step is to clearly verbalize the desired objective. Once the objective has been defined, in the situations involving compromise solutions or conflict, there will be at least two requirements that must be satisfied in order to reach the objective and to satisfy these requirements there will be some pre-requisites. These pre-requisites may require sharing of the same resource that is available in limited quantity or these prerequisites may be contradictory to each other. This is where the conflict arises. Diagrammatically, it can be explained as below

Let the objective is A and the requirements to meet the objective are B and C. The prerequisites for the requirements B and C are D and Not D or D and some more amount of D as shown below

These requirements and prerequisites are always based upon certain assumptions. If we carefully analyze and challenge these assumptions, we will find that some of these assumptions are invalid or can be invalidated. Once this is done both the requirements can be satisfied simultaneously and the conflict is resolved without any compromise. Goldratt [4] explained it by taking the example of compromise-based approach conventionally used in determining economic batch quantity

While determining economic-batch-quantity, the objective is to minimize the total inventory related costs. For this purpose there are two requirements. One, reduce the setup cost. For this purpose the number of setups should be small and the batch size should be large. The second requirement is to reduce the inventory carrying cost. For this purpose, the prerequisite is that the batch size should be as small as possible. This conflict is shown in the following diagram:-
Conventional solution method is to find the solution where the sum of these two costs is minimum. In this solution, we compromise a bit on carrying cost and a bit on ordering cost. Then we carry out the sensitivity analysis to state that the total cost is not badly affected by small deviations (either positive or negative) from economic batch quantity. The assumption that we use in calculating the total ordering cost is that the setup cost per set up is fixed. JIT challenged this assumption and showed that the setup time and cost can be reduced substantially and thus we can move to smaller batch size. TOC challenges setup cost by questioning whether setups cost us anything at all by using the concept of operating expenses. It questions whether an additional setup increase our operating expense at all. On bottlenecks, it does. Not by increasing operating expenses rather by decreasing our throughput. On non-bottlenecks, since we have spare capacity and by additional setups, we will be using that capacity only So, the operating expenses do not increase. Thus the conflict can be resolved without compromise at least on the non-bottleneck operations. Similarly, the conflict between the larger batch size and the smaller batch size can be resolved by using larger process batch (it will reduce the number of setup changes) and smaller transfer batches (It will make the material quickly pass through the production system and the carrying cost will be less). This, again, is a solution without any compromise.

According to sensitivity analysis, equal deviation above or below economic batch size has same impact on the total inventory related costs and hence any one of these can be chosen. Let us change it a bit and see the impact. We know that the profit is equal to selling price minus cost. If we assume that the selling price per unit is constant then as the cost per unit goes up the profit per unit will go down. We replace cost per unit with profit per unit on y-axis. Similarly, we take investment on x-axis in place of total quantity. Now, if we choose smaller quantity, we will require lesser cash but if we choose larger quantity with the same profit, it will require more cash and may mean liquidity problem. So, the impact of larger and smaller batch will not be same on the working of the organization as seen from the global point of view of having sufficient cash in hand to meet the impending liabilities.

Future reality tree: - It tries to portray the future situation, when the solution identified in the previous step is implemented. This will help in judging the suitability of the solution before spending time, money and energy in implementing it. Since future reality tree is a sufficiency-based diagram, it points out the deficiencies in the solution, if any. Similarly, it points out the negative effects of the proposed solution so that the solution can be suitably modified before implementation.

Pre requisite tree: - This tree helps to surface and eliminate the obstacles in the implementation process of a chosen solution. To overcome the obstacles; the intermediate steps/ objectives are defined. To build the prerequisite tree , we begin by listing all the obstacles that stand between the organization and its stated objective. Then for each obstacle we identify a condition that overcomes the obstacle. This identified condition becomes the intermediate objective.

Transition tree: - This tree is generally plotted when the people implementing a solution are not the same who developed it. This tree highlights the steps needed to take the organization from current problem situation to the desired future. To build the transition tree, we identify those actions that we need to take, given our current environment, to achieve the intermediate objectives that we identified in prerequisite tree and the final objective to transition the organization from its current state to the desired future state.

First, CRT should be plotted to identify the core problem, and then this problem should be analyzed and solved by using the evaporation cloud. To prove the effectiveness of solution, FRT is drawn. It presents the future that will exist, once the solution is implemented. FRT is presented to the employees of the organization to get their criticism about the solution (negative branch reservations). To plug these negative holes pre-requisite tree is plotted. Finally transition tree is plotted, which shows the complete steps to take an organization from the current state to the desired future state.

Three steps improvements process: The process of improvements is to be inspired by the following three simple questions

What to change: Every organization in a real environment is overwhelmed with problems and/or opportunities, which needs the manager’s attention and/or corrective actions. However, limited time, efforts and resources make it difficult to act on all such problems or opportunities. Hence, the manager has to find what should be changed (the core problem) to effectively improve the performance [12]

What to change to: Once the core problems have been identified, the next step is to find the solution. If sincere efforts are not directed towards finding solutions of the core problems, chaos and panic will result. [12]

How to cause the change:Perhaps the most difficult of the three questions is to find out how to cause the change in a system? In addition to time, efforts and capital required, the managers often face the problem of emotional resistance from the people in the organization who perceive change as a threat
to their security. If “To what to change to” is identified, but it is not possible to cause that change, then the solution is not of much use. [12]. So, it is necessary to gain the required buy-in and approvals to implement the developed solution. Finally, a detailed action plan has to be prepared using project management technique to successfully implement the solution. The plan should detail the actions to be taken, the person responsible for each action and a time schedule for each action.

Current Reality Tree can be used to answer the first question, while Evaporating Cloud and Future Reality Tree can be used to answer the second question and Prerequisite Tree and Transition Tree can be used to answer the third question. These tools are explained in the above sections. The current reality, the future reality tree and the transition tree are sufficiency based logic diagrams. They consist of a collection of simple declarative statements that are linked with cause and effect relationships. A sufficiency-based diagram is one that identifies all the conditions that are necessary and sufficient to cause a particular effect. On the other hand, the evaporation cloud (conflict resolution diagram) and pre-requisite tree are necessity based logic diagrams. A necessity based logic diagram is one that identifies the conditions that are merely necessary for a particular effect to exist. However, these conditions are not sufficient to cause the effect. e.g. for survival, it is necessary that a person ingests food but the mere fact that someone is ingesting food is not sufficient to ensure the survival of the person. The cause and effect relationships between the statements in logic diagrams are established by connecting them with and/or logical connectors.

IV. TOC BASED MEASUREMENT SYSTEM

To align the efforts of employees to the purpose of the organization, TOC suggests the following measurement systems at corporate level, plant level and process level

**Global performance measures**

- **Net profit:** It is defined as the difference between throughput and operating expenses
- **Return on investment:** It is the ratio of net profit to the inventory
- **Cash flow:** It is a red line of survival which is an on/off type measurement i.e. when a company has enough cash, it is not so important. But when it is not enough, nothing is more important than cash for the survival of the company.

**Plant level performance measures**

- **Throughput:** It is defined as the rate at which a system generates money through sales. [1].
- **Inventory:** All the money that a system invests in purchasing the things, which it intends to sell [1].
- **Operating expenses:** All the money the system spends in turning inventory into throughput. [1].

These definitions are significantly different from those used in a more traditional setting. Throughput is defined as a rate, thus introducing time as a significant factor in measuring throughput. Similarly, inventory is an all-inclusive term defining everything that a firm may choose to sell. Investment in building, property, machines etc as well as raw material is all categorized as inventory by this definition. TOC does not consider value added costs as part of inventory valuation. Similarly, in operating expenses, no distinction is made between direct or indirect, long and short-term expenses. The operative metric used in all these definitions is money [13].

**Process level performance measures:**

- Throughput dollar days
- Inventory dollar days
- Local operating expenses

**Method to allocate operating expenses among the products:**

All plant expenses are considered operating expenses. Divide the total plant expenses by total scheduled hours on bottleneck/constraint to get the cost per constraint hour. Multiply the cost per constraint hour by the number of hours of processing required at the constraint to get the operating expenses for an individual product and add the cost of material to get total cost of a product.

So, when the constraint changes or the processing time per unit changes, the basis of allocation of plant expenses among the products should also changes.

V. DRUM- BUFFER–ROPE METHOD

The scheduling technique of TOC is called Drum–Buffer–Rope. The meaning of the terms Drum, Buffer and Rope is explained below:

**Drum:** The constraint of the system is identified. The rate of output of the constraint determines the output rate of the system. Therefore, a schedule is prepared for the constraint. This schedule is called as Drum schedule.

**Buffer:** To protect the output of constraint from disruptions in upstream operations, sufficient in process material is kept between the material release point and the constraint. This WIP protects the constraint from starvation in case of disruptions up stream. This material is called Buffer or more precisely Constraint Buffer. It is defined in terms of Time rather than number of units.

**Rope:** This is the feedback mechanism used to control the timing of release of material to the actual rate of production at the constraint. When Constraint finishes processing on one unit of material, one more unit of raw material is released into the production system.

**To successfully implement DBR:**

1. It is very important for the key managers to be actively involved in the entire process.
2. It is worth spending significant amount of time to develop realistic scheduling rules for the drum by using simulation.
3. During transition period, some management changes may be required.

VI. BUFFER MANAGEMENT

It is a method of monitoring the presence or absence of material in a buffer and taking actions to prevent disruption of a system’s throughput. [14], The Buffer is of four types

- Constraint Buffer
- Shipping Buffer
- Assembly Buffer
- Capacity buffer

These Buffers refer to time margin in the release of material on the shop floor

Constraint Buffer: - The difference between the time of release of a material into the production system and the time when this material is scheduled to be processed on the constraint is called constraint buffer. This early release provides a safety to the constraint against the disruptions in the upstream operations.

Shipping Buffer: - The difference between the promised due date and the scheduled time by which the material should be ready for dispatch to the customer is called shipping buffer. This buffer is used to ensure the delivery of products to the customer in time and to protect the due dates against the disruptions in the processing operations downstream from the constraint.

Assembly Buffer: - It is the time margin added to those components, which will not be processed on the constraint/ bottleneck but will be assembled with the components being processed at the bottleneck. These components should reach the assembly area before the bottleneck components. So, the release of these components into production system is advanced by some time. This amount of early release is called assembly buffer. If we provide this buffer, the bottleneck components will not be kept waiting in the assembly area for non-bottleneck components.

WIP inventory equal to the Constraint/shipping/assembly buffer should lie in front of the Constraint/shipping/assembly area for most of time. If this inventory level starts reducing, it is an indication of some problem in some upstream operation. This temporary reduction in inventory is called a hole in the buffer. The size of the hole determines whether some immediate corrective action is required or not. Small temporary variations are very common and we need not to bother about them. These small statistical variations are called hole in the region 2. A hole in the region 3 require immediate expediting. Once a region 1 hole has developed. It is likely to prevail for quite sometime. This time depends upon the amount of protective capacity available with the upstream operations. We may have to resort to the actions like overtime at upstream operations to restore these time buffers again.

If there are no holes in the buffer for an extended period of time then it is an indication that the buffer size is unduly high and can be safely reduced without adversely effecting the performance. Tracking the source of region two holes will pin point the potential improvement opportunities in the upstream operations.

Traditional performance measures like efficiency, equipment utilization etc. takes an organization away from customer focus while buffer management and hole system makes every one work for the customer orders. [Wahlers & Cox, 1994]

Capacity Buffer: - It is the amount of extra capacity available with a non-constraint resource over and above the capacity of a constraint resource. This capacity can be used to nullify the effect of a disturbance (e.g. machine break down) quickly and re-build the constraint buffer.

VII. CONTROL POINTS TO MANAGE THE FLOW OF PRODUCTS

Following are the five control points to manage the flow of parts through a manufacturing system:

- Gating operation: - Points of material release into the system.
- Constraint operation: - Bottleneck point
- Point of divergence: - Where an individual part can be used in two or more different products.
- Point of convergence: - Assembly operation
- Termination point: - Shipping area

Specific scheduling information as to product, quantity, time etc. must be provided at the control points to meet the production schedule. The instructions at the other non-critical centers is simply “ Work if work is available, otherwise be ready to work.”

VIII. NINE OPT RULES [GOLDRATT & FOX, 1986]

These are a set of guiding principles used in the development of OPT (optimized production timetable) scheduling software. These rules are

- Balance flow, not capacity
- Level of utilization of a non-bottleneck is determined not by it’s own potential but by the speed and utilization level of some other constraint in the system.
- Utilization and activation of a resource are not synonymous.
An hour lost at a bottleneck is an hour lost for the total system.  
An hour saved at a non-bottleneck is just a mirage.  
Bottlenecks govern both throughput and inventory in the system.  
A transfer batch may not, and many times should not, be equal to the process batch.  
The process batch should be variable and not fixed.  
Schedules should be established by looking at all of the constraints simultaneously. Lead times are a result of a schedule and cannot be predetermined.

IX. TOC HEURISTICS:

If the total demand exceeds the production capacity of a system, then this heuristics can be used to determine the optimal product mix to be manufactured. The steps in this heuristics are;

Determine the system constraint
Determine the throughput per unit for all the products.
Divide throughput per product by the time it takes to process it on the bottleneck to get throughput per unit time of the bottleneck.
Manufacture as much quantity of product with highest throughput/time ratio as feasible (Limited by the bottleneck capacity or market demand).
If some idle capacity is available after the previous step, allocate it to manufacture the product with next highest throughput/time ratio
Repeat the previous two steps till the capacity of bottleneck is exhausted.
If there is a single constraint, then this heuristics gives optimal solution but when there are more than one constraint resources, it either gives sub-optimal or infeasible solution[Plenart]. So, it needs to be modified a bit. Since we have to take care of all the constraints simultaneously.

X. CONCLUSION

Over the years, starting from a simple production scheduling software, Theory of Constraints has developed into a complete management philosophy. But the information on various concepts of TOC is not available at a single place. This paper describes the main concepts like five step focusing process of on going improvement, thinking process tools, nine OPT rules, Drum-Buffer-Rope, TOC product mix heuristics etc. TOC is a complete tool kit in itself and can be used in identifying and resolving the problems faced by any organization. Many of these concepts are contradictory to the orthodox cost accounting based thinking. If properly learnt and practiced, these can be very important tools in the tool kit of a manager.

REFERENCES