Theory of Constraints
and
Lean Manufacturing:
Friends or Foes?

By Richard Moore, Ph.D. and Lisa Scheinkopf

Copyright © 1998, Chesapeake Consulting, Inc.

All Rights Reserved
Introduction

The Theory of Constraints (TOC) and Lean Manufacturing are two popular business philosophies that have received a great deal of attention in recent years. Their objectives and underlying assumptions are at the same time strikingly similar and in stark contrast.

TOC is a popular business philosophy that first emerged with Dr. Eliyahu Goldratt’s landmark book, *The Goal*. One of the strengths of the TOC approach is that it provides focus in a world of information overload. It guides its practitioners to improve their organizations by focusing on a very few issues—the constraints to ongoing profitability.

In recent years, Lean—derived from the methods of the successful Japanese automobile manufacturer Toyota—has also gained popularity in manufacturing circles. Lean became internationally recognized as a result of the Womack & Jones book, *The Machine That Changed The World*. The Lean approach guides its practitioners to improve their organizations by focusing on the elimination of any and all waste.

Both philosophies focus on improvement and advocate techniques to control the flow of material on the shop floor. Both have demonstrated dramatic results of implementations—profitability skyrockets, inventories and lead times are slashed, and operations are drastically simplified, just to name a few. At the same time, both movements recognize that in order to achieve and sustain such improvement trends, the change agent’s perspective must expand beyond the walls of manufacturing to include the rest of the enterprise. As a result, TOC and Lean movements have expanded their scope to encompass principles and practices of the entire system to enable continuous system-wide improvement.
Implementation success forges strong loyalties to the techniques used to bring about the success. Occasionally, loud voices from each of the two camps—the TOC camp and the Lean camp—claim their philosophy is the philosophy, and that there is no room for both philosophies to exist in an organization. Their feeling is that if you advocate Lean, you must disavow TOC, and if you practice TOC, you must disavow Lean. We disagree.

In reality, the techniques of TOC and Lean are for the most part in harmony. The purpose of this paper is to highlight the points of agreement as well as the areas of potential conflict. In this paper, we will:

- Provide an overview of the basic principles and practices of TOC (Section One) and Lean (Section Two), focusing on their application to manufacturing organizations.
- Establish common ground shared by both philosophies (Section Three).
- Examine some of the obvious and some of the not so obvious critical differences between TOC and Lean implementations (Section Four).
- The overall objective is to seek a symbiotic relationship that captures the best elements of each. The end result will be to introduce an approach to Lean and TOC that highlights and addresses the actual and perceived conflicts (Section Five). It is our belief that this approach will provide a path for companies to reap even greater benefits in the short and long term.

We have written this paper to serve as a single point of reference. Those readers who are well versed in TOC but not Lean may want to skip Section One and proceed by reading Sections Two through Five. Those versed in Lean but not TOC may choose to
read Section One, and then Three through Five. Those readers who are unfamiliar with either approach or would first like a good overview of both should read the entire paper.
Section One: TOC (Theory of Constraints): The Fundamentals

TOC views organizations as systems consisting of resources, which are linked by the processes they perform. The goal of the organization serves as the primary judge of success. Within that system, a constraint is defined as anything that limits the system from achieving higher performance relative to its purpose. The pervasiveness of interdependencies within the organization makes the analogy of a chain, or network of chains, very descriptive of a system’s processes. Just as the strength of a chain is governed by its single weakest link, the TOC perspective is that the ability of any organization to achieve its goal is governed by a single, or at most very few, constraints.

While the concept of constraints limiting system performance is simple, it is far from simplistic. To a large degree, the constraint/non-constraint distinction is almost totally ignored by most managerial techniques and practices. Ignoring this distinction inevitably leads to mistakes in the decision process. The implications of viewing organizations from the perspective of constraints and non-constraints are significant. Most organizations simultaneously have limited resources and many things that need to be accomplished. If, due to misplaced focus, the constraint is not positively affected by an action, then it is highly unlikely that real progress will be made toward the goal. Given this perspective, TOC’s 5-step process offers a systematic and focused process which organizations use to successfully pursue ongoing improvement:

The Five Focusing Steps
1. **Identify** the system’s constraint.
2. Decide how to **exploit** the system’s constraint.
3. **Subordinate** everything else to the above decisions.
4. **Elevate** the system’s constraint.
5. Don’t allow **inertia** to become the system’s constraint. When a constraint is broken, go back to step one.

However, prior to identifying the constraint, two prerequisites must be satisfied to gain perspective for the analysis.
a) Define the **system** and its **purpose (goal)**.

b) Determine how to **measure** the system’s purpose.

When this process is applied to manufacturing, the following usually unfolds:

a) Define the **system** and its **purpose (goal)**

Given that the roots of TOC are deeply embedded in manufacturing, often the system is initially defined as the manufacturing operation, or the plant. The purpose of the manufacturing operation is to enable the entire organization to achieve its goal. It is important to have clear definition of the organization’s goal. One goal shared by most manufacturing companies is to make “make more money now as well as in the future.” While this goal may be arguable in special circumstances, making money certainly provides the funds to fuel ongoing operations and growth regardless of other stated goals. As such, making money is at least a very tight necessary condition in almost every organization. As a result, it is appropriate to continue this example using “making more money now as well as in the future” as the goal of the manufacturing organization. The next question to be answered is **how do we measure “making money?”**

b) Determine how to **measure** the system’s purpose.

Manufacturing organizations purchase materials from vendors, and add value by transforming those materials into products their customers purchase. Simply stated, companies are “making money” when they are creating value added at a rate faster than they are spending. In order to calculate “making money,” TOC starts by categorizing what a firm does with its money in three ways:

1. **Throughput (T)** is defined as **the rate at which the organization generates money through sales**. The manufacturing process adds value when customers are willing to pay the manufacturer more money for the products than the manufacturer paid its vendors for the
materials and services that went into those products. In TOC terminology, this “value-added” is the **throughput**.

2. **Operating expense (OE)** is defined as *all of the money the organization spends in order to turn inventory into throughput*. Operating expense includes all of the expenses that we typically think of as “fixed.” It *also* includes many that are considered to be variable, such as direct labor wages. In order to be profitable, the company must generate enough **throughput** to more than pay all of the **operating expenses**. As such, profit is calculated simply as T – OE.

3. Rate of return is also an important measure of profitability. An unacceptable profit is made when it’s bringing a poor rate of return on investment—and this return is greatly impacted by the amount of money that is *sunk in the system*. In TOC terminology, this is “**Inventory**.” Formally, inventory (I) is defined as *the money that the system spends on things it intends to turn into throughput*. Return on investment, then, is Net Profit (T-OE) divided by Inventory (I). Inventory, as used in this equation, includes what is known as “passive” inventory such as plant and equipment. However, in improving manufacturing operations, the focus is much more on reduction of “active” inventory—the raw material, work-in-process, and finished goods needed to keep the system running.

Often, it is easy to lose sight of the goal in the process of making day-to-day decisions. Determining the impact of local decisions is complicated by the fact that measuring the net profit of a manufacturing plant in isolation from the larger system is impossible (though many organizations fool themselves into thinking they can). In practice, productivity and inventory turns may be more appropriate measures than profit at the plant level. The TOC approach to measuring productivity and turns uses the same three fundamental measures—T, I, and OE. Productivity is measured as T/OE—in essence, the ratio between money generated and money spent. Meanwhile, inventory
turns is measured as $T/I$—the ratio between money generated and level of investment required to generate it.

The concept of allocating all of the money in a system into one of three mutually exclusive and collectively exhaustive categories of throughput, inventory, or operating expense may appear unconventional at first. Why would one do such a thing? The real power lies in using $T$, $I$, and $OE$ to evaluate the impact of decisions on the goal of making money. When we want to have a positive effect on net profit or return on investment, on productivity or turns, we must make the decisions that will increase throughput, decrease inventory, and/or decrease operating expense. The cause-effect connection between local decisions and impact on the intermediate measures of $T$, $OE$, and $I$ is usually much more clearly defined. These intermediate measures can then serve as direct links to the more traditional global financial measures.

Given three measures, one naturally takes priority over the others. One of the distinguishing characteristics of managers in TOC companies is that they view throughput as the measure with the greatest degree of leverage in both the short and long term. This is largely due to the fact that, of the three measures, opportunities to increase throughput are virtually limitless. Certainly, reducing inventory and/or operating expense cannot go less than zero, and in many cases, reducing one or both may have a significant negative impact on throughput.

An overriding principle that guides TOC companies is that ongoing improvement means growth. They believe that growth doesn’t happen by concentrating on what to shrink, but rather by concentrating on what to grow. That means concentrating on the means by which they choose to increase Throughput. This emphasis on throughput first (inventory second and operating expenses as third) is referred to as “Throughput World Thinking,” and is often held in contrast with the common managerial obsession with cost reduction, hence the term “Cost World Thinking.” (For a more exhaustive discussion of these measures and their implications for management see The Haystack Syndrome pages 14-64.)

With the prerequisites of defining the system and its measures fulfilled, let’s move on to the Five Focusing Steps.
1. **Identify** the system’s constraint. For the manufacturer, the question to be answered here is “what is physically limiting our ability to generate more throughput?” The constraint will be located in one of three places: 1) the market (not enough sales), 2) in the vendors (not enough materials), or 3) in an internal resource (not enough capacity of a resource or skill set). From a long-term perspective, an additional question must be answered—if not immediately, then as soon as the operation is under control by implementing the next two steps. That question is *where does our organization want its constraint to be?* From a strategic perspective, where *should* the constraint be?

2. Decide how to **exploit** the system’s constraint. When we accept that the rate of throughput is a function of the constraint, then the question to be answered at this step is, “*What do we want the constraint to do,* in order that the rate of throughput generated by it is maximized (now and in the future)?” The following activities and processes are typically implemented in association with this step:

   - **When the constraint is internal:**
     - The resource is considered as “a most precious and valuable resource.”
     - Wasted activity performed by the constraint is eliminated.
     - People focus on enabling the resource to work on the value added activities that it alone is capable of doing. This often means that the constraint resource offloads other activities to non-constraints.
     - Attention is paid to setup, and efforts are made to minimize setup time on the constraint resource.
     - Utilization and output of the constraint is measured. Causes for downtime on the constraint are analyzed and attacked. Care of the constraint resource becomes priority number one for maintenance, process engineering, and manufacturing engineering.
     - Inspection steps may be added in front of the constraint, to ensure that only good material is processed by it. Care is taken at the constraint (and at every step after) to ensure that what the constraint produces is not wasted.
• Often, extra help is provided to aid in faster processing of constraint tasks—such as setup, cleanup, paperwork, etc.
• Steps are taken in sales and marketing to influence sales of products that generate more money per hour of constraint time.

**When the constraint is raw materials:**
• The raw material is treated like gold.
• Reducing scrap becomes crucial.
• Work in process and finished goods inventory that is not sold is eliminated.
• Steps are taken in purchasing to enhance the relationships with the suppliers of the constraint material.
• Steps are taken in sales and marketing to influence sales of product that generate more money per unit of raw material.

**When the constraint is in the market** (and we claim that the major constraint is almost always in the market):
• The customers are treated like precious gems.
• The company gains an understanding of critical competitive factors, and takes the steps to excel at those factors. From the manufacturing perspective, this usually means
  • 100% due date performance
  • Ever faster lead times
  • Superior quality (as defined by customer need)
  • Adding features (as defined by customer need)

• While a discussion of strategic constraint placement is a topic beyond the scope of this paper, suffice it to say that there are advantages to strategic selection of an internal “material flow control point.” When the constraint is internal, the constraint resource is almost always selected as the control point.
• To exploit the constraint/control point, it is finitely scheduled to maximize output without overloading it. Overloads serve only to
increase lead-times as work queues back up in front of the constraint. The schedule defines precisely the order in which that resource will process products. It serves as “the drum” for the rest of the manufacturing organization. The drum is based on real market demand (in other words, the market demand is what pulls the schedule). This schedule serves as the backbone of an operations plan that meets due date performance while simultaneously maximizing throughput and minimizing inventory. It is the first element of the “drum-buffer-rope” process for synchronizing the flow of product. The buffer and rope aspects will be discussed below under subordinate.

3. **Subordinate** everything else to the above decisions. Step one identifies the key resource determining the rate of throughput the organization can generate. In step two, decisions are made relative to how the organization intends to maximize that rate of throughput—how to make the most with that it has. In this step, the organization makes and implements the decisions to ensure that its own rules, behaviors and measures enable, rather than impede its ability to exploit the identified constraint. “Subordinate” is the step where the majority of behavior change occurs. It is also in this step that we define “buffer” and “rope:”

- The ability of the company to maximize throughput and meet its promised delivery dates hinges first on the ability of the constraint/control point to meet its schedule—to march according to the drum. TOC also recognizes that variability—in the form of statistical fluctuations everywhere—exists in every system. It is crucial that the drum be protected from the inevitable variability that occurs. The means by which it attempts to assure this is the buffer. A TOC company does not want to see its drum schedule not met because materials are unavailable. Therefore, work is planned to arrive at the constraint/control point some period of time prior to its scheduled start time. The buffer is the amount of time between the material’s planned arrival time at the control point and its scheduled start time on the control point. The same concept is put to work in what is called the
shipping buffer. In companies where it is important to meet the due
dates quoted to their customers (and we can’t think of any companies
where it’s not important), work is planned to be ready to ship a
predetermined amount of time prior to the quoted ship date. The
difference between this planned “ready to ship” time and the quoted
ship date is the shipping buffer.

• In a TOC company, work is released into production at the rate
dictated by the drum and is timed according to the predetermined
length of the buffer. This mechanism is called the rope, as it “ties” the
release of work directly to the constraint/control point. This third
element ensures that the TOC plant is operating on a pull system. The
actual market demand pulls work from the constraint/control point,
which pulls work into the manufacturing process.

• It is important to note that at all places other than those few requiring
buffer protection, inventory is expected to be moving. There is no
planned inventory anywhere else. The end result is very low total
inventory in the manufacturing operation. Low total inventory in turn
translates into shorter lead times as work center queues are minimized.

Several additional activities and behaviors that are required to support the
*subordinate* rule include:

- **Roadrunner** mentality takes over. The analogy of the roadrunner cartoon character is used to portray the approach to work. The roadrunner operates in two speeds—full speed ahead or dead stop. In a TOC plant, if there is work to be worked on, work on it at full speed ahead (of course, the work is to be of high quality as well). If there is not work to work on, stop. Congratulations for emptying your queue. Take the time you have with no queue and use it for learning, for cleaning your work area, for helping another team member, or on another activity that will ultimately help the organization. It’s even OK to take a break. The workers’ purpose is to turn inventory into throughput, not simply to produce more inventory. Workers are responsible for ensuring that the drum of the organization doesn’t miss a beat.

- Performance measures change. For instance, in many TOC companies everybody is measured on constraint performance to schedule. Maintenance is measured on constraint downtime. Gain sharing programs are modified to include constraint and throughput-based measures. The old measures of efficiency and utilization are abandoned at non-constraints.

- Protective capacity is maintained on non-constraint resources. We have already established that manufacturing organizations have both dependency and variability. Buffers are strategically placed to protect the few things that limit the system’s ability to generate throughput and meet its due dates. If we have a system in which the capacity of every resource is theoretically the same, then every instance of variability (e.g. breakdowns, slow processing times, defective raw material) will result in some degree of buffer depletion. After some period of time, the buffer will be depleted enough that the constraint shuts down—this is the throughput equivalent of shutting down the whole system. Unless, of course, heroic (and expensive) efforts such as overtime, outsourcing, or customer cancellations readjust the system. In a TOC environment, additional capacity is intentionally maintained on non-constraint resources for the purpose of overcoming the inevitable variations (a.k.a. instances of Murphy’s Law) before the system’s constraint notices.
The combination of a few strategically placed buffers and protective capacity results in a predictable, stable overall system that has immunized itself from the impact of the inevitable variations that occur.

- Buffer management is used as a method to ensure that constraint and shipping schedules are met, and to focus improvement efforts. In a TOC plant, a short 10-15 minute meeting occurs every shift, and replaces the typical production meeting. Called a buffer management meeting, its participants:
  - Check the release schedule, and keep a record of early, on-time, and late releases.
  - Identify any work that is part of the planned buffer that is not yet at the buffered resource.
  - Identify the current location of the missing work.
  - Assign appropriate personnel (usually, someone from the current meeting) who will make sure the work moves quickly from its current location to the buffered resource. This action becomes their first priority upon leaving the meeting.
  - The current location of the missing work and the amount of drum-time that work represents is recorded. This step is key to continuous improvement. Periodically (weekly or monthly) this data is analyzed to determine where work meant for the drum is stuck most often. This becomes the focus for improvement effort. Causes are identified and removed. Some of the “exploit” techniques are employed to ensure that wasteful activity is removed from the processes performed by that resource. If these activities don’t create sufficient protective capacity (enough capacity that this resource is no longer the major cause for “holes” in the buffer), additional capacity may be acquired. The intent is to increase the velocity of the flow of material (the transformation of inventory into throughput). Once the obstruction to flow is resolved, the size of the buffer may be decreased.

4. **Elevate** the system’s constraint. The first three steps represent the TOC approach to maximizing the performance of the initial system. In the “elevate” step, the
constraint itself is enlarged. If the constraint is capacity of an internal resource, more of that capacity is acquired (additional shifts, process improvements, setup reductions, purchasing equipment, outsourcing, hiring people, etc.). If the constraint is materials, new sources for material are acquired. If the constraint is in the market, then sales and marketing bring in more business. At some stage during the elevate step the constraint will move to another location in the system.

5. Don’t allow inertia to become the system’s constraint. When a constraint is broken, go back to step one. This step reminds us to make it an ongoing improvement process. It also reminds us that once the constraint is elevated, to ensure there is sufficient protective capacity surrounding it. And that as the constraint changes, so must the rules, policies, and behaviors of the people in the organization.
Section Two: Lean Fundamentals

The Lean enterprise is focused on the elimination of muda—waste. Drawing on a rich history of Japanese manufacturing techniques, Lean advocates also outline a five-step process. Based on a customer-focused systems view, this process provides the foundation of any lean enterprise.

1. Specify Value

A waste-free process is a process that is working correctly. It takes time and effort to get the waste out of a process, so it is important to work on processes that create value. A firm’s customers are the final judges as to whether or not the firm has created value. Therefore, one category of muda (waste) is having the “right” process for a product or service that the customer doesn’t want. Lean companies therefore work to precisely define value in terms of specific products with specific capabilities offered at specific prices through a dialogue with specific customers. In other words, they work to understand and deliver what the customer wants to buy. Lean companies often restructure on the basis of product line, organizing managers and employees into product teams.

2. Identify the Value Stream

In the first step, the products that are waste (customers don’t want them), and products that are value (customers want to buy them) were identified. Now it is time to begin the process of identifying and removing the waste that is involved in providing the products that are value to your customers.

It is frequently amazing the things that are discovered in a good process review. Not only are there unanticipated problems that surface, but also processes are often riddled with solutions to previous, but now non-existent, problems. John Covington tells “the story of the stick-mark.” Once upon a time, John worked as an engineer at a company that produced x-ray film. One of all engineers’ early assignments was to try to solve the “stick mark problem.” The company had been scrapping millions of dollars worth of film every year, due to what they called a “stick mark.” A stick mark was a discoloration on the film that occurred during the production process. The first thing John did was to take samples of good film and scrapped film from the scrap bins. He then brought them
to a local hospital, where he asked the x-ray technicians to look at the film and see if they could find anything wrong with it. “No,” was their response. They compared it to good film, and could not detect any differences. They then took some x-rays with “stick mark” film and “good” film. The radiologists were unable to detect any deficiencies in the x-rays, and were unable to detect any difference between the x-rays shot with either of the films.

John went back to his company to try to understand why they were throwing away film that customers considered to be good. Way back when, in the early days of film production, the process included a step where the film needed to be dried. The way in which it was dried was by hanging it over a stick. Sometimes, this process resulted in a long mark where the stick was. This was unusable film, and had to be thrown away. The company then instituted an inspection process so that good film could be sorted from “stick mark” film. As time went on, the company made better and better use of technology. Not only did the manufacturing process improve, but also as the stick marks became harder to detect, the inspection technology “had to be” improved as well, so that stick marks that were invisible to the eye could be caught. The defect had been eliminated, but the wasteful inspection and sorting process didn’t. Wasteful? Yes, absolutely!

Womack and Jones¹⁵ define the value stream as “The set of all the specific actions required to bring a specific product through the three critical management tasks of any business: ... problem solving, ... information management, ... physical transformation.” As you gain an understanding of what the value stream for a product is, you will discover three categories of activities:

- **Steps that definitely create value.** In the manufacturing process, these are the steps that are actually transforming the fit, form, or function of the raw material, and bring it a step closer to the finished product.

- **Steps that create no value but are necessary, due to current state of the system.** In the manufacturing process, these might include inspection, waiting, and some transportation steps.

- **Steps that create no value and can be immediately eliminated.** If the activity clearly does not fall into one of the above categories, it should be stopped.
During this step in the process of becoming lean, detailed process flow diagrams are created for each product, highlighting all of the steps that are considered to be muda. This is usually done in the context of Kaikaku—lean’s term for radical improvement. Contrasted with kaizen, or continuous improvement, kaikaku, also known as breakthrough kaizen, is an intense questioning and reexamining of every aspect of a process. Any steps that can be eliminated immediately are stopped. Any activities that are identified as “non-value but currently necessary” become targets for improvement.

This is also the point at which “target costing” is implemented. In short, target costing is a methodology in which the cost of a product is established based on its “muda-free” process. What if we didn’t have scrap? What if we didn’t have to conduct receipt inspections? This is now the cost that the company strives to achieve through the elimination of muda. As it gets closer to the target cost, the lean philosophy suggests that the company will then be able to enjoy increased profits, or to reduce its selling prices to its customers, thereby increasing value in the customers’ eyes.

3. Flow

In order to document the process, Lean teams will physically walk the process, noting the distance the product must travel in order to go through its entire process. Some very small operations report that their process is over a hundred miles long, and it is estimated that the process of producing aircraft is tens of thousands of miles long! With the process-specific muda identified and on its way to elimination, the purpose of this step is to encourage organizations to focus on rapid product flow, unencumbered by the walls and the physical distance that exist between typical functional departments. “Lean enterprises” are created for each product. The physical layout of the people and equipment involved in the process is changed. Factory floors are laid out in cells rather than in functional groupings, thereby reducing the distance parts must travel. Where before there were departments for engineering, scheduling, and customer service, lean enterprises have teams of people from each of those disciplines comprising the team responsible for the business of specific products.
It is here that lean enterprises implement what is called 5S, a methodology used to reduce the slack hidden in plants. (Monden, 1996, p. 199.) 5S is comprised of the activities listed below, which collectively translate to a cleanup activity at the workplace. The intent of 5S is to remove the muda associated with clutter and disorganization.

- **Seiri** – separate the necessary things from the unnecessary and discard the unnecessary
- **Seiton** – Neatly arrange and identify things for ease of use (a place for everything, and everything in its place)
- **Seiso** – to always clean up; to maintain tidiness and cleanliness—to clear your workplace thoroughly
- **Seiketsu** – To constantly maintain the 3S mentioned above, Seri, Seiton, and Seiso. Keeping a clean workplace without rubbish or oil leakage is an example of Seiketsu.
- **Shitsuke** – to have workers make a habit of always conforming to rules.

4. **Pull**

In the lean enterprise, inventory is considered to be waste. Therefore, producing anything that is not sold is waste as well, for if it’s produced but not sold, it remains as finished goods inventory. Thus, it is important that real customer demand pull product through the system. This is in stark contrast with the traditional push approach to manufacturing where the system encourages each resource to produce as much as possible, thus, pushing products through its system. Once the first three steps are implemented, this concept is especially important. Because the process is shortened when wasteful steps, wasteful activity within steps, and distance parts must travel is removed, lean organizations usually find themselves with the capability to produce more than before. In a push environment, such capability would translate into increased inventory—not exactly what you would call lean. In a pull environment, this tendency to overproduce is controlled. Activities may then be directed toward either removing excess capacity or increasing the rate of pull.
Today’s information technology makes it possible for more and more systems to transition from the push mentality embodied in the traditional approach of manufacturing and distributing products based on forecasts. In an era of dynamic markets, where last year’s demand in no way reflects what will happen this year, the traditional push approach places undue weight on historically based forecasts. In today’s world, to the extent that the manufacturing and distribution system is responsive, it is far more effective to manufacture based on actual customer demand. Point of sale terminals provide the capability to capture in detail exactly what was sold and pass that information back through the supply chain to the distributors manufacturers, and even to vendors. The practice of pull is made operational in lean enterprises with two methods, *takt time* and *kanban*.

**Takt Time**

*Takt time*\(^1\) is used to set the pace of production by matching production rate with the rate of customer demand. The *takt* time is calculated by dividing the available production time by the rate of customer demand. As an example, for a plant that operates on a single 8-hour shift (480 minutes) with a demand of 120 units/day, the *takt* time is four minutes. Knowing this time is significant, in that it provides a sense of the desired pace of a plant’s output. The point is always to define *takt* time precisely at a given point in time in relation to demand and to run the production sequence precisely to *takt* time. In a lean enterprise, the goal of every resource at every step along the process is to produce at the rate demanded by *takt* time. Often the *takt* time and each resource’s progress relative to this target is posted and displayed. Meanwhile, many manufacturing environments currently lack the flexibility to respond to frequent changes in *takt* time. The variation is considered to be *muda*, and becomes a candidate for improvement teams to eliminate.

**Single Piece Flow**

The following quote describes material flow after a lean implementation in a bicycle plant. “In the continuous-flow layout, the production steps are arranged in a sequence, usually within a single cell, and the product moves from one step to

---

\(^1\) Spelled *tact* time by Toyota Production System advocates (Monden, 1998, p303)
the next, one bike at a time with no buffer of work-in-process in between, using a range of techniques generically labeled “single-piece flow.” The lean philosophy considers any idle inventory to be muda. With the combination of takt time and single piece flow, the lean enterprise strives to achieve no idle inventory. Often, companies implementing lean begin with kanban systems. Kanban places small queues of inventory that are of a predetermined size at every resource. The job of each resource is to work to fill the queues to their predetermined size. When the queue is full, the preceding resource stops. In single piece flow, the queue is zero. The preceding operation works when the resource it feeds is ready to pull the piece directly. Single piece flow and kanban enable “pull” by effectively “stopping the push.” Workers know when to start working, and they know when to stop working. Idle time on the part of workers is considered to be muda in a lean environment. When a worker is prevented from working due to a full queue in a kanban system, or a downstream resource not ready to pull a part in a single piece flow system, idle time occurs. Elimination of this idle time is yet another candidate for “muda-attack” in a lean environment.

5. Perfection

The initial successes that are achieved as a result of implementing the first four steps highlight new opportunities for improvement in reducing effort, time, space, cost, and mistakes while offering products and services that more closely reflect what the customer really wants. This step serves to remind us that continuous improvement is possible, and is the desired state of any change in any environment. Dreaming about perfection is fun and useful, because it shows us what is possible and helps us to achieve more than we would otherwise. To keep the pump primed for perfection, mature lean organizations practice open book management and work hard to instill the spirit of kaizen or continuous improvement.
Section Three: TOC and Lean – Areas of Agreement

There are many areas of agreement between the advocates of lean thinking and TOC.

Value

Both TOC and Lean embrace the value principle in that the customer’s perception of value is crucial. Lean’s perspective is that “Value can only be defined by the ultimate customer.” Similarly, TOC suggests that throughput is not generated until a customer’s check for the product has cleared the bank. For TOC, the customer’s perception of value is a major determining factor in increasing a product’s throughput. In It’s Not Luck, Goldratt states “A product that relieves prospect’s problems brings benefits—the more and bigger problems that it relieves the greater the benefits.” Clearly the focus is on the customer’s perception of value.

The Value Stream

Whether called a value stream by Lean or a value-added lane by TOC the concept is the same—an explicit acknowledgment that customer value is created by a chain of interdependencies that extends far beyond the walls of the manufacturing plant. Both approaches also inherently recognize that the job of every person in the organization is to turn inventory into throughput. Defining the system and creating an understanding of the actual process flow (with process flow and other diagrams) is an early step in both TOC and Lean implementations.

Flow

Flow has long been a buzzword with respect to improving manufacturing operations. Schonberger’s 1982 book Japanese Manufacturing Techniques: Nine Hidden Lessons in Simplicity stressed the importance of flow in a chapter entitled Simplify, and goods will flow like water. Synchronous Manufacturing devotes a section to “a river analogy” which describes the flow of goods through the system. Synchronous management Principle One is Balance flow not capacity. It is clear that both TOC and Lean philosophies advocate flow. In Lean Thinking, Womack and Jones have logically
expanded the concept well beyond the manufacturing plant and back into the feeding processes of design and order entry. There is absolutely no conflict between the two philosophies on the concept of the importance of flow.

**Pull**

Both TOC and Lean embrace the *pull* principle, and offer techniques to control the flow of product based on pull from the market. Lean approaches the pull concept sequentially, meaning that no one upstream should produce a good or service until the customer downstream asks for it. Pull is the driving force behind TOC’s Drum-Buffer-Rope (DBR) methodology as the market demand provides the basis for scheduling the constraint (drum), which is the basis for releasing any material (rope) into the manufacturing process.

**Perfection**

When the 1984 version of *The Goal* was revised two years later, the book’s title changed from *The Goal: Excellence in Manufacturing* to *The Goal: A Process of Ongoing Improvement*. In this revision, an epilogue was added in which the leading character observed, “The only way for an organization to survive and prosper ultimately is to change . . .” This organization will be one where the energy is directed not at achieving stability, but at pursuing ongoing improvement. In the sense that the end-state of ongoing improvement is an endless pursuit of perfection the two movements are in total agreement.

Meanwhile, both Lean and TOC tap the work force as a source of the improvement effort. The importance of the worker participation in improving systems cannot be understated. It is the ownership of the ideas, coupled with the demonstrated results, which ignites the fire of continuous improvement.
Section Four: The Differences—from Paradigms to Practice

The advertised objective of improvement efforts is one of the major differences in TOC and Lean. In this area, Lean and TOC have completely different obsessions: reduce *Muda* versus increase throughput.

*Muda* is described by Womack in *Lean Thinking* as “the one word of Japanese you really must know.”\(^{15}\) *Muda* means waste, and lean goes after waste with vengeance. Monden’s book *Toyota Production System* provides additional background underlying lean thinking.

“The Toyota production system is a viable method for making products because it is an effective tool for producing the ultimate goal—profit. To achieve this purpose, the primary tool of the Toyota production system is cost reduction, or improvement of productivity. Cost reduction and productivity improvements are attained through the elimination of various wastes such as excessive inventory and excessive work force.”\(^{8}\) (emphasis added.)

The statement is clear and unequivocal. For the cost conscious manager, it provides strong support and validation of the correctness of his cost-cutting activities. In both worlds, net profit is calculated the same way—throughput minus operating expense. Improving net profit by reducing costs seems like a logical and irrefutable concept. However, to the ears of TOC advocates, this statement is like fingernails being run down the blackboard. The TOC advocate is obsessed with improving net profit by increasing throughput. All decisions are evaluated based on the impact of the decision on Throughput, Inventory, and Operating Expense. The primary emphasis is placed on increasing throughput. Decisions made to change inventory or operating expense are done so with the express purpose of increasing throughput as a result of those changes.

This single fundamental difference in paradigm—the view of continuous improvement via reducing costs (waste) or increasing throughput—is at the root of very different behaviors and practices in organizations that embrace Lean or TOC. Let’s examine some of these differences.
Implementation Approach—Perspectives on Waste

While it is clear that products and services should be tailored to match the needs and desires of the customer, the focus in most TOC efforts is to improve the system *beginning with the current state of the system*. If the constraint is in the market, then it may be that the company is not adding value as defined by the customer. If the constraint is internal (the market wants more than can be supplied) then, at least to a portion of the market, the company is doing something correctly. Meanwhile, there is strong agreement on customer value—the long-term health of any company depends on the company’s ability to meet the needs of its customers.

In practice, however, many TOC practitioners miss opportunities to eliminate the very real waste of premature and excessive spending on capacity. Sometimes, in the face of capacity shortfalls, many managers in TOC companies jump first to purchase capacity before fully exploiting a constraint resource. The exploit step is highly focused and seeks to eliminate the most devastating form of waste—waste that not only causes unnecessary actions or cost, but also that inhibits throughput. Items like setup time on the constraint, quality problems that impact the constraint, or material shortages are attacked directly in an attempt to improve the constraint output.

*Lean Thinking* 15 outlines seven types of *muda*:

- Human activity which absorbs resources but creates no value.
- Mistakes requiring rectification.
- Production of items no one wants so that inventories pile up.
- Processing steps that aren’t needed.
- Movement of employees or goods from one place to another without purpose.
- Idle capacity – groups of people waiting downstream because an upstream activity has not delivered on time.
- Goods and services that don’t meet the needs of the customer.

In a lean implementation, all instances of *muda* reduction are celebrated. This way of thinking is very appealing to those who live in the Cost World. *Unfortunately, very few successful companies have saved their way into prosperity.* While eliminating waste is a noble task, not all waste is created equal. We believe that as with all things, the prioritization of activities should be based on the impact of the action on the company’s
bottom line (via impact on throughput, inventory, and operating expense) as well as the impact on any necessary conditions.

**Value Stream Definition**

While both TOC and Lean look at the enterprise from the perspective of the chain of events that create the value proposition for a customer, they differ greatly in how the boundaries are defined in the value stream itself. Lean enterprises organize themselves around specific products. The danger in this perspective occurs when the organization’s resources are shared among several products.

Taking time out of one step in a product’s process will not enable the organization to increase throughput, unless that time is constraint time. Taking time out of one step in a product’s process will not enable the organization to reduce inventory, unless that time is taken from a resource that was responsible for major fluctuations in the buffer. Taking time out of one step in a product’s process will not enable the organization to reduce operating expense, unless (unfortunately) a person is laid off as a result.

A less-heralded aspect of lean applies to the reorganization and dedication of resources to a product line. The practice of dedicating resources to specific products will lead to a potential underutilization of machines and equipment. Unfortunately, this underutilization sometimes is sufficient to convert non-constraints into constraints. Although Lean advocates suggest reducing the size of special purpose high-volume machines, it is not always practical to do this. In the absence of downsized machines or operations, systems still have to operate using current resources.

The TOC perspective of the value stream includes resources that are shared across value streams. Thus, a value stream (also referred to as value lane, throughput lane, or throughput channel) in a TOC enterprise will often encompass several products that are produced through the interaction of resources common to those products.

One of the first steps in a TOC implementation is to define the system to be improved. Although this may initially limit system definition to something less than the entire value chain, the other parts of the value chain—design, order entry, manufacturing—will be addressed as needed. Nonetheless, TOC proponents would agree that the entire value stream should be both considered and improved. The value-chain
The concept was first tested in the TOC world in the apparel industry and is outlined in a book called *Tough Fabric*.1

Meanwhile, the logistics and politics of analyzing an entire value chain can be daunting and are often limiting factors in even getting started. Value chains frequently include multiple functions, different companies, and/or different divisions—all with different priorities and measures. The entire chain may never agree to improve until it is severely damaged by the competition. The question to answer may be, *is there value, at least initially, to improving manufacturing (or distribution, or engineering) in isolation and accept the other parts of the system as a given?* We believe there is. Any systems analysis will highlight those areas that constrain operations (e.g. batch sizes driven by slow changeovers driven by lack of engineering of dies; or quality problems due to poor manufacturing design.) In the best of all worlds, dealing with the larger system is preferred. However, the complexity of analyzing systems interactions in large systems should never be underestimated. The concept of narrowing the focus to a specific value stream (inclusive of shared resources) is a systems view in the best sense of the concept.

**Inventory**

TOC’s perspective on inventory is simple—the sole purpose of inventory is to protect throughput—period. Under some circumstances, it makes sense to maintain a buffer inventory to protect against variation in upstream processes. This inventory makes up the buffer portion of the drum-buffer-approach to synchronous flow. Lean takes a somewhat less charitable view toward buffer inventory, considering all inventory waste. This view underlies the concept of single piece flow.

Both philosophies advocate reducing the variability that makes buffer inventories necessary in the first place. One major difference is that TOC will leave the buffer in place *until* the variability is reduced. Lean removes all buffers and attacks the variability as it visibly surfaces.

If the system’s constraint is internal to a cell’s operations, the concept of no buffer will always cause the constraint to be starved at least part of the time. This is due to the reality of variability in any production system. One concern is that proponents of lean—seeing the low hanging fruit of inventory reduction—slash WIP inventories significantly without close examination of the impact on throughput. While in some cases, inventory is
truly waste, in other cases, it provides a crucial role in keeping the constraint operating during flow disruptions upstream. Ideally, the disruptions that cause inventory to be needed can be removed. However, inventory reduction is a process that could involve prioritizing many different improvement efforts. During the cleanup period, and maybe even after, it will be necessary to maintain strategically located buffers to protect the system’s throughput.

In the face of a less-than-perfect system, buffer management provides a means to prioritize improvement efforts. The foundation of buffer management is the recording and analysis of causes of shortages in the buffers that threaten to shut down the constraint. This analysis pinpoints areas most in need of improvement efforts. By starting with those items that frequently cause penetration into the buffer will both protect throughput as well as allow the buffer size to be reduced after things are improved.

**Capacity**

As with idle inventory, Lean views idle capacity as *muda*. “By design, flow systems have an everything-works-or-nothing-works quality which must be respected and anticipated.”\(^{15}\) Frankly, these are the words of someone who has not had to face the customer or boss as the bearer of bad news that their order was not shipped due to the fact that *our system design makes interdependencies crucial—everything works or nothing works*. We would imagine this approach would be hard to sell to the grizzled manufacturing manager whose current system is currently fraught with variability on every hand.

A system in which everything and everybody are working requires:

- Cross-trained workers—it will be impossible to perfectly balance lines. Workers will have to move.
- Machinery that is 100% available and accurate (TPM)
- Rigorous standardization of work by the work team
- Mistake proofing—design of work that is totally fool proof

Even when each of the above elements is improved dramatically, we have yet to discover a manufacturing system in which they were improved totally. Variability exists.
Period. A non-chaotic, predictable system, in which an ongoing, proactive process of continuous improvement progresses, must have protective capacity. That means some non-producing time is sought, expected, and celebrated.

As *muda* is eliminated, what happens to the people no longer needed for their traditional tasks? Most of the marketing we encounter for lean emphasizes head count reduction. While such activity may provide a very short-term positive impact on the bottom line, such practice will not yield long-term improvement. *Protecting jobs by finding people other productive tasks is a central part of any successful lean transition.*

(p.63) Unless activity is undertaken to increase throughput, organizations will continue to “have no choice” but to eliminate jobs. Eliminating the jobs of those who led the improvement inevitably leads to a “process of one-time improvement.” It also gives away one of the hardest things to replace—a trained worker.

**Cost**

TOC advocates view cost from a bizarre perspective by asking the question, *What is the impact on throughput of adding this cost?* In many cases—especially those where money or manpower is added to a constraint—the resulting analysis makes the decision extremely simple. Case in point. There once was a company whose engineering department had a backlog of over two years of projects in support of the plant’s production lines. Manning restrictions of corporate cost reduction programs prevented hiring even one more engineer. This is, by the way, a perfectly defensible cost reduction strategy, after all engineers are expensive. However, at the same time, the queue of engineering projects contained relatively quick, but lower priority projects, which would significantly improve constraint output—which in turn would increase line output. The market wanted more products and the throughput associated with any additional output was nothing short of phenomenal. One project—designed to increase the calibration speed (the constraint on the line) would have allowed the line to produce two additional units per hour—production that could be easily sold to an eager market. There is approximately $500/unit in throughput associated with each unit. Say that, for example you must pay as much as $100,000 per year to hire an electrical engineer with the needed skills. Should the company hire the engineer?
The TOC based decision would compare the $100,000 expense with the throughput that can be reasonably associated with the hiring. If the money for an additional EE were spent what is the impact on T and I? Completing this one project would allow the line to produce two additional units per hour—at $500 throughput each, that’s $1,000 per hour that won’t be there until the project is complete. This project alone would pay back the engineer’s annual salary in 100 hours. Four days—that’s not a bad payback period on a line that runs 24 hours per day.

The reality: The expenditure of $100,000 was not allowed.

**The Danger of Partial Understanding**

Having had discussions with numerous managers over the years who professed to be practitioners of the latest management approach, we have both seen huge misapplications of fundamentally sound ideas. It doesn’t matter if it was MRP, TQM, JIT, TOC, DFT, or Lean—like with many things “a little knowledge is a dangerous thing.” One manager, (a Harvard MBA by the way) proudly pointed out that his inventory system combined JIT and MRP. The JIT parts were produced in 6-month batch sizes (this is JIT?). The large “JIT” production runs tied up critical resources for several days at the expense of orders that were needed to ship that very day. But somehow in his mind, he had implemented JIT. His situation brings to mind Senge’s statement “Today’s problems are caused by yesterday’s “solutions”” (Senge, 1990, p.57).

TOC has its share of misapplications as well. We have also talked with managers who pride themselves in daily or weekly finding and smashing the ever-changing constraint. This is certainly not TOC.

**Lean’s Impact on Traditional Measures: Don’t Ask, Don’t Tell?**

One of the questions that remains unanswered is the impact of organizations’ measurement systems on lean implementations. The cost accounting battleground should be faced in both settings. Little has been written about “lean accounting.” However, Womack and Jones\textsuperscript{15} may have understated the difficulty in changing decades of traditional cost world ideas. They implied that an organization’s own accounting group should be able to figure out how to answer the question “what kind of management
accounting system would cause our product leaders to always do the right (lean) thing? One company cited in *Lean Thinking*\(^{15}\) went to the extreme. “Standard cost and variance analysis were declared dead as concepts immediately…” (p. 136). While this is laudable, it isn’t every day that the financial arm of the organization rolls over as easily.

Direct challenges to the product cost bureaucracy in large companies are often major undertakings. TOC advocates have been fighting this battle for some 15 years and are just now making inroads. *The Theory of Constraints and Its Implications for Management Accounting*\(^9\) provides an excellent source of information concerning how current accounting systems lead to erroneous decisions. Many of the lessons learned in TOC can be applied to lean implementations as well.
Section Five: TOC and Lean’s Synergy – Capturing the Best of Both Worlds

As TOC practitioners, there are many parts of lean thinking that are completely consistent with our view of the systems that make up our world. It can be enlightening to evaluate lean practices using TOC measures.

TOC’s prerequisites and five focusing steps provide a highly effective paradigm for managing the enterprise and focusing its approach to improvement. Combining many of the Lean practices and principles strengthens the paradigm even more. It is this combination that enables organizations to improve and sustain greater benefits. We suggest the following:

- **Adopt the Throughput World perspective.** The bottom line is the purpose of inventory and operating expense is to create throughput. People will rally around and stick with a mutually beneficial common goal of growth much easier than they will a corporate goal of cost cutting. In addition, the Throughput World paradigm forces people to look outside their own box and find out how their organization creates value, as well as how the organization can create greater value for its customers. It puts people in relation with their customers, and it helps to take away the fear of being rewarded for improving the company with a pinkslip.

- **Define the system to be improved, its purpose, and the measures of its purpose.** We recommend that you start with your system—the system that you control—and gain internal control before trying to change your customers and/or vendors. Our experience is that as you gain internal control, you gain the credibility needed to influence your customers or vendors in the right direction.

- **Identify the system’s constraint.** Use the *kaikaku* process to define the current process flow. Chart the process and eliminate the steps that are totally unnecessary and very easy to eliminate. Not to reduce cost, but rather to quickly reduce the number of dependencies and increase protective capacity. Fewer dependencies mean less opportunity for Murphy, less Murphy and increased protective capacity means less disruption to the system, providing for shorter buffer times.
• **Decide how to exploit the system’s constraint.** The process of exploiting the constraint is a perfect application for lean techniques. The system’s constraint is that resource which limits the output of the entire system. We frequently hear statements, like *We need to buy another machine or hire another worker; we just don’t have enough capacity.* No matter what the team says concerning a resource being maxed out, never take their word for it! Always observe the resource with someone who knows the process. Invariably, there is a tremendous amount of waste associated with constraint operations. Waste associated with setup. Waste associated with skilled personnel performing unskilled (and sometimes menial) tasks while at the same time the skilled workers are in short supply. There are scores of reasons constraint operations may be limited. Constraint operations should be the primary target for breakthrough kaizen sessions. It is often useful to establish a throughput per hour value for constraint operations. What is the throughput associated with the resource when it is operating? This dollar value can be as much as several thousand dollars per hour. Knowing this dollar value often helps modify the perspective on what changes make economic sense. Most organizations would refuse to hire a person who would be kept busy only 25% of the time—two hours out of an eight-hour day. However, if that two hours of work increased the amount of time worked on the constraint (at say $1300/hour throughput) then the decision to add this “unproductive and under worked employee” would be viewed in a much more favorable light.

• **Subordinate to the system’s constraint.** The concept of subordination is included in both lean and TOC approaches. A major part of subordination in a DBR system is the idea of limiting the release of material into the system to avoid overloading the constraint. This is accomplished in lean systems by the use of kanbans, which are analogous to a series of short “ropes” (communication links) between all resources. The concept of the rope in DBR, ties the material release all the way back to the gating operation. Intermediate operations are to process parts as rapidly as they can once they appear at the work center.

There are many other activities that should be subordinated to the constraint as well. Maintenance should subordinate to the constraint operation. That means dropping everything (yes even if they are in the middle of an operation on a non-constraint) and
rushing to get the constraint back on line. The same goes for analyzing quality problems that impact constraint output or for reducing constraint setup time. In short, the entire organization is subordinated to the constraint. This is based on the knowledge that increased output at the constraint increases output for the entire system.

- **Elevate the constraint.** As mentioned earlier, this is often the first step taken when a constraint is uncovered. In reality, many times it is never needed and organizations incur unnecessary spending. In such an instance, operating expense and or investment is higher than necessary. Many times the exploit and subordinate steps increase the capacity of the constraint by a significant amount.

  The process of elevating the constraint increases the overall resource capacity to a higher level, usually by adding capital equipment. The equipment doesn’t even have to be new in that alternate routings often offer an effective means to offload the constraint. Incidentally, in “Cost World” companies, if a routing change increases the *total* labor or machine content of a part (as when a slower non-constraint operation is used to offload a constraint operation) the apparent impact of such change on product price calculations will result in disapproving the change. This is often true even if the change has a significant positive impact on the company’s bottom line!

- **Avoid inertia. Identify the next constraint if the constraint is broken.** Organizations will always have at least one physical constraint—raw materials, internal resources, or lack of markets. As one is broken, the constraint will shift to another place in the organization. Knowing this, and never resting on the laurels of success, keeps the process of continuous improvement focused and goal directed.
Summary

TOC’s concept of ongoing improvement by focusing on and managing constraints may not be as inspiring as the Lean goal of seeking perfection. However, the complexity of modern organizations and systems leaves managers with an almost unlimited number of things to improve. The magnitude of the task is sufficient to paralyze even the most conscientious manager. Meanwhile, in reality, only a handful of those hundreds of potential improvements will make a real difference in achieving an organization’s goal. TOC’s constraint-focused approach is both logical and pragmatic. Identifying and addressing the constraints provides the fastest and lowest cost means for increasing the throughput of any organization.

Meanwhile, lean offers a different approach to improvement—that of trimming the organizational fat by removing waste at every turn. To the degree that an organization is morbidly obese (imagine the most absurdly bureaucratic system that you can think of) the idea of ripping the waste out of the system is not only easy, but also, has great appeal. However, in most organizations, waste reduction takes concerted effort over a long period of time. Where do you start? Removing variability is neither a trivial task nor is it quickly accomplished in most settings. The tools invoked during lean implementations are useful. However, they are far more useful when they are focused on the organization’s limiting factors.
References


The theory of constraints (TOC) is a management paradigm that views any manageable system as being limited in achieving more of its goals by a very small number of constraints. There is always at least one constraint, and TOC uses a focusing process to identify the constraint and restructure the rest of the organization around it. TOC adopts the common idiom "a chain is no stronger than its weakest link". This means that processes, organizations, etc., are vulnerable because the weakest person or part The Theory of Constraints and Lean Manufacturing are both systematic methods for improving manufacturing effectiveness. However, they have very different approaches: The Theory of Constraints focuses on identifying and removing constraints that limit throughput. Therefore, successful application tends to increase manufacturing capacity. Lean Manufacturing focuses on eliminating waste from the manufacturing process. Therefore, successful application tends to reduce manufacturing costs. Both methodologies have a strong customer focus and are capable of transforming companies to be faster, strong