

Using Stock, Time and Capacity Buffers to Build the World's Most Resilient Supply Chains

Dawid B Janse van Rensburg

INTRODUCTION

Without analyzing the subject field of resilience (broader subject area is risk management) too deeply, the objective of this paper is to explain/illustrate how stock, time and capacity buffers can be used and combined to build highly resilient supply chains.

However, the context of the paper will be positioned within the subject area of risk management, seeing that resilience has to do with vulnerability of supply chains when exposed to large scale disruption, mostly external.

The paper will therefore aim to set the context for supply chain resilience (what it is), the factors that makes supply chains vulnerable, and then most importantly, how to build a resilient supply chain.

The case is made that agile supply chains almost certainly ensures a resilient supply chain, and then the author will explain how buffers are used to build agile supply chains, and therefore resilient supply chains.

DEFINITION OF A RESILIENT SUPPLY CHAIN

According to dictionary.com (2014) the pure definition of resilience is as follows:

“Noun: the power or ability to return to the original form, position, etc., after being bent, compressed, or stretched; elasticity.”

And considering that the subject area is supply chain management:

According to a Global Risks Report, resilience is the ability of a global supply chain to reorganize and deliver its core function continually, despite the impact of external and or internal shocks to the system.

“[...] being better positioned than competitors to deal with – and even gain advantage from – disruptions” is a perspective from Sheffi and Yossi in a Harvard Business Review article on building a resilient supply chain (2005).

From a macro perspective, “Resilience is about building capacity, through the collective and simultaneous efforts of those in and out of government towards a shared result”, according to an article written by the Pacific Institute for Public Policy (2001).

Another view on economic resilience to shocks according to the Organisation for Economic Co-operation and Development (2007): “[The] ability to maintain output close to potential in the aftermath of shocks.”

Lastly, another perspective from Sheffi and Yossi (2008): “[...] the ability to bounce back from large-scale disruptions”

Keywords from the above definitions of resilient supply chains include the following:

Reorganise, bounce back, return to original form – clearly to emphasis is here on the ability to recover quickly from disruptions.

Deliver core function continually, maintain output – to be able to maintain the product supply or service delivery to customers and/or consumers.

Competitive advantage – probably true that those supply chains that already have a competitive advantage over the competition will be able to recover the quickest from disruptions.

Building capacity – resilience follows after the required capacity has been put in place.

FACTORS THAT DISRUPT SUPPLY CHAINS

All too often resilience is viewed in the context of large-scale external disruptions, but the factors that more frequently test the resilience levels of supply chains are sometimes very close to home. Following are the top five vulnerabilities that disrupt supply chains, as identified by supply chain experts in a recent survey conducted by the World Economic Forum (2013):



Figure 1. The Top Five Vulnerabilities that Disrupt Supply Chains.

In the WEF report referred to above, it is stated that disruptions to the supply chain can harm organizations, especially private and listed companies, enormously. Research has indicated that the average drop in shareholder value returned is around 7%, and in many instances even start declining before a public announcement about the disruption is made. In most cases, it takes months for stock prices to return to original values.

WHY ARE SUPPLY CHAINS SO VULNERABLE TO DISRUPTIONS TODAY?

Should there be a scale that indicates that supply chains are generally more exposed to disruptions today than previously, then there must be a set of reasons for this phenomena. According to Khan (2005) supply chains are perhaps more susceptible to disruptions than previously, due to the following factors:

- Widespread adoption of 'lean' practices
- The move to off-shore manufacturing and sourcing
- Out-sourcing and reduction in the supplier base
- Global consolidation of suppliers
- Centralised production and distribution

The following illustration shows that resilience is an optimum point on an arbitrary scale of "leanness", where leanness refers to the level of waste that has been removed from the supply chain. Such "waste" will almost certainly be in the reduction of inventory, cycle-times, and capacity. The argument is that a supply chain that is too lean, logically becomes highly vulnerable.

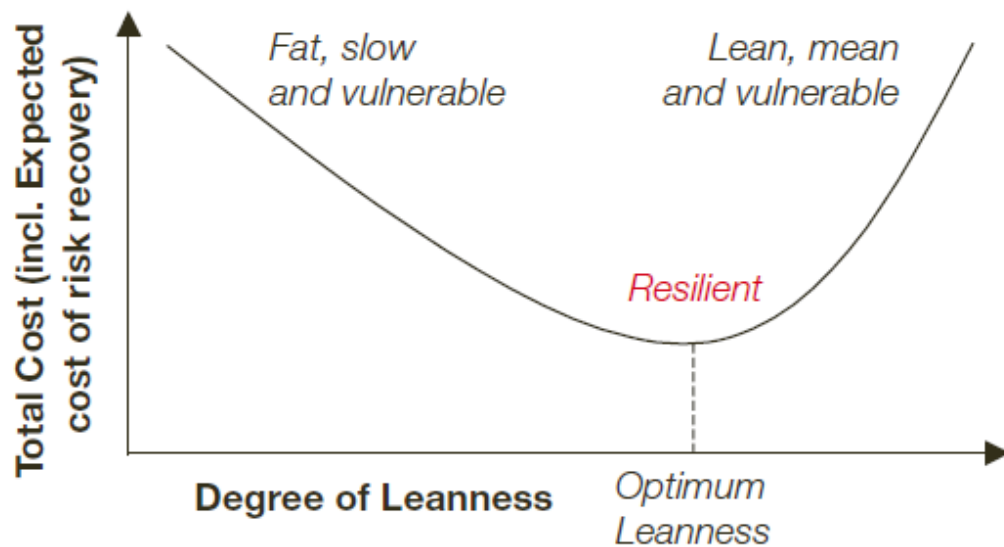


Figure 2. Creating Supply Chain Resilience through Lean Six Sigma (Christopher and Rutherford, 2004)

IDENTIFYING AREAS OF RISK

When designing and/or building a supply chain with the objective to make it resilient, it is important to consider risk areas, in order to be able to counter such risk areas.

We should strive to identify vulnerabilities by asking questions such as:

- What has disrupted operations in the past?
- What known weaknesses do we have?
- What 'near misses' have we experienced?
- What would be the effect of a shortage of a key material?
- What would be the effect of the loss of our distribution site?
- What would be the effect of the loss of a key supplier or customer?

Khan (2005) has identified the following sources of risk, which serves as a handy guideline to focus the process of risk identification:

Demand Risk

- Loss of major accounts
- Volatility of demand
- Concentration of customer base
- Short life cycles
- Innovative competitors

Supply Risk

- Dependency on key suppliers
- Consolidation in supply markets
- Quality and management issues arising from off-shore sourcing
- Potential disruption at second tier level
- Length and variability of replenishment lead-times

Process Risk

- Manufacturing yield variability
- Lengthy set-up times and inflexible processes
- Equipment reliability
- Limited capacity/bottlenecks
- Outsourcing key business processes

Network/Control Risk

- Asymmetric power relationships
- Poor visibility along the pipeline
- Inappropriate rules that distort demand
- Lack of collaborative planning and forecasts
- Bullwhip effects due to multiple echelons

Environment Risk

- Natural disasters
- Terrorism and war
- Regulatory changes
- Tax, duties and quotas
- Strikes

ROBUST VERSUS RESILIENT SUPPLY CHAINS

Within the subject area of resilience and supply chains, authors often emphasize the fact that a distinction must be drawn between a robust and a resilient supply chain. The important point is that a collection of robust processes that links together to form a supply chain is not necessarily a resilient supply chain.

In order to explore the difference between a robust and a resilient chain, the definitions are again referenced.

Robust: “Strong in constitution, hardy, or vigorous”

A robust process or chain enables a firm to manage regular fluctuations in demand efficiently under normal circumstances regardless of occurrence of a major disruption, but this does not in itself make a resilient supply chain.

Resilient: “The ability of a system to return to its original [or desired] state after being disturbed”

The core concept of resilience is that it (a) encourages a whole system perspective, (b) it explicitly accepts that disturbances happen, and (c) it implies adaptability to changing circumstances.

A robust process can be defined as “a process able to deal with reasonable variability”

A resilient supply chain can be defined as “a supply chain with the ability to recover quickly from unexpected events impacting supply chain performance”

AGILE SUPPLY CHAINS – LINK TO RESILIENT SUPPLY CHAINS

Based on the aforementioned statement of a resilient supply chain, the concept of an agile supply chain logically comes to mind, due to the emphasis on the ability to recover quickly from disruptions.

Agile supply chains are designed to respond rapidly to unpredictable change, and Kahn [n] lists the principles based upon which agile supply chains are designed:

- Very close connection to final marketplace
- Visibility of real demand
- High levels of synchronicity upstream and downstream
- Organisational focus on processes rather than functions
- Advanced level of collaborative planning with supply chain partners
- Continuous search for time compression opportunities

CRITERIA FOR A RESILIENT SUPPLY CHAIN

Based upon the principles of an agile supply chain, the following criteria, or key performance indicators (KPI's) can be defined:

After a disruption, the agile supply chain can return very quickly to a state where:

- Very high availability levels of product across the supply chain, i.e. very little out-of-stocks of own product/service
- Very high levels of end product delivery reliability, i.e. excellent due-date performance
- Replenishment based partnerships with resilient suppliers, i.e. based on the concepts of vendor managed inventory, suppliers are closely linked to the supply chain, and have full visibility of their product in the client's supply chain
- Short cash-to-cash cycle times
- Optimum stock levels (note, not low stock levels)
- Visibility of the spare capacity that exists at any point in time within the supply chain with emphasis on the capacity constrained resources.

HOW TO BUILD A RESILIENT SUPPLY CHAIN USING BUFFER MECHANISMS

In theory, the supply chain can only be as good as it was before the disruption, and not better (as a result of the disruption), and the state should ideally be returned as rapidly as possible. In order for this to be possible, the supply chain should ideally be in a very healthy state just before the disruption. Logically, the nature of disruptions is that the time of it taking place is unknown, and this therefore implies that the supply chain must always be in a healthy state.

A healthy state would describe a supply chain that meets the criteria as listed for an agile supply chain previously. In order to build and operate such a supply chain, appropriate levels of inventory at the right places are required, and must be maintained at an optimum level – not too much, and not too little.

Similarly, excessive lead times in the supply chain must be reduced, but again by not too much. As will be explained, appropriate lead times, coupled with an appropriate degree of loading on capacity constrained resources (e.g. in the factory) will ensure good flow through the supply chain, again resulting in faster cash-to-cash cycles.

Lastly, there is a requirement for capacity constrained resources, especially in a manufacturing environment, to maintain an optimum level of workload on such resources. This will be explored in the following section titled capacity buffers.

STOCK BUFFERS

In the Theory of Constraints (Schrage, 2010), the practice is to define a level of safety and constantly monitor how the safety is being used. This safety is called a **buffer**. In a distribution environment, the quantity to be maintained at the stock locations (including the central or plant warehouse (PWH) as well as regional warehouses) is defined as a **buffer size**, and this is a **stock** type buffer. The buffer size in a distribution environment (make-to-stock buffer size) is the number of units that needs to be maintained overall in the supply chain for the specific stock location for the specific stock keeping unit (SKU). For example – if the stock buffer size is 100 units and currently at the stock location there are 40 units, 60 units should be on order or on the way from the feeding

stock location to this one (the feeding stock location for the PWH is the factory). If those 60 units are not on the way, a replenishment order of 60 units should be issued immediately.

Note: different stock locations will have different buffers for the same SKU, since the supply and/or demand pattern might be different between them.

Buffer penetration is defined to be the number of missing units from the buffer divided by the buffer size. For the above example the buffer penetration for the stock at site is 60% $((100 - 40) / 100)$. The buffer size is divided into 3 equal zones. The buffer penetration sets the color of the buffer according to the different zones:

- Less than 33% buffer penetration: Green
- Between 33% and 67% buffer penetration: Yellow
- Between 67% and 100% buffer penetration: Red
- 100% buffer penetration: Black



Figure 3. Colour coding of a Stock Buffer.

The buffer penetration color gives an indication regarding the urgency of replenishing this stock:

- Green – the inventory at the consumption point is high – providing more than enough protection for now
- Yellow – the inventory at the consumption point is adequate – there is a need to order more units from the upstream supply chain
- Red – the inventory at the consumption point is at risk of depletion – units in transport/manufacturing (depending on which consumption point it is) - should be considered for expediting efforts and an urgent replenishment order must be put to the supplying source if nothing is available on the way to the consumption point
- Black – the stock has run out at the consumption point, meaning every hour passed at this stage is lost sales opportunities – this situation must be resolved ASAP as it represents real damage, especially at the most downstream links in the supply chain

There could be several buffer views on the same buffer, and this is sometimes referred to as the **Virtual Buffer** concept. Let's examine the following:

The buffer size for this SKU at the stock location is 100 units. There are 25 units in stock at the stock location, and there is a shipment on the way from the PWH to the stock location for 25 units. The virtual buffer penetration is taking into account the aggregated stock of downstream stocks, i.e. the 50 units.

The priority is determined by the virtual buffer penetration of the next downstream stock. The virtual buffer penetration provides a very powerful tool – there is a single measurement with different views, but all the decision makers involved in the supply chain can get the priority according to the need:

- The manager of the stocks at the stock location can see clearly that the priority of this SKU is red (75% Buffer Penetration) – he needs to find out how to get more stock of this SKU ASAP
- The transportation manager can get the priority of the shipments – what shipments need to be expedited – in this case the shipment needs to be expedited (75% buffer penetration)
- The Plant WH manager can get the replenishment priority of this SKU at this stock location – in this case 50% of the buffer size of the SKU in this stock location and the priority of this replenishment shipment is yellow (50% buffer penetration).

Dynamic Buffer Management

TOC advocates simple and straightforward methods to use in an inventory replenishment, or distribution environment. Due to the many shortcomings of forecasting techniques, which can be described as a “push” supply chain, an excellent alternative is the concept of replenishment based on consumption. Forecasting techniques requires deep levels of expertise, and in most cases the forecast is eventually highly inaccurate. The TOC logic dynamically measures the actual usage of the stocks and readjusts the inventory levels accordingly. This method is referred to in TOC literature as **Dynamic Buffer Management (DBM)**.

By monitoring the buffer penetration at each stock location for each product, it can be confirmed whether the buffer size for this product at this stock location is about right. The Dynamic Buffer Management (DBM) approach argues that by monitoring and adjusting the buffer sizes the "real" stock that needs to be kept can be easily and rather accurate determined at the site in order to cover for the demand, taking into consideration the supply side (how fast stock can be delivered to the stock location).

DBM concerns itself with two different occurrences – one is whether the buffer size is too large and the other is when the buffer size is too small. When trying to measure whether the buffer size is too high, the indication is when the buffer penetration at site of a stock keeping unit (SKU) in a certain stock location has been Too Much in the Green (TMG) – meaning being in the green for several consecutive days (when the cumulative green zone penetration above the yellow band exceeds a certain target, typically about 500% of the green zone). This implies that the current buffer is too high for the current demand pattern, at least for this time period, which suggests several alternatives:

- Demand has reduced
- The supply side has improved
- The initial buffer size was too high
- Demand fluctuates severely (and then the cumulative target percentage should be increased rather than decreasing the buffer). The default recommendation for handling the too much green is to decrease the buffer. The basic principle says that the buffer be decreased by 33% when needed, but this is a guideline and depends on several factors:
 - How rapidly buffers must be reduced once sensed that demand is reducing
 - How risky/important this SKU is perceived to be
 - How risky/important this stock location is perceived to be.

A very similar mechanism is used for determining whether the buffer is too low – determining whether this SKU in this stock location has been Too Much in the Red (TMR). However, the algorithm

usually differs, since in this case the algorithm needs to be very responsive to the depletion of stock, not like in the too much green case in which we would like to take his time and play it safe. The most basic algorithm for the TMR calculation is to determine whether an SKU is in the red for several days (again using a target for cumulative red zone penetration below the yellow zone, typically 100% of the red zone). Typical reasons for red zone penetration include the following:

- Demand has gone up (the preferred reason)
- The supply side response has deteriorated
- The initial buffer size was too low
- Demand fluctuates severely

The general rule in a TMR situation is to increase the buffer, the default being by 33%, and again this is just a guideline and each case may be dealt with differently.

After adjusting the buffer, the SKU needs to get into a "cooling period" in which no buffer suggestions in the same direction are given (until the system adjusts to the revised buffer size). This cooling period should be long enough to let the adjustment take place (the new quantities ordered to arrive to the stock location) and yet short enough so that a sudden real change in the market demand will not occur without someone noticing. For the TMR – the cooling period is a full replenishment time, and for the TMG the cooling period is letting the inventory at site cross over to the green from above (since lowering the buffer size probably caused the current inventory at site to be above the buffer size level).

Manufacturing priorities according to urgency in PWH

Manufacturers usually manufacture to order. That means that each work order on the shop floor is for a specific customer for a given due date. TOC for that environment prioritizes the production orders based on their due dates (for more details please refer to published literature on Simplified DBR).

When manufacturers embrace the TOC solution for distribution, another angle should be thought of. In this case the production orders are not for a specific customer, and are just covering for consumption from the PWH. Therefore, the right priority should be set not according to time, but rather in the same way the priority in the stock locations for the SKUs was defined – the best priority mechanism is to take the buffer penetration at the warehouse as the priority for the work order that needs to replenish the stock buffer. If there is more than one WO for the same SKU – the best priority mechanism is to take again the Virtual BP in the following way:

The priority of every WO is based on the virtual BP of the next WO in production (the one which was released before it) to get its' production priority. This ensures that the production is in line with the actual usage of the stock – if the stock is depleted fast the WO will be expedited through production and otherwise it will float in production on the excess capacity of the production system. Every entity in the supply chain is fully aligned and synchronized with the goal of the system – to be responsive to the actual consumption of stocks from the next link in order to create availability otherwise unattainable.

Rules for setting up initial buffer sizes

The first step in moving from push distribution to pull distribution is setting up the plant warehouse (PWH) and starting to build inventories to fill the initial stock buffers. The decision of what the initial stock buffer should be might seem a very complex decision – the amount of uncertainty is huge; so fear is very natural. A golden rule here is to derive quickly at a target level, and let the algorithm guide the process towards an optimum buffer size, as opposed to try and calculate a correct starting level. Since the DBM mechanism will adjust the buffers according to real consumption, the initial estimate must be more or less correct.

The recommended practice is to start with an initial estimate: taking the replenishment time from the source to the destination and multiplying it by the average consumption and by a factor of 1.5 for safety.

The replenishment time to use should be:

- For a production environment (plant warehouse) – taking the current quoted production lead time for this item (after implementing TOC in the manufacturing environment the lead time will usually be cut in half and then the DBM will automatically suggest lowering the buffer)
- For a transportation environment (central warehouse, regional warehouse and consumption points): mainly transportation time plus something to account for shipping only on certain days during the week.

A simple rule can be used also to determine whether an SKU has some yearly seasonality effects: if looking back on last year's consumption (and possibly the year before) one month's sales are more than twice the monthly average of the total sales (somewhere between 15-20%), this SKU should be defined as seasonal in that month. For the seasonal SKUs, a different initial buffer can be defined for the seasonal months and for the regular months (using the same rules stated above). The difference in the buffer sizes can be calculated and fed into a seasonality model – in which the buffers are set manually or automatically by software before the season starts/ends. The monthly/weekly seasonality can be detected using a similar mechanism.

TIME BUFFERS

For Make-to-Order (MTO) environments, a similar concept of buffer penetration is used as for stock buffers. In other words, the lead time for manufacturing or ordering from a supplier is used, and divided in the colour zones that is used for stock buffer penetration:

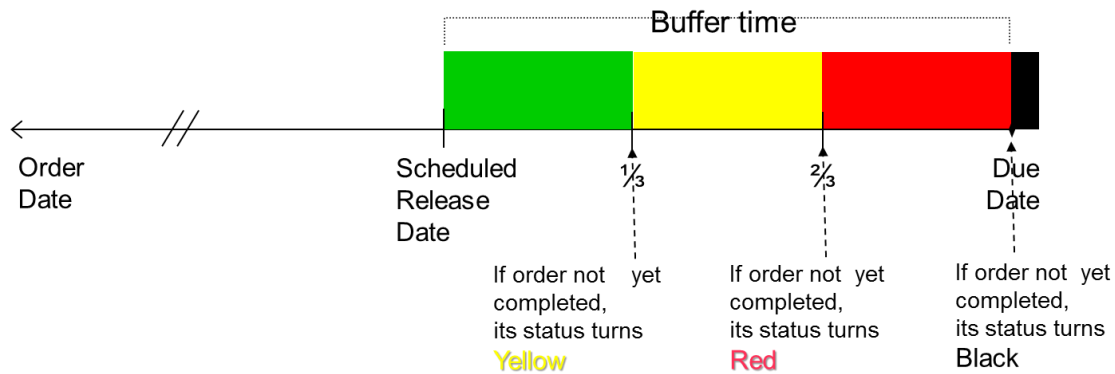


Figure 4. Colour Depiction of a Time Buffer.

The principle here is again to prioritize the MTO work orders with the deepest penetration into the time buffer. A WO with a red priority will therefore be produced before a WO with green priority, and so on.

In TOC practice, lead times are traditionally reduced substantially, and if the factory follows the scheduled release date recommendations, the end result is almost certainly that flow through the factory will be improved substantially.

The following diagram illustrates how the initial time buffer should be determined:

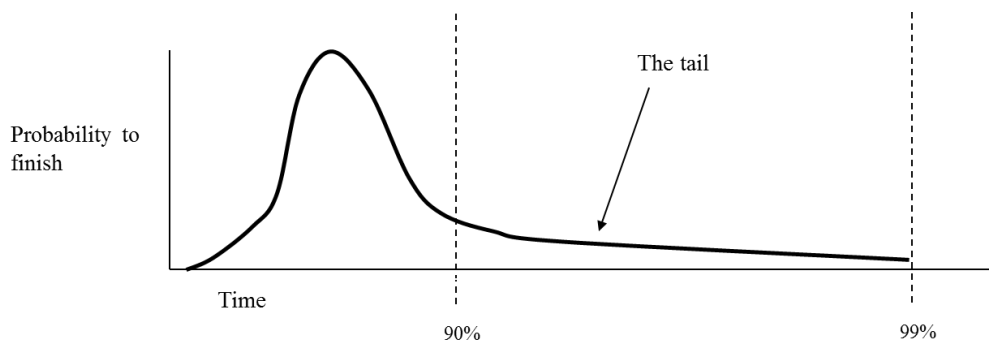


Figure 5. Distribution of Lead Time taken to complete Work Orders.

The lead time in which most (90%) MTO WO's are completed is determined. Almost always there will be a long "tail", i.e. the time that it takes to complete most WO's.

The red zone (i.e. 33% of the time buffer) would typically be set from the 90% mark, i.e. the WO's falling in this red zone will be expedited, and the end result is a reduction in the number of WO's (approximately 10% at the outset) that is finished too late, or close to the due date.

Again using the simple colour priority system, the end result is a reduction in lead times, but not to the extent that due dates are compromised, as well as dramatically improved due-date performance.

CAPACITY BUFFERS

The application of TOC in the production environment dictates that capacity constrained resources (CCR) be identified and managed carefully. A CCR is typically a machine or work area where bottlenecks first appear, and as a result very often determine the flow through the facility, i.e. causing lead times to extent.

It has been proven that in production systems the flow is substantially impacted from the 80% loading factor onwards. This rule, which is derived from Little's Law, holds that in most production environments, inventories (work in progress) will start increasing, slowing down flow through the system, and extending lead times.

TOC practice is therefore to monitor the CCR's in the production environment carefully, and when the workload starts to approach the 80% mark, the release of new or additional work in the system is delayed. This will ensure that current WO's in the system are not impacted, and become late.

Calculating the workload is based upon the average rate of units produced in a certain time unit, e.g. hours or days. Typically, all MTS and MTO work orders will be included in the workload calculation, and the criteria for MTO work orders is that the recommended release date must fall within the time horizon that is selected.



Figure 6. Illustration of a Capacity Buffer (Inherent Simplicity (2013))

Having visibility of the workload on the CCR's in the factory typically enables planners and sales teams to decide if another sales order can be accepted or not, and what the implication will be on the due dates of other work orders should a work order with a very ambitious due date be accepted (or even normal lead time).

Again, TOC practice dictates that the promised due dates for new orders in the factory first be confirmed in the case where loading is excessive (i.e. above 80 or more percent).

Also, in order to maintain a healthy level of agility, CCR resources (and by implication other resources) should always have a level of spare capacity (approximately 20%). This statement is

central in the debate between the conflict of managing efficiencies in the factory, and due-date performance, or flow.

PUTTING THE SUPPLY CHAIN TOGETHER

Designing an agile supply chain must ideally incorporate the use of these three buffers as described.

Stock buffers are positioned at strategic stock locations in order to ensure high availability of product at all times, and monitored and adjusted dynamically on a continuous basis to prevent overstocking. This also applies to the management of raw materials which qualifies to be kept in stock at all times. A single priority system comprising of colours as described will then ensure that the supply chain can be maintained at a healthy level, and almost certainly with an adequate level of protection against disruptions.

Time buffers are used in the production and purchasing areas, and monitored constantly on the same colour coded priority system, eliminating conflicting priorities across the supply chain.

Capacity buffers are designed and monitored to ensure that the reliability of due-date promises are maintained at a high level.

The diagram below illustrates how the combination of these buffers are designed into the system to ensure a resilient supply chain:

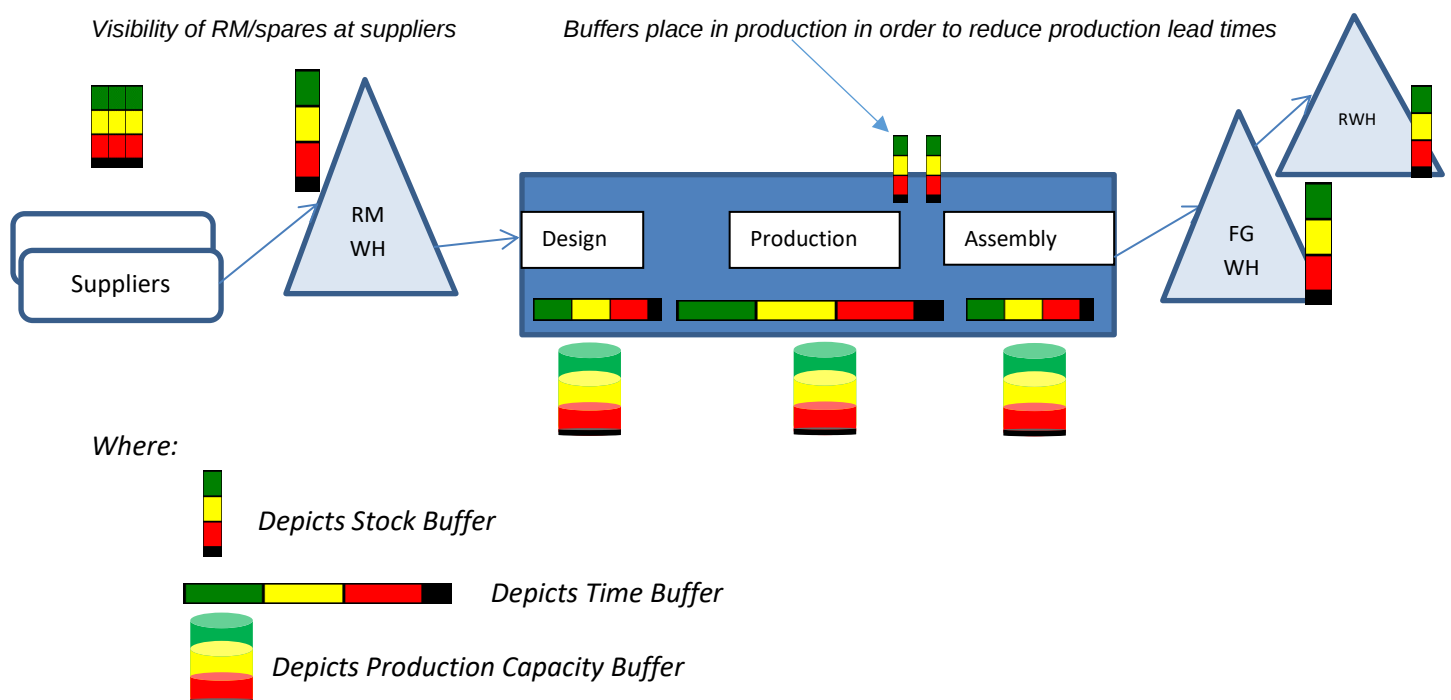


Figure 7. Typical deployment of Stock, Time and Capacity Buffers in a Supply Chain build for Resilience.

CASE STUDIES

The author has been involved in the deployment of the buffer concepts in a number of supply chains in South Africa. These have been across various industries, including footwear, building supplies, FMCG, stationary, printing, and others.

The following improvements have been documented in research study that was printed in book form by Mabin and Balderstone (2009). Figures represent the mean improvements of eighty companies documented in the study:

- Lead times: 70% reduction
- Due-Date-performance: 44% up
- Inventory levels: 49% reduction
- Revenue: 63% increase

Similar remarkable results have been achieved in the SA based implementations.

CONCLUSION

The objective of the paper is to demonstrate how stock, time and capacity buffers can be combined in the design of a supply chain to ensure a high level of agility, and by implication resilience.

The assumption is that resilience is linked, and actually ensured by having an agile supply chain.

References

Christopher, M & Rutherford, C, 2004, Creating Supply Chain Resilience through Lean Six Sigma; www.criticaleye.net, June-Aug 2004.

Global Risks Report. *World Economic Forum*, 2011.

Khan, O., 2005, *Managing the Supply Chain in an age of Uncertainty*, Cranfield University School of Management.

Mabin and Balderstone, 2000, *The World of Theory of Constraints*, St. Lucie Press.

Organisation for Economic Co-operation and Development, 2007, *Structural policies and economic resilience to shocks*.

Pacific Institute of Public Policy, 2011, *Food for Thought: Exploring Food Security in the Pacific*.

Sheffi and Yossi, 2005, Building a Resilient Supply Chain, *Harvard Business Review*. *Supply Chain Strategy*.

Sheffi, Yossi, 2008. *Resilience – What is it and how to achieve it*. http://web.mit.edu/scresponse/repository/Sheffi_Congressional_Testimony.pdf. Last accessed: 20/12/2012.

Schragenheim, A, 2007, *Managing Distribution According to TOC Principles, Inherent Simplicity*, Tel Aviv, Israel.

World Economic Forum, 2011, *Global Risks Report*.

World Economic Forum, 2013, *Industry Agenda: Building Resilience in Supply Chains*, January 2013.

www.dictionary.com; last visited on 1 May 2014.