Wysowl Pty Ltd
ACN 010 677 022

18 Miller Street
Chermside QLD 4032
Ph: Intl+ 61 7 3359 9918
Mob: Intl+ 61 407 149 966
john@wysowl.com.au
www.wysowl.com.au

Wysowl Pty Ltd
Newsletter Number 25
December 2010

THE UTILISATION TRAP

By John McConnell

In many operations, scheduling a plant or a supply chain for maximum equipment utilisation is a terrible mistake. Such an approach to scheduling has the potential to actually increase costs.

BACKGROUND

This newsletter outlines some of the common mistakes when scheduling/planning for factories and supply chains. Too often, executives and managers embark on initiatives to improve performance without a good understanding of the science involved or the potential consequences of their decisions. To be successful as a manager or scheduler in a manufacturing or supply chain operation, it is essential to understand the fundamental relationships that exist between throughput volume, capacity/utilisation, inventory/Work-In-Progress (WIP) and cycle time.[1] [2]

A manufacturing example. At one factory the practice was to schedule production a week at a time. Customer demand and stock levels were two obvious key determinants for the schedule. However, once these had been factored in, any spare capacity was filled by adding high volume commodity type product to the schedule. As will be shown later, this guaranteed that WIP soared such that the aisles were clogged with trolleys of product, cycle time grew to intolerable levels and customer service levels were maintained only by holding very high levels of finished goods inventory.

A supply chain example. The distribution network for a construction materials company was attempting to reduce costs by reducing the total amount of product between the factories and the customers. This meant reducing inventories and WIP (in this case, WIP is product in transit). In the following months customer service levels fell dramatically and predictably.

When Stephen Hawking was writing *"A Brief History of Time"* his publisher told him that every equation he put in the book would halve sales. Eventually, Hawking settled for one equation, Einstein's famous description of the equivalence of mass and energy, $E = mc^2$. Often, books and articles about scheduling are riddled with equations. This is no surprise given that many of the concepts hail from Operations Research, a science with heavy mathematical and statistical elements. However, I have decided to follow

Hawking's lead and include only one equation, Little's Law, which is essential for any understanding of scheduling. Little's Law states: (1) (2)

Ave Throughput Volume = Ave Work In Progress/Ave Cycle Time

Of course, we can express this relationship many ways. In this case we could transpose the terms and create:

Work In Progress = Throughput Volume x Cycle Time

Because manufacturing and distribution can never *exactly* meet market demand some sort of buffer is unavoidable. It is possible to reduce buffers such as Work In Progress, but not to eliminate them. Other buffers are inventory (both raw material and finished goods) capacity/utilisation and time. For instance, if you decide that meeting very short lead times is necessary, inventory will be higher than if lead times are such that time exists to manufacture at least some of the product after the order is received. If you plan to make much of your product to order, greater capacity is necessary unless you are prepared to live with very slow delivery time. In this case, time is used as a buffer. The reasons for these tradeoffs will be explained soon. At this stage it is sufficient to make the point that managers are constantly faced with these tradeoffs. The problem arises when they do not understand these tradeoffs and the ramifications of their decisions.

The terms in Little's Law are averages. They must be, because some degree of variation is inevitable. It is this variation in volume of material flowing through the factory or supply chain that significantly determines performance. To illustrate, imagine a manufacturing process that is operating perfectly. There are no delays, no unplanned downtime and no rejects. Imagine that the throughput volume is a constant 500 units per hour, the WIP is 2,000 units and the cycle time is four hours. Now let us add some variability to this situation. Perhaps we introduce some rejects or unplanned downtime. By necessity, performance must degrade. Allowing downtime to occur is a departure from perfection, and cycle time must increase. In addition, rejects must reduce throughput volume because they consume resources and capacity but are not included in the throughput volume.

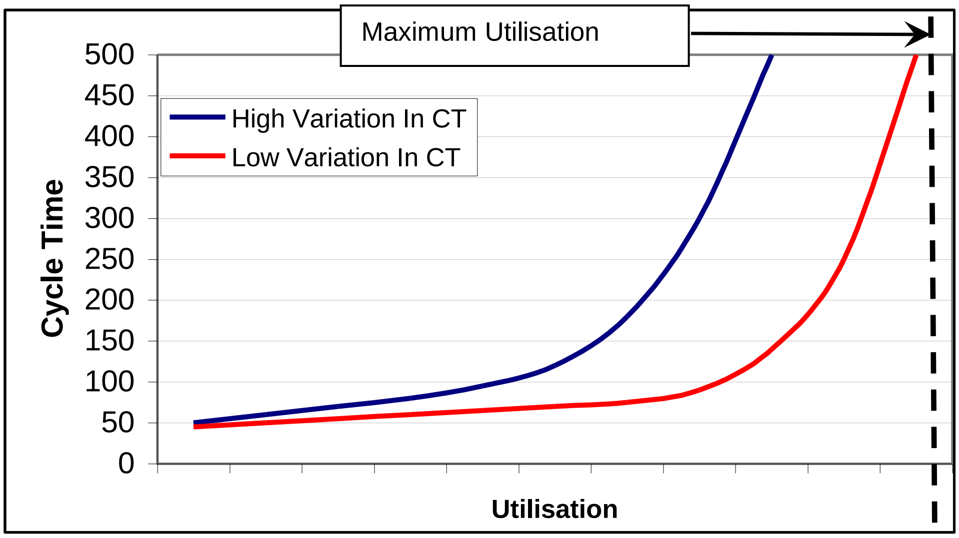
The first corollary to Little's Law states that as variation in the flow of material through the process increases, so too does cycle time. If variation and cycle time increase, either WIP must increase to keep throughput constant, or if WIP is held constant, throughput must fall (or some combination of both).

This is why so many world class operations drive rejects, delays and downtime to what sometimes seems like absurdly low levels. When costing rejects most companies include only the manufactured cost in their calculations. This is a mistake, because such an approach fails to cost the consequent variation which is either driving WIP up or choking throughput, and these can be considerable costs.

THE UTILISATION LAW

Nonetheless, it is fair to ask what all this has to do with scheduling. Many plants schedule to maximise utilisation because the managers and schedulers are attempting to reduce unit costs, but there is a price to pay for high utilisation. The reason for this is linked to variation and part of the answer can be seen at Figure 1. This chart illustrates what happens to cycle time as utilisation increases. As maximum utilisation is approached, cycle time increases. The greater is the variability, the longer is the cycle time for any given utilisation level (and usually, the higher will be the WIP). At low levels of utilisation, spare capacity exists to allow the various steps in operations to "catch up" after delays/disturbances. As maximum utilisation is approached, this capacity buffer begins to disappear, and the process responds by using other buffers, and it either builds WIP or lengthens cycle time, or both.

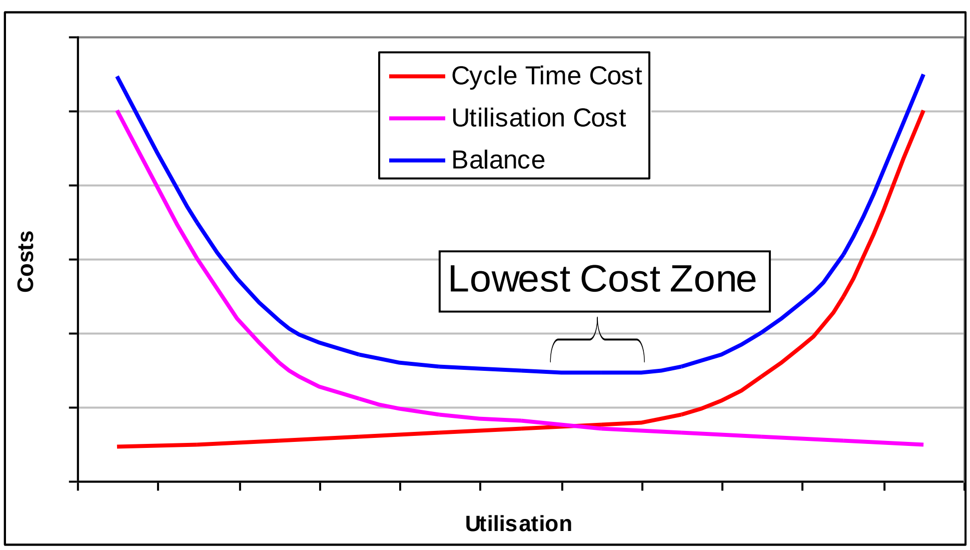
FIGURE 1
THE UTILISATION LAW IN ACTION



This is exactly what happened in the factory example given earlier. As yields fall from 100% towards 50%, random variation about the mean increases. This is part of the binomial law. Because in this case yields were low and variation was therefore high, maximising utilisation in scheduling lengthened cycle time which in turn exploded WIP. If throughput remains reasonably constant, increased cycle time **must** increase WIP, in accordance with Little's Law. All of this happened because the schedulers were trying to keep the factory running at maximum capacity as a cost control measure.

When we combine the costs of utilisation and cycle time/WIP a graph similar to that at Figure 2 is obtained. It shows that for any given degree of variation, there is an optimum utilisation zone if our objective is to reduce costs.

FIGURE 2
BALANCING UTILISATION AND CYCLE TIME COSTS



However, in some cases our overriding objective will not be to minimise costs. In some operations, a superior approach will be to raise both utilisation and cycle times/WIP to increase throughput in order to meet market demand because this is less expensive than adding capacity to increase throughput. This is a business decision, but it can best be made by understanding the tradeoffs involved.

Little's Law is a law, not just a good idea. It is as unforgiving as the law of gravity. Whether we understand it or not, it applies. The utilisation law is in fact a special case of Little's Law. The utilisation law states that as maximum utilisation is approached, cycle time (and usually WIP) increases. Either adding capacity or reducing utilization brings cycle times down just as rapidly, but by then the damage to customer confidence may be done. This utilisation law should be understood by all schedulers and managers, especially those managers in the financial control areas. Throughput, inventory/WIP, capacity/utilisation, cycle time and variation are inexorably interlinked as a system. They behave like a fluid enclosed in a balloon. If we press at one place, the balloon bulges somewhere else…so we press that bulge…and so on.

Many supply chains schedule the movement of materials through a process based on historical average performance of key events in the process. On paper, the schedule works, but often it fails in practice. Nearly always the cause of this failure is the inherit variation in the system. Because supply chains seldom have significant spare capacity, variability consumes resources and lengthens cycle times. In a supply chain impacted by significant variation, costs will increase due to wasted capacity and/or customer service levels will decline.

THE THEORY OF CONSTRAINTS

Every process has a bottleneck. This bottleneck determines the capacity of the entire system. Sometimes the bottleneck is a function of capacity, and sometimes it is caused by poor utilisation due to downtime etc. It is the utilisation of the bottlenecks that is critical. For all other steps in the process it is important only to utilise these steps such that they guarantee supply to the bottleneck. Any utilisation above this level only increases cycle time and adds WIP. For example, in an electronics factory the managers had spare hardware capacity in the assembly process, so what governed assembly capacity was the staffing level. This gave the schedulers and the managers the ability to quickly raise or lower capacity by changing the staffing levels by using more or less casual labour. The problem was testing. At four different stages of the process, every product was tested. These test stations were less flexible than the assembly process because there was less spare instrument capacity than there was spare assembly hardware. There was no point scheduling more throughput than the testing instruments could handle. Moreover, every minute of lost time at these test stations was permanently lost production.

Management decided that for the entire product range excellent customer service levels were essential. Many customers that purchased low cost commodity products also purchased much more complex high cost products. Both types of products needed to be delivered rapidly in order to be able to fill a minimum of 95% of all orders in-full-on-time. The numbers were crunched. For low value-added commodity type products (where labour was a high proportion of total cost) the best solution was to minimise labour costs and to bear the consequent longer cycle times, WIP and higher inventory of finished goods. For high value-added products (where labour was a low proportion of total cost) the chosen solution was to add capacity slightly to guarantee that WIP always existed in front of every test station, ensuring the test stations were never idle, to maximally staff the test stations, and to ensure an instrument fitter was available (in the factory on duty, not on call) to minimise repair/calibration time of the test instruments. This minimised finished goods inventory and provided rapid cycle time. Simultaneously, management launched a raft of activities to improve quality and yield and to reduce variation in the flow of material through the process.

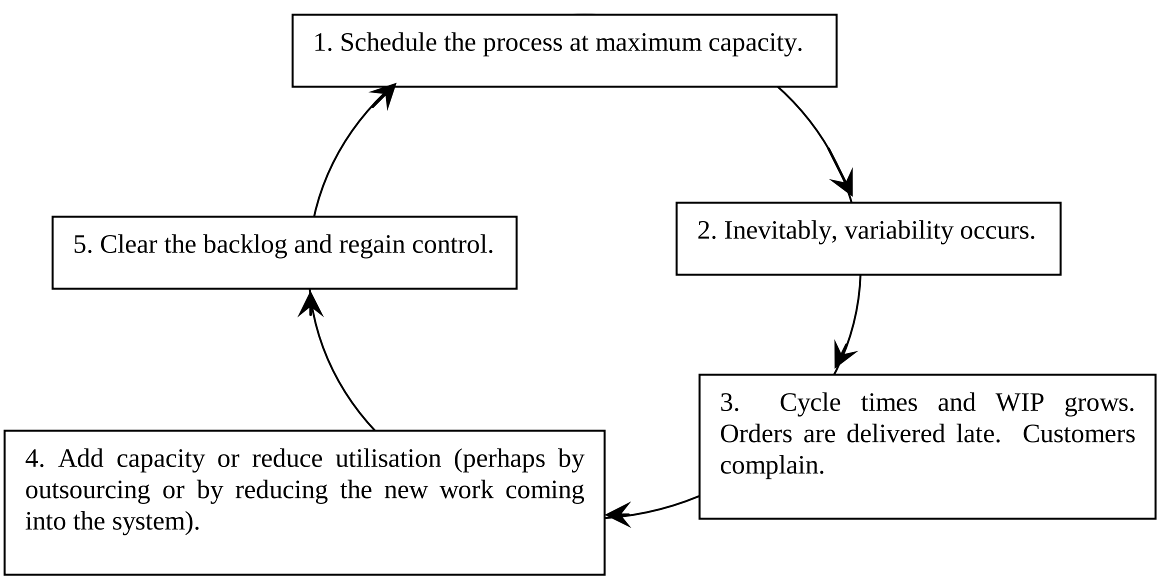
This is a good example of managers making informed management decisions to meet strategic objectives with maximum profitability. There are only two broad courses of action that make scientific and economic sense. The first is to find the optimum balance of the cycle time, capacity/utilisation and inventory/WIP buffers for any particular process. The other is to work relentlessly at reducing variation in the flow of material through the process. As variation is reduced, opportunities exist to either lower WIP/inventory or to increase throughput.

In one supply chain involving imported product, managers had been plagued with slow delivery times and supply failure as imported batches failed specification when tested. In addition, the various steps in the process showed significant variation around the planned schedule averages. These issues occasionally starved the downstream plants of feed. The most significant cost was gross lost margin when the business failed to deliver product to retailers and their competition grew market share at their expense. Management addressed the issues by introducing independent testing before batches of product were loaded onto a ship, by deliberately building some WIP buffers at two key locations and by reducing variation, especially in the early steps of the process. The increased WIP introduced some additional carrying costs, but it all but eliminated failure to deliver product to the market which had a much higher cost it terms of lost gross margin and lost customer satisfaction.

Problems tend to occur when managers and schedulers expect manufacturing and supply chain operations to behave in ways not permitted by natural laws. For example, if we schedule for close to maximum capacity/utilization we ought to expect that cycle time and WIP will grow significantly and plan accordingly.

TQM, Six Sigma, Kaizen, Kanban and Lean Manufacturing have all provided useful tools for improving business performance. However, the utility of these tools is severely limited in the hands of managers and schedulers who do not understand the implications of Little's Law and the Utilisation Law. In particular, it is common to find people skilled in the use of some of these tools, but who continue to try to defy these laws, at a significant cost. There is no substitute for knowledge. Figure 3 illustrates what happens in many operations when they try to maximise utilisation.

FIGURE 3
CONSEQUENCES OF MAXIMISING UTILISATION



LATE SCHEDULE CHANGES

Many plants constantly make schedule changes when they allow customers to make last minute changes to orders or when they allow customers to drop orders into a production schedule that is already underway. These are business decisions, but the manager who makes these decisions should understand that allowing these disturbances introduces variation (and probably increases utilisation). Therefore, either this manager needs to make additional capacity available or be prepared to ship orders late.

At one electronics factory the plant manager understands this. So too do his schedulers. As orders are dropped into the schedule at the last minute, additional staff hours (over and above that required to cover the additional order) are also scheduled across the line to provide the extra capacity necessary to absorb the additional variation. (They are in the fortunate position where the available production equipment has significant spare capacity, making employee hours on the line the constraining factor.) Their assessment is that the additional costs are a necessary evil, because they believe that their responsiveness to customers gives them a marketing advantage. The salient point here is that these managers/schedulers are making a carefully considered business decision. They are aware of the costs of their decisions and the tradeoffs involved. This is not always the case.

FORCING INVENTORIES/WIP DOWN

Some managers use Enterprise Resource Planning (ERP) systems to limit the amount of WIP that is allowed in the system or force down inventory levels in an attempt to reduce costs. In accordance with Little's Law customer service levels will fall unless there was excessive inventory present at the outset. If variability and cycle time are not reduced then throughput must decrease because there is less WIP in the system and as a consequence some capacity is wasted. This is a common problem in operations that become obsessed about reducing WIP. There is a big difference between being Lean and becoming anorexic, which is as bad for operations as it is for individuals. Managers making decisions about inventory and WIP levels need to understand how much WIP is needed to maintain the necessary throughput given the variability present.

Managers everywhere want minimum levels of inventory and WIP. Also they want to optimise production and to provide brilliant customer service levels; but in many quarters these things have become disconnected false idols. Sometimes, these idols are worshipped out of blind faith with inadequate understanding of the potential implications. Too often managers pursue what are seen as cost reduction opportunities, such as inventory and WIP reductions, unaware of the inherit dangers. Managers who want to be demand driven and agile are *de facto* stipulating that they require a very small time buffer. If they attempt to have a very small time buffer as well as very low inventory, and they do not reduce variability, a large increase in capacity buffer is necessary. Unless the spare capacity already exists, this can be a very expensive option. Having much spare capacity is in itself a form of waste, because the unused capital is not earning income. If managers do not add the necessary capacity and do not reduce variability, but nevertheless reduce inventory, the time buffer will increase, resulting in longer response times and a reduction in in-full-on-time delivery (IFOT, sometimes expressed as OTIF). The cost in lost customer satisfaction can be far greater than the losses due to carrying inventory. Again, it is a business decision.

First we must decide how we want to perform in the customer's hands (for example; product range/variety, promised lead times, quality performance) and to allocate priorities. Then we undertake the analysis to decide which mix of capacity/utilisation, inventory/WIP and cycle times provide the optimum solution. Then we invest in reducing variation (for example, quality improvement, improving the uniformity of flow of material through the process, maintenance procedures and spare parts holdings to reduce downtime, improve start-up and changeover processes)

For instance, if a management team chooses a certain level of in-full-on-time performance for a given volume throughput required, they can then manipulate the other key variables (capacity/utilisation, cycle time, inventory/WIP) and decide which scenario provides superior financial performance.

Alternately, they may need to optimise utilisation to meet market demand. In this case, they almost certainly will need to accept longer cycle times and higher WIP and finished goods inventory, at least until they have successfully conquered variation.

LARGE CONTINUOUS OPERATIONS

In large continuous operations such as mines, metallurgical plants, refineries and cement plants the problems outlined in this paper are often mitigated by large stockpiles. Unlike manufacturing, most mining/milling and similar operations have significant buffer stocks at least one and sometimes two or three places in the process. These stockpiles put considerable WIP in the process and do a pretty good job of absorbing most of the variation in the process. Unlike manufacturing, usually the major threat in mining is not the minute to minute or hour to hour variation; it is the long stops that can starve the mill of feed. Manufacturing seldom has the catastrophic failures that we see in mining (and in early stage continuous processing operations like cement and metallurgical plants), so whilst the concepts/laws are the same, the strategy will be different. Nevertheless, even in large continuous operations reducing variation by reducing stoppages is nearly always a successful improvement strategy.

CLOSE COUPLED OPERATIONS

All operations can be characterised as close coupled, coupled or decoupled. When machines are tied closely together such that the WIP between them is so low that a stoppage at one machine very soon shuts down the entire line, the process is said to be close coupled. Examples include bottling lines and high speed assembly lines such as are commonly found in electronics or medical devices. When large levels of WIP exist between elements of the process and where the various elements of the process are largely free to operate independently of each other, the process is said to be decoupled. Examples include mining and metallurgical operations and some pharmaceutical operations where the WIP and finished goods inventories are measured in months of stock. Between these extremes we find coupled operations. The more closely coupled are the elements of the process, the more damaging is the impact of variation. Sometimes the better solution is to conquer variation, but sometimes decoupling the operations is a superior approach. For instance, at a close coupled port facility that loads coal onto ships, the variation in train arrivals was beyond the port operator’s control. In addition, some elements of the operation were so dependent on each other that a stoppage/failure at one machine prevented another machine from passing it and soon stopped the flow of coal to a ship entirely. In this case the superior option was to decouple the elements of the process to reduce dependency. The cost of decoupling to increase throughput was far less than the cost of increasing capacity.

SUMMARY

Every day managers and schedulers are faced with the business decisions discussed here. Those managers and schedulers who understand the tradeoffs involved are in a far superior position to optimise their business in terms of customer service and profitability. As attempts to reduce variation in the flow of material through a process are successful, managers and schedulers will have new options. They can increase throughput whilst holding WIP steady, or they can hold throughput steady and decrease WIP (and perhaps finished goods inventory).

References:

1. *Factory Physics*, W. Hopp & M. Spearman, McGraw-Hill, 1996
2. *Six Sigma in the Pharmaceutical Industry*, J. McConnell and B. Nunnally, CRC Press, 2007