On Work In Progress Optimisation

Bruce Gunn and Saeid Nahavandi
School of Engineering and Technology, Deakin University, Australia

ABSTRACT
In this paper, we investigate how to best optimise the level of Work In Progress (WIP) in a real world factory. Using a simulation model of the factory, we have been able to show that an optimum level of WIP can be attained. By systematically varying the maximum allowable level of WIP within different model runs, results show that the throughput reaches a high level very quickly and then tapers off. The production lead times, in contrast, begin at relatively low levels and increase after the optimum WIP level has been reached.

KEYWORDS: Optimum WIP, Manufacturing, Discrete-event simulation

INTRODUCTION

Many manufacturing operations maintain high levels of Work In Progress (WIP) in order to maximize production output. While this is an admirable goal, many studies [1-3] have suggested this is achieved at significant cost, both in terms of WIP and production lead times through the manufacturing facility. While attempts have been made to investigate the optimal buffering problem [4-5], these studies are aimed at solving the complex mathematics that arise from these difficult problems. What they fail to address is the complexity associated with large numbers of workcenters and/or products found in many manufacturing enterprises. The concept of dynamic buffering has also been studied [6], which allows for improved delivery performance through the use of variable buffers to cover short-term fluctuations in demand. Following on from these studies, using discrete-event simulation and some fundamental concepts from production management, this paper aims to determine if an optimum WIP can be found in practice. At this optimum WIP level, a drop in maximum throughput is more than offset by a significant reduction in the cost of WIP and production lead times.

Hopp and Spearman [7] provide a base analysis of the relationship between WIP, throughput and lead times. This is an idealised analysis of a perfectly balanced transfer line, containing four machines with buffers in front of each. The study investigates the throughput and lead time along the line as the level of WIP is increased. Initially, with only one workpiece, the throughput of the line is low and the lead time the summation of the total processing work time through all four machines in the line. As the number of workpieces increases, the throughput increases but the lead time remains constant. When the number of workpieces reaches four, the throughput reaches a maximum, whilst the lead time maintains the same constant value. As more work pieces are added, the throughput plateaus, but the lead time begins to increase linearly with increasing WIP. This is best shown in figures 1 and 2, which plot the throughput against WIP level, and lead time against WIP level respectively.

The optimal WIP level therefore resides at the point where the throughput is at or near a maximum, and the lead time is at or near a minimum. Whilst the analysis is for a very simple, idealised case study, the results highlight the importance of the concept of an optimal WIP level.

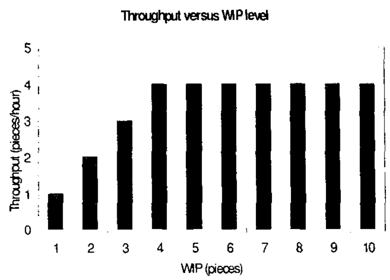

Figure 1. Throughput versus WIP

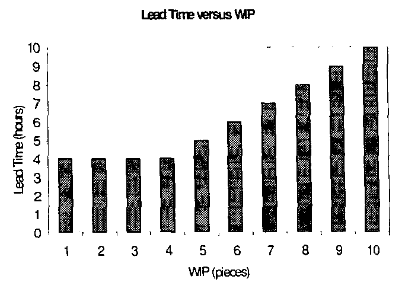

Figure 2. Lead time versus WIP

THE REAL-WORLD SYSTEM: Fastener Manufacturing Facility

The analysis provided in the introduction is good for developing understanding of the macro performance of manufacturing systems, but the question remains: Does such analysis apply to "real-world" factories? Two main methodologies could be employed to determine whether this is the case: trial a reduction in WIP within a production environment, or use computer simulation. Due to considerations of risk and cost, a simulation model of an existing factory has been developed. Using a series of techniques, the model has been developed, validated and verified against plant data, to attain good representation of both the production throughput and lead time.

The facility under investigation is a typical batch manufacturing plant, laid out along functional lines, with very little streamlining of operations, as described previously [8]. The factory layout and large product range means that the operational characteristics lie somewhere between job shop batch manufacturing and high volume repetitive manufacturing. Goods are typically transferred between operations via metal buckets, through the use of forklifts and overhead cranes. In order to study the optimal WIP problem for this factory, a schedule was extracted from the company databases, containing over 4000 work orders.

A series of experiments were designed to examine a scenario suggested in Hopp and Spearmann [7], where the level of WIP is maintained at a constant value, called the CONWIP (for Constant WIP) case. In this scenario, the level of WIP is capped at a maximum value and not allowed to increase beyond this point. This cap takes the form of a variable, which places a limit on the total number of buckets on the simulated shop floor, and can be altered at run time through a text file.

The basic experimental design is to systematically run the simulation model, gradually reducing the maximum allowable WIP and measuring the results, as summarised in Figure 3. Starting with an initially high level of WIP, the simulation model is run with the same schedule file used to validate the model. At the end of the simulation run, a custom written macro is executed to calculate the average production throughputs, lead times and levels of WIP. These values are calculated only during some pre-defined period of time, in order to filter out the end effects so often observed with simulation results [9]. This process is repeated over, but each time the level of maximum WIP, in terms of number of buckets, is reduced by a systematic amount. This systematic reduction was chosen to be 200 buckets, as this gave a good representation of the plant performance. The process ends when the maximum WIP level is below some pre-defined minimum level.

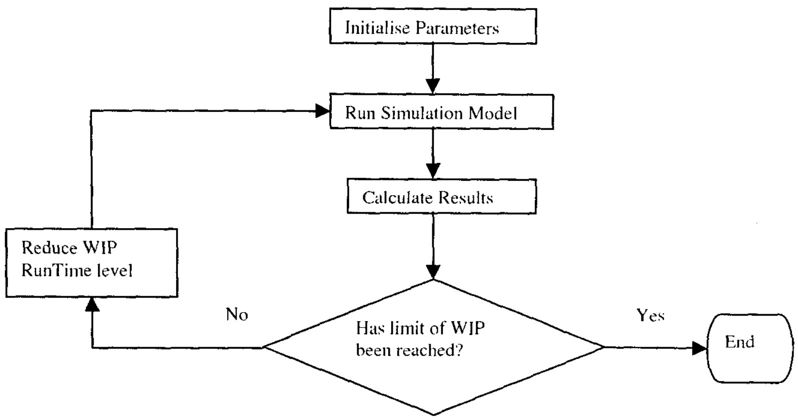

Figure 3. CONWIP experimental design.

RESULTS

Several different sets of results can be obtained from the one set of data. The prime area for confusion is how best to measure throughput and WIP levels. Throughput can be viewed as pieces (or number of product) per unit time, mass (tonnes) per unit time or dollars (cost) per unit time. Because different parts of the plant perform at different rates and are controlled via different mechanisms, this measure is not as simple as it appears. Similarly, with the level of WIP. For example, in the factory that was studied, some parts of the plant are controlled in a 'pieces per hour' manner, whilst other parts of the plant are controlled in a 'kilogram per hour' manner. This dichotomy is best solved by examining what matters most to business, the dollars per hour rates, but these numbers are, by definition, variable, and subject to changes from both internal and external factors. In essence, all three variables will be examined and some conclusions drawn.

Figure 4 provides an overview of the effect modifying the maximum level of WIP on both the pieces throughput (LHS y-axis) and the production lead time (RHS y-axis). Both are measured against the pieces WIP on the x-axis. Figure 4 is essentially an amalgamation of Figures 1 and 2, where production throughput and lead time have been plotted on the same graph. Also plotted is the level of 95% throughput, which is the cut-off for determining the point of optimum WIP. Throughput of 95% is considered the best choice, as most production managers would give up some percentage of throughput for greater flexibility in production.

Starting with the initial number of buckets set at 1800, the reduction in WIP can be observed at every point in the throughput and lead time curves. Each diamond (throughput) and square (lead time) is a reduction, of 200 buckets, in the maximum allowable level of WIP in the simulated plant.

Examination of Figure 4 leads to some conclusions about the factory performance. As the maximum number of buckets is reduced, the average level of WIP also reduces. Throughput, however, shows only a slight decline up to around 1000 buckets, from it highs just above 1 million pieces per day. Average lead time, however, indicates a monotonically decreasing function from highs of around 28 days, down to around 16 days. Thus, as expected, the throughput does not decrease dramatically, whilst the lead time decreases sharply at high levels of WIP. Below 15 million pieces in WIP, the throughput begins to decrease at a more rapid pace, whilst the lead time begins to level out. At this point the factory is past the point of optimum WIP. In terms of this study, the optimal WIP is defined as the point where the throughput drops below 95% of the maximum throughput. This is indicated in Figure 4 with the use of the vertical arrow. At this point, the throughput is just over 1 million pieces per day. This is just under 16 million pieces of WIP. This represents a potential saving of some 10 million pieces of WIP, and around 12 days in lead time.

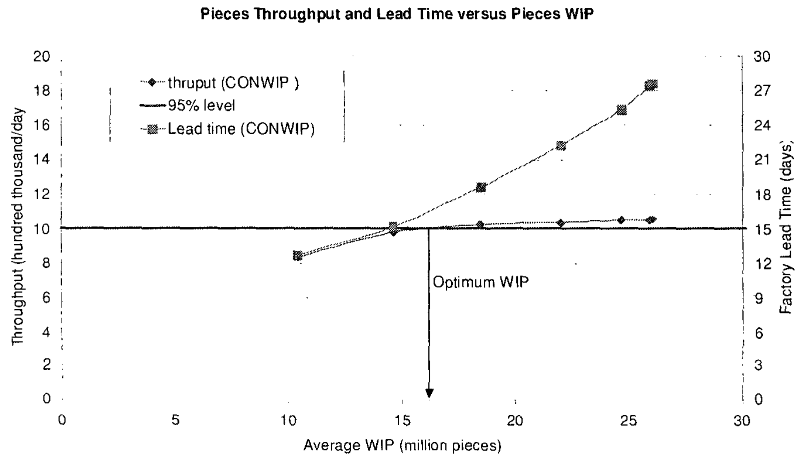

Figure 4. Pieces throughput and lead time versus pieces WIP.

The above behavior is reasonably simply described by examining the process flow and capacity limits within the plant. When the levels of WIP are low, very few of the downstream processing units are constraints to the flow of product through the plant. As quickly as the buckets are completed at the first operation, forging in this case, they are processed almost immediately by the next work center on the routing. Little queue time exists on the input buffers of the downstream work centers. Lead times through the plant are then the summation of the setup, processing and move times through each work center and are at, or near, a minimum. The throughput of the plant, however, is low as the primary machines being blocked from operating, due to the small number of buckets to manufacture product into.

As the level of WIP is increased, then queues begin to develop in front of the key processing work centers, and the percentage of queue time in the total lead time value increases. The primary machines are now less constrained, due to the greater availability of buckets in the plant. When all the key bottleneck resources have significant queues on their input buffers, then the increase in throughput tapers off with increasing WIP, while the lead time begins to increase. The key difference between this analysis from the model, and the simple scenario suggested earlier [7], is that the real world system has a number of different bottlenecks dependant upon the product mix being manufactured. A sharp change in rate of change of throughput is not observed in this real world case. Instead when all the key downstream bottlenecks have significant queues, the throughput can still slowly increase due to the product being manufactured but not passing through the key bottlenecks. If only one bottleneck existed in the plant, for all the products, then this bottleneck would control the throughput, and lead times, throughout the factory. Real world batch manufacturing, however, usually has several bottlenecks, and these bottlenecks are different for different product groups.

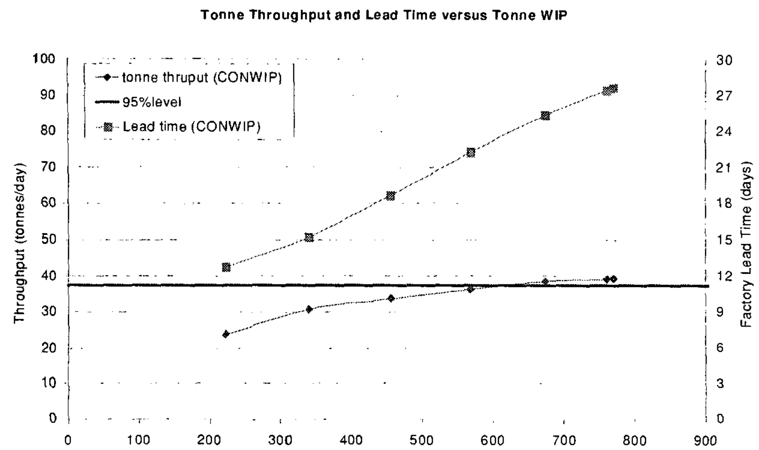

Figure 5. Tonne throughput and lead time versus Tonne WIP.

The throughput and WIP levels can be described in terms of quantity (i.e. pieces per day) or in terms of mass (i.e. kilograms per day). This is due to the batch nature of the manufacturing facility, where some functional areas are rate controlled via mass, and other functional areas are rate controlled via pieces. A graph can be plotted similar to Figure 4, where the throughput and WIP levels are measured in terms of tonnes rather than pieces. Such a graph is presented in Figure 5.

The dichotomy of determining whether throughput should be measured in terms of kilograms per day or pieces per day can be resolved by examining what counts most to a business, the value of the product.

Dollar costs are a key factor in the determination of most business decisions. By focussing on dollars, rather than pieces or tonnes, this study looks directly at one of the key business drivers. When the dollars are considered, a graph similar to Figures 4 and 5 can be plotted. This information is provided in Figure 6, and this shows a similar pattern to that observed in figure 5. The throughput increases up to around 1000 buckets, at which point the throughput is at 95% of the maximum throughput. Any subsequent increase in WIP leads to very little increase in throughput. The average lead time performs in a similar fashion to the previous graphs.

Figure 6 suggests that the optimum WIP level, based upon the 95% throughput criteria is around 2.6 million dollars. Such a decrease represents a saving of $700,000 in WIP from the maximum throughput case. This would again provide a lead time reduced from 27 to 21 days, with significant savings in terms of increased plant flexibility.

CONCLUSIONS

The above results show that an optimum level of WIP can be calculated for an existing factory given the scenario suggested by Hopp and Spearman. Using a discrete-event model developed for a fastener manufacturer, it has been shown that an optimum level of WIP exists. The results show that the factory throughput only drops slowly as the level of WIP is halved, from the high initial levels. As the level of WIP drops, the factory lead time also drops. At the optimum level of WIP, the reduction in throughput increases dramatically, while the rate of change in lead time decreases by a corresponding amount. The results of this modelling have been successfully implemented in practice.

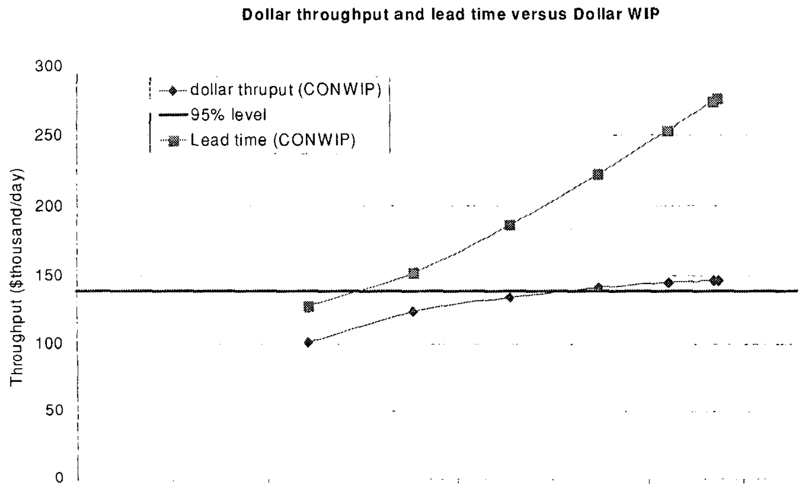

Figure 6. Dollar throughput and Lead time versus Dollar WIP.

REFERENCES

1. Goldratt E.M. and Cox J. *The goal. 1st edition*, Gower Publishing Company, England, UK. (1986)
2. Andijani A. "Trade-off between maximizing throughput rate and minimizing system tune in Kanban systems.", *Int J Ops. and Prod. Management, Vol. 17, No. 5, pp 429-445* (1997).
3. Hall R.W. *Zero Inventories*, Dow Jones-Irwin, Homewood, Illinois, USA.(1983).
4. Akella R. and Kumar P.R., "Optimal control of production rate in a failure prone manufacturing system.", *IEEE Transactions on Automatic Control, Vol. 31, pp 116-126.*(1986)
5. Bielecki T. and Kumar P.R., "Optimality of zero inventory policies for unreliable manufacturing systems.", *Operations Research, Vol. 36, pp 532-541.*(1988)
6. Gardiner S. C. and Blackstone J. H., "Dynamic buffering", *Int. J. Prod. Res., Vol 36, No. 2, pp 333-342*, (1998).
7. Hopp W.J. and Spearman M.L., *Factory physics, foundations of manufacturing management.*, First Edition, IRWIN Publishing, Chicago, USA.(1996).
8. Gunn B. and Nahavandi S., "Determining the optimum level of work in progress using constraint analysis and computer simulation", *Assembly Automation, Vol 20, No4, pp305-312, (2000).*
9. Kleijnen J.P.C., *Statistical tools for simulation practitioners.*, M. Dekker, New York, USA.(1987).