Roser, Christoph, Masaru Nakano, and Minoru Tanaka. "Holistic Manufacturing System Analysis: Optimization Based on a Single Simulation." In International Congress on Logistics and SCM Systems. Tokyo, Japan, 2004.

Holistic Manufacturing System Analysis: Optimization based on a Single Simulation

C. Roser, M. Nakano, M. Tanaka
Toyota Central Research and Development Laboratories, Japan

Abstract

Manufacturing systems are crucial for the Industry. A frequent objective is therefore to improve the throughput of the system, which can be done by allocating buffers in the system or by improving the machine performances. Yet, while an increase in the buffer size usually increases the throughput, it often also increases the inventory and the makespan. Furthermore, improving the wrong machines will only incur cost but not improve the system performance. This paper describes a general prediction model of these performance measures for different buffer allocations and machine performances based on only a single simulation using a holistic approach.

1. Introduction

The understanding and optimization of manufacturing systems is a frequently researched subject in discrete event simulation (Boesel 2001; Fu 2000) (Azadivar 1999; Swisher 2000). Yet, most methods analyze the simulation results in an atomistic view, and study each machine separately. However, if the correlated manufacturing system is broken into its independent machines before analysis, it is difficult to later combine the analysis of the machines into a understanding of the entire system. The presented methods here all use a holistic approach by studying the interactions between the machines in order to understand the system. Subsequently, the methods using this holistic approach yield a much better understanding of the manufacturing system.

Applying the holistic approach to the active times of the machines evaluates the shifting bottleneck in the system. The presented shifting bottleneck detection method is based on a detailed analysis of the relationships of the working times of the different machines and provides a very accurate and quantitative measure of the constraints in the machines.

Applying the holistic approach to the idle times of the machines enhances the understanding of the blocking and starving relations in the system and allows a better understanding of the effect of buffers. The presented buffer allocation method is based on a detailed analysis of the blocking and starving relationships in manufacturing systems and provides a usable prediction model of the effect of buffers in a manufacturing system, subsequently allowing the optimization of the system using only a single simulation.

2. A New Frontier: Holistic Manufacturing System Analysis

The newly developed methods described below all use a holistic approach to manufacturing system analysis compared to the atomistic view of the current simulation analysis methods. The holistic methods analyze not only independent machines, but the interaction between machines. The Merriam-Webster online dictionary defines holistic and atomistic as follows:

> **holistic***: relating to or concerned with wholes or with complete systems rather than with the analysis of, treatment of, or dissection into parts*
> **atomistic***: characterized by or resulting from division into unconnected or antagonistic fragments*

2.1. The Status Quo: Atomistic Analysis

What is called generally "a simulation" consists of many different steps, as for example building a model, verifying the model, simulating, collecting data, and finally analyzing the collected data. A manufacturing system is a network of interconnected entities, as for example machines, buffers, workers, or AGV. These entities affect each other and interact with each other. To understand the system it is crucial to analyze these interactions.

However, the status quo of current manufacturing simulation analysis is usually an atomistic analysis, completely ignoring the dependencies and interactions and achieving only a fraction of the possible conclusions which are hidden in the simulation data. This approach is visualized in Figure 1, where the manufacturing system is first simulated, and then broken into its individual elements for an atomistic analysis..

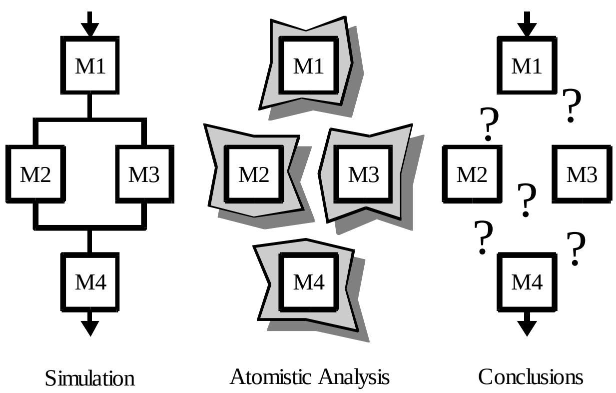

Figure 1: Atomistic Analysis Process

2.2. Holistic Analysis

The analysis presented here goes beyond the standard and is fundamentally different from the standard output analysis. The holistic analysis analyzes the interactions between the machines and statistically represents the correlations between the machines in order to understand the system. This results in a much better understanding of the system, and also allows for a prediction of the system performance based on the changes in the different entities. This approach is visualized in Figure 2, where after the simulation of the manufacturing system, the complex interactions of the system are analyzed to obtain a holistic understanding of the system. Subsequently, valuable information about the system is obtained and it is easy to make valid conclusions for the correlated system.

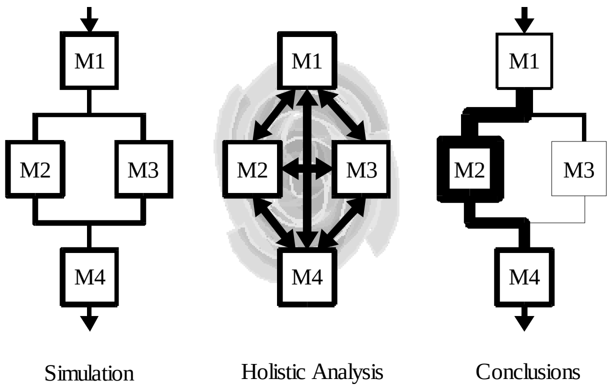

Figure 2: Holistic Analysis Process

3. Shifting Bottleneck detection method

The presented method will be able to detect and monitor the shifting momentary bottleneck of a production system, and also determine the average bottleneck over a selected period of time based on the duration the machines are active without interruption. Subsequently, a sensitivity analysis of the machines with respect to the throughput can be performed and a basic prediction model established. More details for the different sub-methods can be found in previous publications (Roser 2002a; Roser 2002b; Roser 2002c; Roser 2002d; Roser 2003b).

3.1. Holistic Analysis Of The Active Duration

The presented method is based on a holistic analysis of the duration a processing machine is active without interruption. A machine is active whenever the machine may cause other machines to wait. For example working on one part may cause a subsequent idle machine to wait for the completion of the part, or a machine under repair may block previous machines. A state is inactive if the associated machine is not active but instead waiting for the completion of another task, for example the arrival of a part or service, or for the removal of a part. The bottleneck detection method then compares the durations of the active periods of the different machines.

3.2. The Momentary Shifting Bottleneck

The underlying idea of the method is that at any given time the machine with the longest uninterrupted active period is the momentary bottleneck at this time. Figure 3 visualizes the method using a simple example consisting of only two machines. The figure shows the active periods of the machines over a short period of time. At the selected time t, both machines M1 and M2 are active. Yet, as M1 has the longer active period, M1 is the bottleneck machine for the time t. As there is no machine active before the current bottleneck period, there is no overlap and no shifting at the beginning of the current bottleneck period. However, at the end of the current bottleneck period, M2 is active and has the longest active period. Therefore, the subsequent bottleneck machine is M2. During the overlap between the current bottleneck period and the subsequent bottleneck period the bottleneck shifts from M1 to M2. Now, M2 is the bottleneck machine. Similarly, at the end of the bottleneck period of M2, the bottleneck shifts back to M1. Processing all available data using this method shows at what time which machine is the bottleneck machine, when the bottleneck is shifting, and when there is no bottleneck at all. Therefore, it is possible to detect and monitor the bottleneck at all times.

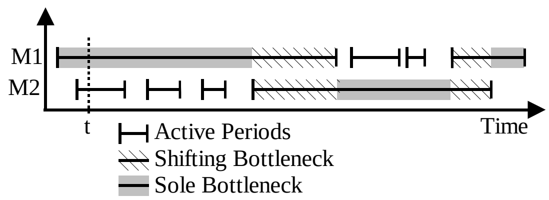

Figure 3: Shifting Bottlenecks

3.3. The Bottleneck Probability

The above method detects and monitors the momentary bottleneck at any instant of time. However, in many cases it may be of interest not to investigate an instant of time but rather a period of time. To determine the bottleneck probability, the percentage of time a machine is the sole bottleneck machine and the percentage of the time a machine is part of a shifting bottleneck is measured for the selected period of time as shown in Figure 4 for the example of Figure 3.

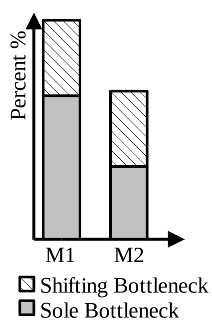

Figure 4: Average Bottleneck over Period of Time

4. Buffer Allocation Model

Buffers improve the system throughput by reducing the idle time (blocking and starving) of the machines. Therefore, to understand the buffers it is crucial to understand the blocking and starving of the machines, the causes thereof, and, most important the path to the causes and the buffer locations in between.

4.1. Holistic Starving And Blocking Analysis

To find the cause of an idleness of a machine, an algorithm has been developed that follows the cause from machine to machine or buffer until the cause of the idle period has been found. This method analyzes every starving or blocking occurrence of every machine in the simulation, and finds the cause of the starving and blocking, and, more important, the buffer locations on the path between the idle machine and the cause thereof. The time a possible buffer location is part of a path is determined for each machine. An example system with 7 machines has been analyzed, and the causes of the blocking and starving of the machines has been established. Figure 5 presents the results for machines M3 and M5 in graphical form, showing the path of the starves (cross-hatched) and blocks (diagonal-hatched) from machine M3 and machine M5 to the machine causing the starve or block. The width of the path represents the fraction of the starves/blocks following this path.

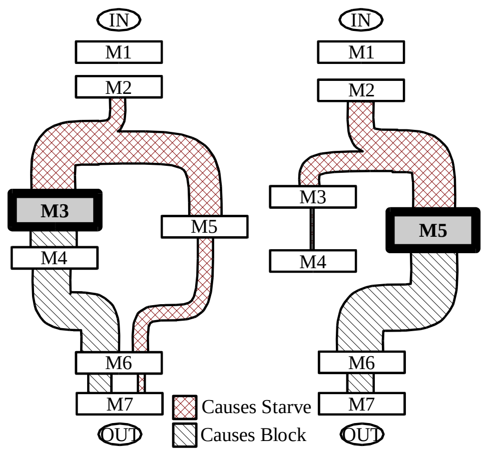

Figure 5: Causes of Blocking and Starving of Machines M3 and M5

In the simulated example, machine M3 is blocked 6.1% of the time. Whenever machine M3 is blocked, the path to the cause of the block leads to the next downstream machine M4 (100% of the blocked time). However, M4 itself is rarely the cause of the block. Most the paths continue to machine M6 (78% of the blocked time), and M7 (46% of the blocked time). Therefore, a buffer increase before machine M7 affects the blocking of machine M3 46% of the time. Machine M3 is also starved for 5.8% of the time. The path to the cause of the starve splits, with 38% of the starving periods caused by machine M2, and 62% following to machine M5. From machine M5 the paths continue to M6 (27% of the starving time), and from there to M7 (15% of the starving time). The buffer before machine M6 has an especially interesting effect, as it not only reduces the blocking of machine M3 by providing spaces to M4, but also reduces starving on the very same machine M3 by providing spaces to machine M5. The causes of the starving and blocking of machine M5 can be traced similarly. The path to the causes of the blocked and starved periods has to be analyzed for all machines to estimate the effect of buffers.

4.2. Buffer Prediction Model

The same buffer can have different effects depending on the number of parts and the number of spaces provided to the machines in the system. Therefore, the first step is to estimate the mean number of parts in a buffer, and subsequently the mean number of additional parts and the mean number of additional free spaces if a buffer is increased. The next step estimates the number of additional parts available in front of a machine to reduce starving and the additional number of spaces available after a machine to reduce blocking. Subsequently, the possible reduction in the time per part of the machines can be estimated.

The mean time that can be reduced therefore depends on the distribution of the starving and blocking times of the machines, and the probability density function of the starving time distribution and the probability density function of the blocking time distribution are needed to estimate the reduction in the idle times of the machines.

The mean reduced idle time can be calculated by integrating the probability density functions multiplied by the time span between the time 0 and the upper limit defined by the cycle time and the additional number of parts or spaces. The mean waiting time of the entire distribution can be calculated by setting the upper limit of the integral to infinite. The ratio of these two integrals is the percentage reduction of the waiting time. Combining this percentage reduction with the percent of the time a machine is starved or blocked gives the overall percentage reduction of the mean starving time per part and the mean blocking time per part. The total percentage reduction in the time between parts for a machine is the sum of the percentage reduction of the starving times and blocking times. The estimation of the system improvement is based on the individual machine improvements and the bottleneck probability.

The prediction model for the effects of buffers has been validated extensively for a number of different systems and buffers (Roser 2003a). For example, Figure 6 shows the comparison of the predicted time per part to the measured time per part for a buffer BM3 located before machine M3 in the same manufacturing system as shown in Figure 5. The continuous line shows the measured data including the 95% confidence intervals, and the dotted lines shows the predicted system performance. The predicted performance follows the measured data very nicely. The overall root mean squared error RMSE was only 0.24s.

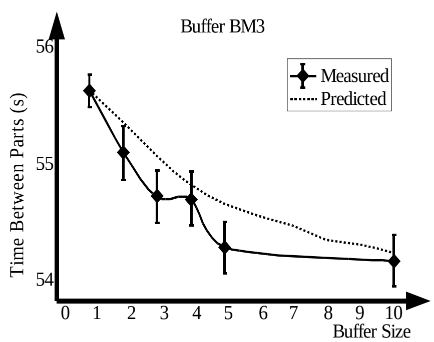

Figure 6: Performance Prediction for Buffer BM3

4.3. Buffer Optimization

The buffer prediction model can subsequently easily be used to optimize a manufacturing system. Two different optimization approaches are possible, one using a single step optimization, based on only a single simulation. The other approach uses a multi-step optimization, where the prediction model is used for a local optimization, after which a new simulation verifies the system and is itself optimized again.

To optimize the system an utility function is needed to create a trade off between the throughput, makespan, and the work in progress. As the prediction model allows the rapid comparison of many different buffer allocations, a wide variety of optimization methods can be used, as for example a gradient based method or a genetic algorithm. A good description of a wide range of optimization methods can be found in (Nemhauser 1994).

In a single step optimization, the buffer allocation of the manufacturing system is optimized to determine the buffer allocation with the maximum utility. Figure 7 shows a multi step optimization for a system similar to Figure 5. The simulation quickly reached an optimal plateau after 4 steps, and no further improvement was possible after step 13.

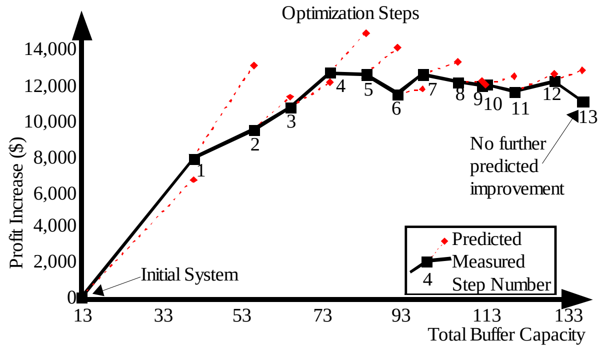

Figure 7: Multi Step Buffer Optimization

5. Conclusions

This paper summarizes a medley of methods recently developed by the Toyota Central Research and Development Laboratories related to the analysis of manufacturing system simulations. Unique to all the methods is the underlying holistic approach of analyzing not only the entities of the manufacturing system but also the interactions between these entities. All of the methods have been validated and implemented in an automated analysis software used by some companies in the Toyota group. This software creates an easy to use MS Excel file containing all the data and prediction models for optimizations.

The bottleneck detection method determines the momentary shifting bottleneck based on a holistic comparison of the machine working times. This provides a quantitative measure of the constraints, allowing the subsequent use of the bottleneck probability for a sensitivity analysis and a prediction model. Overall, this method greatly enhances the understanding of the system by quantitatively determining the primary and secondary bottlenecks and the non-bottlenecks.

The buffer prediction model is based on a detailed holistic analysis of the blocking and starving situations in the manufacturing system, determining the cause of every idle period in the system. This allows for a subsequent prediction model of the effect of buffers onto the system and a optimization of the buffer allocations. This prediction model greatly reduces the time needed to understand and improve the effect of buffers in the system.

Further research includes the development of additional methods in the promising field of holistic simulation analysis. By statistically analyzing the interactions in a manufacturing system or any discrete event system in general, a much deeper understanding of the system can be obtained. If the relations between the system entities is understood, the effect of changes in one entity can be estimated, greatly benefiting the industry by improving their competitiveness.

6. References

Azadivar, Farhad. (1999). "Simulation Optimization Methodologies." *Winter Simulation Conference*, Phoenix, AZ, USA, 93-100.

Boesel, Justin, Jr., Royce O. Bowden, Glover, Fred, Kelly, James P., and Westwig, Erik. (2001). "Future of Simulation Optimization." *Winter Simulation Conference*, Arlington, Virginia, USA, 1466-1469.

Fu, Michael, Andradottir, Sigrun, Carson, John S., Glover, Fred, Harrell, Charles R., Ho, Yu-Chi, Kelly, James P., and Robinson, Stephen M. (2000). "Integrating Optimization and Simulation: Research and Practice." *Winter Simulation Conference*, Orlando, Florida, USA, 610-615.

Nemhauser, G. L., Rinnooy Kan, A. H. G., and Todd, M. J. (1994). *Optimization*, Elsevier Science, Amsterdam.

Roser, Christoph, Nakano, Masaru, and Tanaka, Minoru. (2002a). "Detecting Shifting Bottlenecks." *International Symposium on Scheduling*, Hamamatsu, Japan, 59-62.

Roser, Christoph, Nakano, Masaru, and Tanaka, Minoru. (2002b). "Shifting Bottleneck Detection." *Winter Simulation Conference*, San Diego, CA, USA, 1079-1086.

Roser, Christoph, Nakano, Masaru, and Tanaka, Minoru. (2002c). "Throughput Sensitivity Analysis using a single simulation." *Winter Simulation Conference*, San Diego, CA, USA, 1087-1094.

Roser, Christoph, Nakano, Masaru, and Tanaka, Minoru. (2002d). "Tracking Shifting Bottlenecks." *Japan-USA Symposium on Flexible Automation*, Hiroshima, Japan, 745-750.

Roser, Christoph, Nakano, Masaru, and Tanaka, Minoru. (2003a). "Buffer Allocation Model based on a Single Simulation." *Winter Simulation Conference*, New Orleans, Louisiana, USA, 1238-1246.

Roser, Christoph, Nakano, Masaru, and Tanaka, Minoru. (2003b). "Constraint Management in Manufacturing Systems." *International Journal of the Japan Society of Mechanical Engineering, Series C, Special Issue on Advanced Scheduling*, 46(1), 73-80.

Swisher, James R., Hyden, Paul D., Jacobson, Sheldon H., and Schruben, Lee W. (2000). "A Survey of Simulation Optimization Techniques and Procedures." *Winter Simulation Conference*, Orlando, Florida, USA, 119-128.