Roser, Christoph, Masaru Nakano, and Minoru Tanaka. "Comparison of Bottleneck Detection Methods for AGV Systems." In Winter Simulation Conference, edited by S. Chick, Paul J Sanchez, David Ferrin, and Douglas J Morrice, 1192–98. New Orleans, Louisiana, USA, 2003.

COMPARISON OF BOTTLENECK DETECTION METHODS FOR AGV SYSTEMS

Christoph Roser
Masaru Nakano
Minoru Tanaka

TOYOTA Central Research and Development Laboratories
Nagakute, Aichi 480-1192 JAPAN

ABSTRACT

The performance of a manufacturing or logistic system is determined by its constraints. Therefore, in order to improve the performance, it is necessary to improve the constraints, also known as the bottlenecks. Finding the bottlenecks, however, is not easy. This paper compares the two most common bottleneck detection methods, based on the utilization and the waiting time, with the shifting bottleneck detection method developed by us, for AGV systems. We find that the two conventional methods have many shortcomings compared to the shifting bottleneck detection method. In the example presented here, conventional methods are either unable to detect the bottleneck at all or detect the bottleneck incorrectly. The shifting bottleneck detection method not only finds the bottlenecks but also determines the magnitude of the primary and secondary bottlenecks.

1 INTRODUCTION

Every logistic or manufacturing system has one or more bottleneck. To improve the throughput of the system, it is necessary to improve the bottlenecks. (Blackstone 2001; Goldratt 1992). The problem is to find these bottlenecks in the complex manufacturing systems used today. Numerous definitions of what constitutes a bottleneck can be found in the literature (Lawrence and Buss 1995). In this paper, we use a definition adapted from Kuo et al. (1996). We define a bottleneck as a machine whose throughput affects the overall system throughput, and the magnitude of the bottleneck as the magnitude of the effect of the machine throughput onto the system throughput. In summary, the sensitivity of the system throughput to the machine throughput determines the level of constraint of the machine.

The problem in finding the bottleneck is that manufacturing systems are not static, but instead vary over time. One cause of variation are random events, as for example a machine failure or other types of temporary delay. Thus, a machine may become a bottleneck only for a short period of time before the problem is resolved and the system returns to steady state. Subsequently, a manufacturing system usually has not only one bottleneck machine, but a number of machines that constrain the system at different times. The problem is to determine the average level of constraint of a machine onto the system over a longer period of time.

A second cause of variation are long term changes in the system. For example, the demand may vary seasonal, or a new product is introduced, changing the load of the machines and subsequently the constraints. Also, the manufacturing system itself may be changed, for example by adding, replacing or improving machines. Often especially the bottleneck machines are improved in an effort to chase the bottleneck.

In most systems there is one machine that is the largest, i.e. the primary bottleneck. However, often there are also a number of other machines that constrain the system, although to a lesser extend as the primary bottleneck. Within this paper we call these machines secondary bottlenecks. Also, manufacturing systems may contain machines that do not affect the system performance at all. These machines are no bottlenecks at all, and within this paper will be called non-bottlenecks. The improvement of the system will naturally focus on the primary bottleneck, but in some cases it may also be cost-effective to improve the secondary bottlenecks. It may even be possible to save some money by reducing the speed of the non-bottlenecks. Therefore it is important to find out which machines are primary bottlenecks, secondary bottlenecks, or non-bottlenecks.

There are currently a number of methods in use to detect the bottlenecks. This paper will compare the conventional methods using the utilization and the waiting time with the shifting bottleneck detection method (Roser, Nakano, and Tanaka 2002a; Roser, Nakano, and Tanaka 2002b) for an AGV system. A number of researchers have developed other bottleneck detection methods. Some methods are also based on utilization, using for example a matrix based approach to determine the overall constraint (Luthi 1998; Luthi and Haring 1997) or the ratio of the cycle time divided by the processing time (Delp et al. 2003). Other methods use a system theoretic approach to determine the sensitivity of the machine throughput to the system throughput (Chiang, Kuo, and Meerkov 1998; Chiang, Kuo, and Meerkov 2000; Chiang, Kuo, and Meerkov 2002; Kuo, Lim, and Meerkov 1996; Li and Meerkov 2000). Bukchin compared a number of theoretical estimations of the system performance, and found that an estimator based on the machine bottlenecks works best (Bukchin 1998). The following section will describe the AGV system used as an example before analyzing the system using the different methods and comparing the results. A discussion of the advantages and disadvantages of the methods is also added.

2 AGV SYSTEM

The presented system consists of three machines and three AGV's as shown in Figure 1. The three AGV's bring parts from the "In" station to the first machine M1, then to the second machine M2, to the third machine M3 and then to the "Out" station. The AGV's only proceed to the next stop if the next stop is free, i.e. not blocked by the previous AGV. Each machine also has two buffers of capacity one for unprocessed and processed parts. There is an infinite supply and demand of parts at the "In" and "Out" stations respectively.

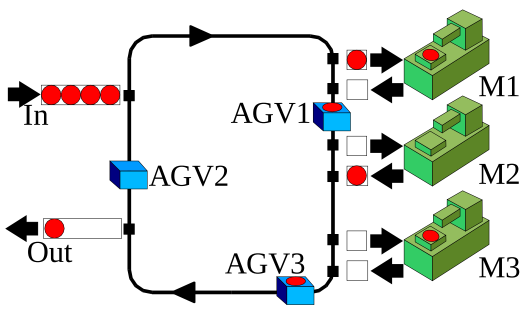

Figure 1: AGV System

Table 1 shows the machine parameters. Besides the deterministic cycle time, each machine has randomly occurring failures, with an exponential distributed mean time between failures (MTBF) of 10,000s and an exponential distributed mean time to repair (MTTR) of 500s. Table 2 shows the distances the AGV has to travel between the stations and the traveling time with a speed of 500mm/s. Loading and unloading is instantaneous. The distance from M3 to the "in" station includes the stop at the "out" station.

Table 1: Machine Parameters

Machine	Cycle Time (s)	MTBF (s)	MTTR (s)
M1	55	10,000	500
M2	60	10,000	500
M3	40	10,000	500

Table 2: AGV Travel

From	To	Distance (mm)	Time (s)
In	M1	34,650	69.3
M1	M2	6,050	12.1
M2	M3	6,050	12.1
M3	In	32,700	65.4

The simulation was implemented using the GAROPS simulation software (Nakano et al. 1994) and run for 400 hours simulation time. The measured production rate of the initial system was one part every 80.5s.

3 BOTTLENECK DETECTION METHODS

Three bottleneck detection methods based on the utilization, the waiting time, and the shifting bottleneck are applied to the system. The true bottleneck is also determined experimentally by improving the different machines or the AGV speeds.

3.1 Utilization

To detect the bottleneck, the utilization method measures the percentage of time a machine/AGV is active, and then defines the machine with the largest active percentage as the bottleneck. For the presented example the utilization of the different machines and AGV's was measured. As both working times and repair times can constrain the system, the utilization method uses the combined working and repair time percentages, resulting in the percentage of the time a machine or AGV is active as shown in Table 3.

Table 3: Utilization

Mean	Working	Repair	Active
M3	49.68%	5.24%	54.92%
M2	74.52%	5.44%	79.96%
M1	68.31%	4.91%	73.23%
AGV1	65.77%	-	65.77%
AGV2	65.78%	-	65.78%
AGV3	65.77%	-	65.77%

The measured active utilization of the system is also shown in Figure 2, including the confidence intervals with a confidence level of 95%. It can be seen clearly that, according to the measured data, machine M2 has the largest utilization of 80% (including both work and repair times), and therefore would be the primary bottleneck according to the utilization method. Machine M1 has the second largest utilization with 73%. The third largest utilization is found at the three identical AGV's with a utilization of 66%. Finally, machine M3 has the smallest utilization with 55%.

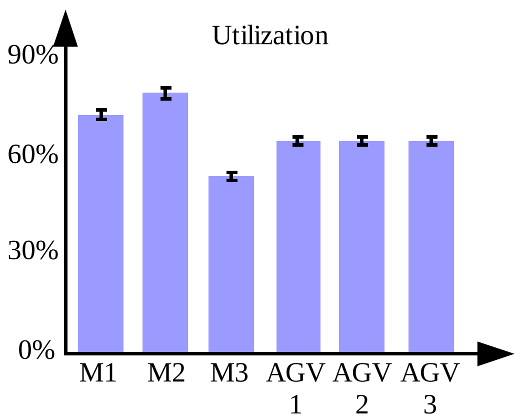

Figure 2: Utilization

Since machine M2 has the largest utilization, this machine would be the primary bottleneck according to the utilization method. Because the confidence intervals do not overlap, machine M2 has truly the largest utilization. Similarly, machine M1 has the second largest utilization, also with non-overlapping confidence intervals. Therefore, machine M1 has truly the second largest utilization. However, it is difficult to say if machine M1 is also a secondary bottleneck. Furthermore, it can only be guessed if the AGV's are also small bottlenecks or not. It is also unknown if machine M3 is a non-bottleneck. In summary, the bottleneck detection method using the utilization finds machine M2 to be the primary bottleneck, yet does not provide any insight regarding secondary bottlenecks or non-bottlenecks.

3.2 Waiting Time

The second bottleneck detection method uses the waiting time of parts in the queue to determine the bottleneck. This method looks for the machine where the parts have to wait for the longest time, and defines the bottleneck according to the queue length. An alternative method may look for the longest queue instead of the longest waiting time; however, this approach works only for linear systems containing only one type of part. If there are multiple part types there might be occasions where a machine with a few parts being processed slowly constrain the system more than a machine with a lot of parts being processed quickly. Therefore, it is recommended to use the waiting time instead of the queue length.

However, there are two additional limitations to the use of the waiting time for bottleneck detection. First, the buffers should have an infinite capacity. Bottleneck detection using the waiting time is difficult if the queue capacity is limited. Second, the system capacity should exceed the supply in the long run to avoid permanently filled queues. Both requirements are not satisfied in the presented example, as all buffers have a capacity of only one and there is an infinite supply. While the waiting time method can still be measured, the results are unusable. Also, the AGV system itself is a type of queue, yet it does not have a queue length or waiting time in the traditional way.

The measured waiting times and queue lengths are shown in Figure 3 for all buffers before and after the machines, where "B" indicates before, and "A" indicates after, i.e. buffer BM2 is before and AM2 is after machine M2. It turns out, that the buffers are empty almost all the time. For the buffers of capacity one, there is in average only 0.05 parts waiting, or a mean waiting time of only 4s. Furthermore, all buffers have approximately the same mean number of parts. Therefore, due to the limitations of the waiting time method the obtained results cannot be used to detect the bottleneck, and the method cannot determine the primary bottleneck, let alone the secondary or non-bottlenecks.

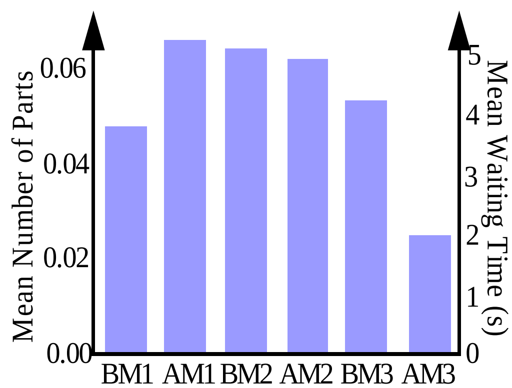

Figure 3: Queue Length and Waiting Time

3.3 Shifting Bottleneck Detection Method

The shifting bottleneck detection method uses the same data as the utilization method in determining the bottlenecks. The shifting bottleneck detection method also investigates when a machine is active or not. However, while the utilization method determines the percentage of time a machine is active, the shifting bottleneck detection method determines the duration a machine is active without interruption. This gives a much better understanding of the constraints within the system, and therefore allows for a much more reliable bottleneck detection. Initially, the average active duration (Roser, Nakano, and Tanaka 2001) has been measured, yet the method has been improved to determine the bottleneck at any given point in time by finding the machine or AGV with the longest active period at that time. This method has been proven to work reliably for non-AGV systems (Roser, Nakano, and Tanaka 2002a; Roser, Nakano, and Tanaka 2002b), and this paper will demonstrate the usefulness for AGV systems.

As the method is described in more detail in the above references, the following description will be brief. The shifting bottleneck detection method determines the periods during which a machine or AGV is active without interruption. The term "Active" includes not only machines working or AGV's transporting, but also breakdown periods, tool changes, or recharging times, i.e. any time a machine or AGV constrains the system. The active periods are occasionally interrupted by inactive periods, where the machine or AGV has to wait for the completion of a process by another machine or AGV, as for example when a machine is blocked or starved.

The underlying idea of the method is that at any given time, the machine with the longest active period is the bottleneck, and the system is constrained by this machine. The method further distinguishes between shifting bottlenecks, where the active period of one bottleneck overlaps with the active period of the next bottleneck, and sole bottlenecks, where the current bottleneck does not overlap with previous or subsequent bottlenecks. Figure 4 shows an example of a two-machine system; where at the beginning machine M1 has the longest active period, and therefore is the bottleneck. Later, the bottleneck shifts from machine M1 to M2, and then M2 is the sole bottleneck. The likelihood of a machine being the bottleneck can be measured easily by determining the percentage of the time a machine is a sole or shifting bottleneck.

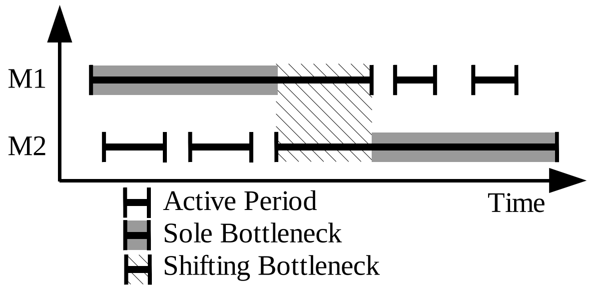

Figure 4: Shifting Bottlenecks

Analyzing the AGV system using the shifting bottleneck detection method as shown in Figure 5 gives a very different bottleneck than the utilization method shown in Figure 2. According to the shifting bottleneck detection method, the primary bottleneck is the AGV system, with each AGV having a total bottleneck probability between 25% and 50%, whereas the machines all have a bottleneck probability of less than 10%. Therefore the AGV's are the primary bottleneck, and the machines are the secondary bottlenecks. The system does not contain any non-bottlenecks. The bottleneck probability of the three AGV's is not equal due to the structure of the system. As the three AGV's always follow the same route in the same order without ever overtaking each other, the last AGV in this particular system is always more likely to be the bottleneck than the first AGV.

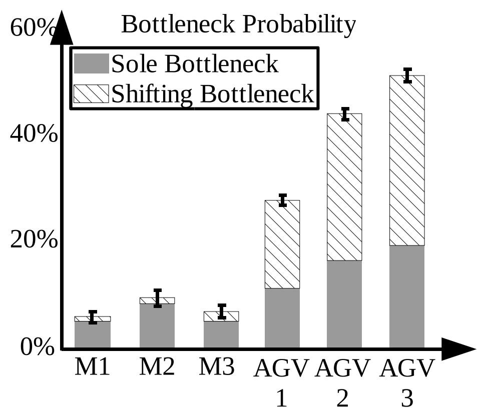

Figure 5: Shifting Bottleneck Probability

Figure 6 shows a graph of the sole and shifting bottlenecks for the original system, and the times of the failures of different machines. It can be seen clearly that every time a machine became a bottleneck, a machine failure has happened at the beginning of the bottleneck period. While Figure 6 shows only a brief period of simulation time, the results are similar throughout the simulation.

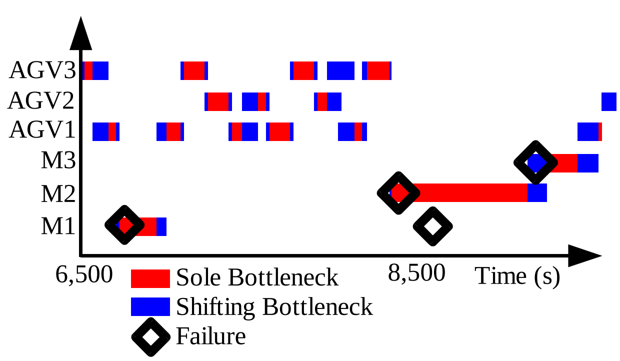

Figure 6: Bottlenecks and Machine Failures

The shifting bottleneck analysis using the shifting bottleneck detection method therefore not only detects the bottlenecks reliably, but also allows a deeper understanding of the underlying causes of the bottlenecks by investigating the temporary bottlenecks and their relation to different system events.

3.4 Experimental Bottleneck Detection

The previous bottleneck detection methods returned quite different results for the same AGV system. While the method based on the waiting time was unable to find the bottleneck, the utilization method and the shifting bottleneck method disagreed if machine M2 is the bottleneck or if the AGV's are the bottleneck. Only the shifting bottleneck method determined the secondary bottlenecks.

So, which machine or AGV is really the bottleneck? To find the bottleneck experimentally, different machines and AGV's have been improved independently, and the resulting production rate has been analyzed. In different simulations, the cycle times of the three machines have been improved, and the performance of the improved system has been measured. Furthermore, the performance of a system with an improved AGV speed has also been simulated. Table 4 shows the results of the different improvements, including the original and improved cycle times for machines and the original and improved speed of the AGV's. These results are also shown graphically in Figure 7, including the 95% confidence intervals.

Table 4: Improved Systems

Machine	Original	Improved	Time Per Part
M1 Cycle	55s	30s	79.6s
M2 Cycle	60s	40s	79.5s
M3 Cycle	40s	20s	80.3s
AGV's	500mm/s	1000mm/s	69.5s

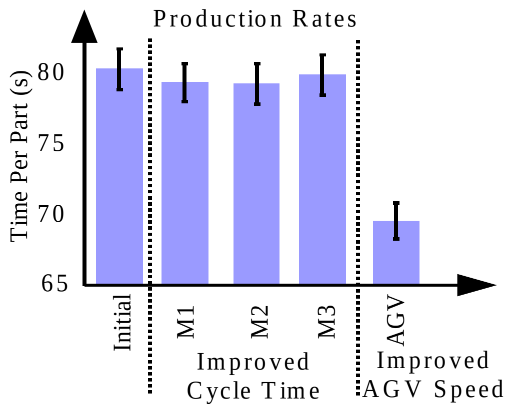

Figure 7: Production Rates

It seems that improving the machine cycle times does not affect the system throughput very much. In fact, the 95% confidence intervals all overlap with the confidence interval of the initial throughput, so it is not even sure if there is any improvement at all. However, improving the AGV speed significantly improved the throughput of the system.

To be more precise, the sensitivity of the throughput to the machines and AGV's has been determined as shown in Table 5. For example, the cycle time of machine M1 has been improved from 55s to 30s, or a 45% reduction. The resulting throughput has changed from 80.5s to 79.6s, or an improvement of 0.9s. Subsequently, the sensitivity of the system improvement to the machine improvement is 0.09s/45% = 0.020 s/%. The speed of the AGV has been doubled, and therefore the travel time has been reduced by 50%. The resulting system throughput has improved by 11s, and therefore the sensitivity of the throughput to the AGV speed is 11s/50%=0.220s/%. The results are also shown graphically in Figure 8.

Table 5: Machine Sensitivities

Machine	Machine Improvement	System Improvement	Sensitivity (s/%)
M1 Cycle	45%	0.90s	0.020
M2 Cycle	33%	1.00s	0.030

M3 Cycle	50%	0.20s	0.004
AGV's	50%	11.00s	0.220

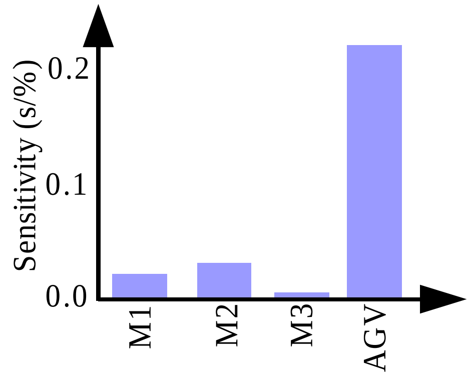

Figure 8: Measured Machine Sensitivities

Overall, the sensitivity of the throughput to the machine cycle time is very small. The most important factor is the AGV speed. Therefore, the primary bottlenecks of the system are the AGV's, and in order to improve the system it is necessary to improve the AGV's. The effect of a machine improvement is very small, and therefore the machines are only secondary bottlenecks. Therefore improving the machines does improve the system only very little.

4 COMPARISON

The different approaches gave vastly different results. The utilization method assumed that machine M2 is the primary bottleneck, and that the AGV's are actually rather insignificant, whereas in fact the AGV's are the primary bottleneck, and machine M2 is rather insignificant. Therefore, the utilization method did not detect the bottleneck correctly as can be seen by comparing Figure 2 with Figure 8. The method based on the waiting time was entirely unable to detect any bottlenecks due to the limitations of the method with respect to the manufacturing system. The method requires infinite or at least very large buffer sizes to determine the bottlenecks, and also a system capacity exceeding the supply, both of which were not given in the presented example. The only method returning correct results was the shifting bottleneck detection method. This method was not only able to detect the bottleneck, but also to estimate the relative constraint of the different machines to the system. Comparing Figure 5 with Figure 2 shows an almost identical behavior of the bottleneck measurement. An overview of the advantages and disadvantages of each method is shown in Table 6.

Table 6: Bottleneck Detection Methods Comparison

Method	Utilization	Waiting Time	Shifting Bottleneck
Accuracy	Medium	Medium	Excellent
Understandability	Medium	Medium	Excellent
Required Data Size	Large	Small	Small
Long Term BN	Yes	Yes	Yes
Medium term BN	No	Delayed	Yes
Short term BN	No	No	Yes
Primary BN	Yes	Yes	Yes
Secondary BN	No	No	Yes
Non-BN	No	No	Yes
Implementation	Very Easy	Easy	Medium
System Limitations	Moderate	Many	Few

All three methods are designed to determine the average bottlenecks over a longer period of time, as for example the entire simulation. However, sometimes not the long term bottleneck but rather the medium term or short term bottlenecks are of interest. A medium term bottleneck is the average bottleneck for a short period of time, for example over 1 hour after a new shift started, or the first 30 minutes after a machine breakdown, etc. The utilization method usually cannot detect these bottlenecks, and the waiting time method detects these bottlenecks only delayed. The shifting bottleneck method, however, is measured directly at the machine, and the bottleneck can be detected for any given period of time. The shifting bottleneck detection method can also detect short term or instantaneous bottlenecks as shown in Figure 4 or Figure 6.

All three bottleneck detection methods are designed to find the primary bottleneck, i.e. the single machine that represents the largest constraint onto the system. However, sometimes not only the primary but also the secondary bottleneck is of interest. The utilization and waiting time methods have difficulties finding the secondary bottleneck, whereas the shifting bottleneck method also finds secondary bottlenecks and determines the level of constraint of each machine onto the system. Furthermore, in some case it might be of interest to determine which machine is no bottleneck at all, for example to reduce the cost by reducing this machine speed. Only the shifting bottleneck method is able to determine the non-bottlenecks reliably. The shifting bottleneck detection method has few limitations and can be applied to almost all systems (Roser, Nakano, and Tanaka 2002a; Roser, Nakano, and Tanaka 2002b), whereas the waiting time method is limited in its use.

5 CONCLUSIONS

It has been shown that the conventional bottleneck detection methods as for example the measurement of the utilization or the waiting time occasionally fail to detect the primary bottleneck reliably, and are usually unable to detect secondary bottlenecks or non-bottlenecks.

On the other hand, the shifting bottleneck detection method based on the active periods is able to measure the likelihood of a machine being the bottleneck reliably for all machines and AGV's. Thus, it is easy to determine which machine is the primary bottleneck, which machines are the secondary bottleneck, and which machines are not bottlenecks at all.

This paper compared the utilization, waiting time and shifting bottleneck detection methods and found that no method excels in every area. However, the shifting bottleneck detection method has the one flaw of being slightly more difficult to implement than the other two methods, whereas the utilization and waiting time method have multiple limitations on usage and accuracy. Overall, the shifting bottleneck detection method is vastly superior to the other methods investigated in this paper. The method has been implemented in a software tool GAROPS Analyzer, automatically analyzing the data of the GAROPS simulation and detection the bottleneck. The results are shown in an easy-to-understand MS Excel spreadsheet.

REFERENCES

Blackstone, J. H. 2001. Theory of constraints - a status report. *International Journal of Production Research*, 39(6): 1053-1080.

Bukchin, J. 1998. A comparative study of performance measures for throughput of a mixed model assembly line in a JIT environment. *International Journal of Production Research*, 36(10): 2669-2685.

Chiang, S.-Y., Kuo, C.-T., and Meerkov, S. M. 1998. Bottlenecks in Markovian Production Lines: A Systems Approach. *IEEE Transactions on Robotics and Automation*, 14(2): 352-359.

Chiang, S.-Y., Kuo, C.-T., and Meerkov, S. M. 2000. DT-Bottlenecks in Serial Production Lines: Theory and Application. *IEEE Transactions on Robotics and Automation*, 16(5): 567-580.

Chiang, S.-Y., Kuo, C.-T., and Meerkov, S. M. 2002. c-Bottlenecks in Serial Production Lines: Identification and Application. *Mathematical Problems in Engineering*, to appear 2002.

Delp, D., Si, J., Hwang, Y., and Pei, B. 2003. A Dynamic System Regulation Measure for Increasing Effective Capacity: the X-Factor Theory. In *IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop (ASMC)*, ed., Munich, Germany.

Goldratt, E. M. 1992. *The Goal: A Process of Ongoing Improvement.* North River Press.

Kuo, C.-T., Lim, J.-T., and Meerkov, S. M. 1996. Bottlenecks in Serial Production Lines: A System-Theoretic Approach. *Mathematical Problems in Engineering*, 2: 233-276.

Lawrence, S. R., and Buss, A. H. 1995. Economic Analysis of Production Bottlenecks. *Mathematical Problems in Engineering*, 1(4): 341-369.

Li, J., and Meerkov, S. M. 2000. Bottlenecks with Respect to Due-Time Performance in Pull Serial Production Lines. *Mathematical Problems in Engineering*, 5: 479-498.

Luthi, J. 1998. Interval Matrices for the Bottleneck Analysis of Queueing Network Models with Histogram-Based Parameters. In *IEEE International Computer Performance & Dependability Symposium*, 142-151, Durham, NC, USA: IEEE Computer Society Press.

Luthi, J., and Haring, G. 1997. Bottleneck Analysis for Computer and Communication Systems with Workload Variabilities & Uncertainties. In *Proceedings of 2nd International Symposium on Mathematical Modelling*, ed. I. Troch and F. Breitenecker, 525-534, Vienna, Austria.

Nakano, M., Sugiura, N., Tanaka, M., and Kuno, T. 1994. ROPSII: Agent Oriented Manufacturing Simulator on the basis of Robot Simulator. In *Japan-USA Symposium on Flexible Automation*, 201-208, Kobe, Japan.

Roser, C., Nakano, M., and Tanaka, M. 2001. A Practical Bottleneck Detection Method. In *Winter Simulation Conference*, ed. B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, 949-953, Arlington, Virginia, USA: Institute of Electrical and Electronics Engineers.

Roser, C., Nakano, M., and Tanaka, M. 2002a. Shifting Bottleneck Detection. In *Winter Simulation Conference*, ed. E. Yucesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 1079-1086, San Diego, CA, USA.

Roser, C., Nakano, M., and Tanaka, M. 2002b. Tracking Shifting Bottlenecks. In *Japan-USA Symposium on Flexible Automation*, 745-750, Hiroshima, Japan.

AUTHOR BIOGRAPHIES

CHRISTOPH ROSER is a researcher at the Toyota Central Research and Development Laboratories, Nagoya, Japan since 2000. Current research topics are simulation output analysis, simulation optimization and logistic systems. He graduated from the Fachhochschule Ulm, Germany with a diploma in automation engineering, and received his Ph.D. in Mechanical Engineering at the University of Massachusetts, Amherst, studying flexible and robust design methods. He can be contacted by email at
<croser@robotics.tytlabs.co.jp>

MASARU NAKANO is a principal researcher at Toyota central Research and Development Laboratories, Inc. (TCRDL) in Japan. He joined TCRDL in 1980 and since then has studied in the field of robotics, computer vision, and manufacturing system design. He is currently research leader and manager of the Digital Engineering Laboratory. He received his B.S. and M.S. in operations research from Kyoto University, and his Dr. Eng. in manufacturing systems from the Nagoya Institute of technology. He can be contacted by email at
<nakano@robotics.tytlabs.co.jp>

MINORU TANAKA is a researcher at the Toyota Central Research and Development Laboratories, Nagoya, Japan. He graduated from Matsue Technical College, joined TCRDL in 1985, and specialized in the field of robotics, manufacturing system modeling, and discrete event simulation. He is a member of the robotics society of Japan. He can be contacted by email at
<tanaka@robotics.tytlabs.co.jp>