

SIMULATION BASED SCHEDULING USING A TWO-PASS APPROACH

Chin Soon Chong

Singapore Institute of Manufacturing Technology
Nanyang Technological University
71 Nanyang Drive
638075 SINGAPORE

Appa Iyer Sivakumar

Mechanical and Production Engineering
Nanyang Technological University
Nanyang Avenue
639798 SINGAPORE

Robert Gay

Electrical and Electronics Engineering
Nanyang Technological University
Nanyang Avenue
639798 SINGAPORE

ABSTRACT

Bottleneck-based scheduling is a popular approach in production scheduling, and it has achieved promising results in industry. To incorporate this approach in discrete event simulation tools is difficult since the approach requires multiple passes, forward and backward, to reach a good solution for the scheduling problem. In this paper, we propose a two-pass scheduling approach using discrete-event simulation that takes bottlenecks into consideration. In the first pass, a simulation run is performed and bottlenecks are determined. If significant bottlenecks are identified, a second-pass simulation is performed to reduce the loading on bottlenecks through specific scheduling strategies.

1 INTRODUCTION

Two broad approaches can be identified for solving scheduling problems: optimization or approximation. Optimization methods, which are directed at finding exact solutions by enumerative algorithms, are often unable to achieve feasible solutions to large problems due to excessive computing requirement. Examples of optimization approaches include Branch-and-Bound (Perregaard and Clausen 1998), and linear programming (Pinedo 1995).

Due to the limitation of exact enumerative techniques, approximation methods become a more viable alternative. In these methods, an optimal solution is not guaranteed but the gain in speed enables larger problems to be solved. The earliest approximation algorithms are priority dispatch rules (Panwalkar and Iskander 1977). These techniques assign a priority value to all the operations that are avail-

able to be sequenced and then the operation with the highest priority value is chosen. During the last 30 years, the performance of a large number of the priority dispatch rules has been studied extensively using simulation techniques (Jones and Rabelo 1998). Some complex dispatching rules comprise multiple levels of heuristics or priorities, with look ahead capability (Sivakumar 1999; Chong, Sivakumar, and Gay 2002).

Another popular example of approximation methods is the bottleneck-based approaches (Jain and Chan 1997). These approaches improve the performance of a manufacturing system through the management of the bottlenecks. In these scheduling approaches, multiple passes are normally performed to develop a good solution to the scheduling problem. This may involve taking one or more lots at a time and developing its complete sequence, backward and/or forward and iterating through all the lots. The algorithm may involve developing an initial sequence based on a simple strategy and then iteratively modifying the sequence to improve the quality of the solution.

In optimized production technology (OPT) for instance, bottlenecks are identified and scheduled first to optimize the throughput of the system. The non-bottleneck operations are then scheduled so as to reduce inventory (Fox 1984). Shifting Bottleneck Procedure (SBP) is another bottleneck algorithm, which is characterized by the following tasks: Subproblem identification, bottlenecks selection, subproblem solution, and schedule re-optimization (Demirkol, Mehta, and Uzsoy 1997). The strategy involves relaxing the problem into m one-machine problems and solving each subproblem one at a time. Each one-machine solution is compared with all the others and

the machines are ranked on the basis of their solution. The machine having the largest lower bound is identified as the bottleneck machine. The bottleneck machine is sequenced first, with the remaining unsequenced machines ignored and the machines already scheduled held fixed. Every time the bottleneck machine is scheduled, each previously sequenced machine susceptible to improvement is locally re-optimized by solving the one machine problem again.

In other bottleneck procedures such as OPIS – Opportunistic Intelligent Scheduler (Smith 1995), new bottleneck is recognized during the construction of schedule. In this procedure, bottleneck analysis is repeated each time a resource or a job has been scheduled.

Bottleneck based approaches have very limited use in simulation-based scheduling due to the multiple passes in time, forward and backward of the approaches. In a recent attempt of simulation based scheduling, Jain and Chan (1997) developed a bottleneck approach for lot release planning. In their approach, bottlenecks in the system were first identified through deterministic forward simulation. The lots that needed the bottlenecks were then determined and grouped together to reduce the time required for setups. Backward simulation was subsequently used to allow backward flow of lots from their operation at the bottlenecks to their first operation to determine their release times.

In this paper, we described a two-pass bottleneck scheduling approach on our earlier work of simulation-based scheduling projects (Sivakumar 1999; Chong, Sivakumar, and Gay 2002). The project was carried out for a test site of a semiconductor manufacturer. The complexity of semiconductor manufacturing with multiple resource constraints, unique equipment configurations, etc. makes it a very challenging scheduling problem (Sivakumar, 1999; Chong, Sivakumar, and Gay 2002). Simulation based scheduling approaches are well suited for such scenarios (Jain and Chan 1997, Gupta and Sivakumar 2002) and have been reported in use in semiconductor industry (Watt 1998, Sivakumar 1999).

This paper is organized to discuss the concepts, the use of concepts in the proposed approach, and examine the performance of the proposed approach in turn. The requirements and scope of the proposed system for scheduling are defined in Section 2. The third section elaborates on the concept of the proposed approach. The fourth section presents the results achieved using the approach for a semiconductor test site. Section 5 discusses the benefits and limitations of the approach that were realized during the development. The last section draws conclusions from this study.

2 REQUIREMENTS AND SCOPE

Semiconductor testing is considered one of the most complex systems in terms of production routes equipment, and interdependency relationship (Sivakumar 1999, Sivakumar 2000). The system is sophisticated because of 1) large and

changing varieties of the products, 2) complex product-to-tester relationships, 3) Multiple level tester-hardware dependency, 4) sequence dependent setup times, 5) re-entrant process flow, and 6) conflicting multiple objectives (Gupta and Sivakumar 2002). These complexities make scheduling problem in semiconductor backend testing a very demanding scheduling problem.

In solving scheduling problems of semiconductor manufacturing, the use of discrete event simulation method offers the advantage of developing a feasible schedule in shorter computation times compared to some other techniques (Sivakumar 1999). In this method each time an equipment becomes available in the simulated time, it selects an operation and all jobs available at that instant of time can be considered. During the selection process, all activities such as preventive maintenance, sequence dependent setups can also be taken into consideration. The focus can thus be on the selection of the most suitable lot on the available machine at decision instance t in simulation clock.

Our work is based on previously developed simulation-based scheduling systems for semiconductor test facility in a semiconductor assembly and test company. The research was initiated with an objective to further improve the system performance by addressing issues in the previous approaches, and incorporating the benefits of bottleneck based approaches.

Previous projects use an auto model generation approach, and have integrated the management policies, rules and algorithms into the systems, translated as parameters. Scheduling logic and algorithms have been implemented to optimize throughput, cycle time and asset utilization. These objectives are implemented by weighted factors.

The projects focus on the detailed modeling of the testers together with the required secondary resources such as handlers, and several other types of hardware. Machine unavailable times such as preventive maintenance are considered in the model. Factors in machine setups such as handler change, handler conversion, hardware change, temperature change and test program change are modeled in the systems.

In these projects, the program performs scheduling optimization by choosing the most suitable lot to load onto the available equipment in future simulated time. In this mechanism, it is difficult to identify bottlenecks in the system during simulation unless the bottlenecks are pre-determined prior to the simulation run. Besides, if detailed logic for look ahead were to be incorporated to recognize bottlenecks during simulation, additional complexity could be overwhelming on already complex semiconductor test modeling.

In our proposed approach, the focus is on the use of simulation output generated relating to lots and testers in the first simulation run, and subsequently use of the information in the second simulation run, in order to improve the system performance. The simulation run uses dispatching rules developed in the previous project without any modifications.

The second run incorporates new heuristics to make use of the details related to the bottlenecks in the first run.

Our approach differs from common simulation based scheduling procedures that use a two-stage simulation (Yang and Chang 1998). In these approaches, offline experiments are performed manually in the first stage with different scheduling rules to determine a condition that can achieve the desired performance of the manufacturing system. This set of dispatching rules will then be used in operations to generate the schedules. In our case, both simulation runs are executed automatically online without human intervention.

3 PROPOSED APPROACH

The approach to develop the two-pass simulation based scheduling consists of the following steps (Figure 1):

1. **First pass deterministic simulation.** This step is to generate detailed schedule and reports including queuing time of lots at testers, utilization of testers and setups. The simulation is run using scheduling algorithms in the previous projects, and results are extracted from simulation event trace.
2. **Capacity Analysis.** The simulation outputs are analyzed to determine the bottlenecks. Machines (i.e. testers) are ordered from the highest to the least constrained. Only those machines that had demand-to-capacity ratio above a pre-defined threshold level are considered to be bottlenecks (M_H). Machines that have demand-to-capacity ratio below a certain threshold level are also identified (M_L).
3. **Bottleneck Analysis.** In this step, lots that need the bottlenecks were identified. Queuing times of these lots are analyzed. Lots with queuing time to

cycle time ratio exceeding a certain threshold limit on the bottlenecks are recognized. This information, together with step and resource details at the bottlenecks is stored as attributes of lots in the second simulation run. The purpose of introducing these attributes is to redirect lots having long queuing time on the bottlenecks to less constrained resources in the next simulation run.

If no resource exceeds the upper or lower thresholds of the demand to capacity ratio, second pass simulation is not required, as no significant improvement could be expected.

4. **Second pass deterministic simulation.** The lot attributes and details on machines that are identified earlier (M_H and M_L) are incorporated in this simulation run. An additional level of priority ordering is implemented into the existing dispatching rules on the machines. On the bottlenecks (M_H), more emphasis is given on processing lots that requires the same setups, whereas on machines marked with M_L , setup rules are relaxed but the lots are still grouped by same setup whenever possible. No change is needed on other machines in the model.
5. **Schedule Comparison.** This step is introduced as a sanity check just to ensure that the better schedule between the first and second pass simulation is chosen based on the performance measures (i.e. throughput, cycle time and asset utilization). The selected schedule will then be used in the shop floor.

4 EVALUATION AND RESULTS

To quantify the improvement in performance through use of this system, different sets of actual data were collected

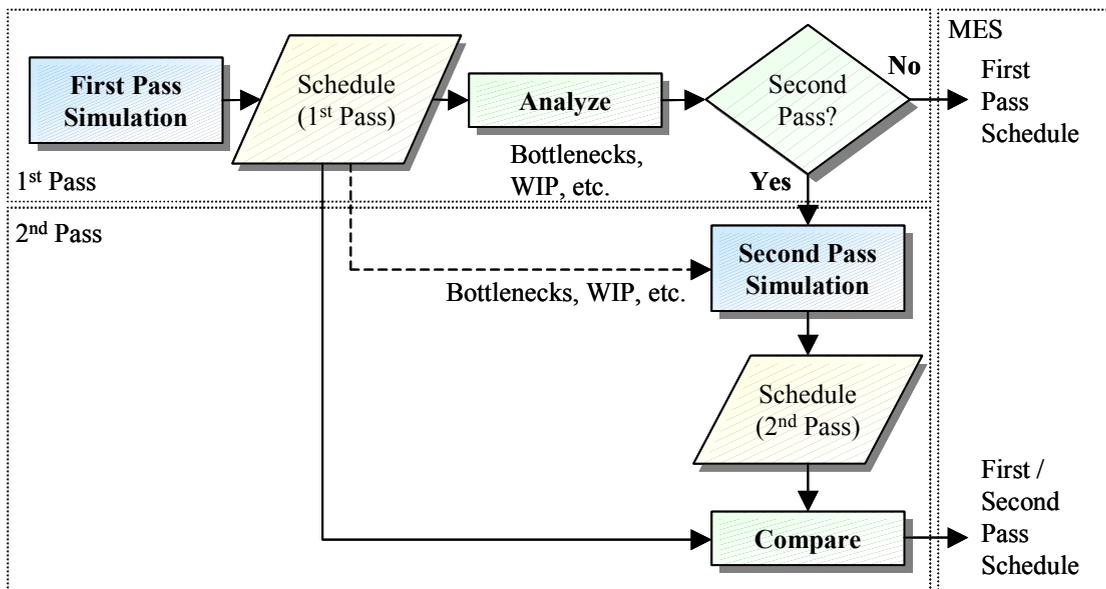


Figure 1: Two Pass Simulation-Based Scheduling Approach

from the manufacturing plant throughout a production period of about six months. The results from the original scheduling system were used as a base case to benchmark the new bottleneck approach. No simulation replication is performed since there is no stochastic element in the simulation model.

The first pass simulation is executed at a run length twice the actual duration of comparison. This is to ensure that the majority of lots have completed their routings and exited the system. Lot details in the first simulation run are needed for capacity and bottleneck analysis.

The results of the evaluation for different data set as compared to the base case are shown in Figure 2. The performance of the system was found to improve for production objectives of throughput (i.e. both total number of pieces and number of lots), cycle time and machine utilization for most of the data set. The level of improvement varies from 0 to 10% across the different data set. Significant reduction is found on the average setup times per machine as compared to other performance measures. This is to be expected because the new approach puts more emphasis on grouping lots based on same setups on the bottlenecks.

Analysis of the results reveals that the amount of setup time reduction across different data set correlates with the number of available lots for selection (see Figure 3). The more lots for selection when a machine becomes available, the better chance of finding lots that require the same setup to the previous processed lot on the machine. However, no obvious relationship can be established between throughput,

cycle time, and percentage of lots scheduled as the interdependencies among the factors could be very complex.

To further verify the proposed approach, a set of experiments was carried out (refer to Figure 4). The experiments involve changes in parameters. The results show that when all machines use the same setup strategy, there is a significant reduction in setup times, some improvement in machine utilization but cycle times and throughput suffer. In the case when all machines are considered as less constrained resources, higher throughput could be achieved but cycle times deteriorate. A better result, in terms of performance objectives, is obtained when bottlenecks use the same setup strategy and less constrained machines have a more relaxed rule on setups.

The last two experiments show that the threshold values (i.e. for bottlenecks and less constrained machines) can have a significant effect on the system performance. These two experiments are variations of the third experiment with less machines classified into the under utilized machine set (50% less) and bottleneck machine set (10% less) respectively. Setup times appear to be more sensitive to the threshold values compared to other performance objectives. However, the two experiments still indicate favorable results in terms of the target performance measures as compared to the base case. More extensive experimentation is needed to confirm the findings.

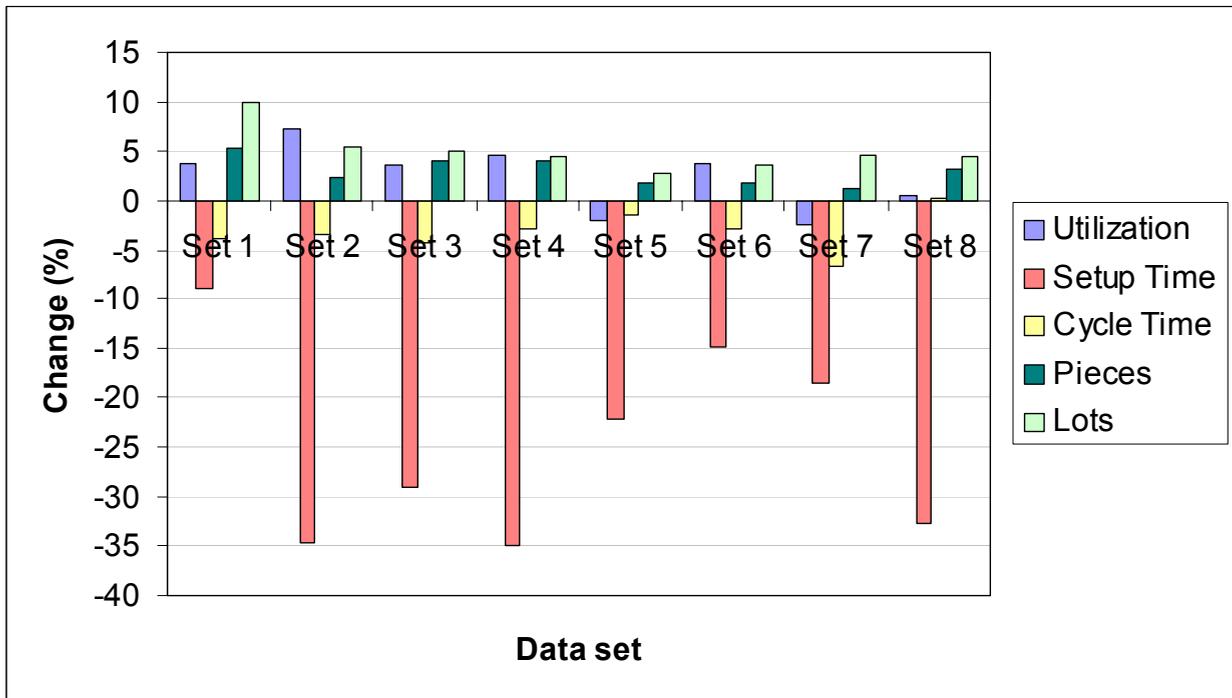


Figure 2: Two Pass Approach Compared to Base Case

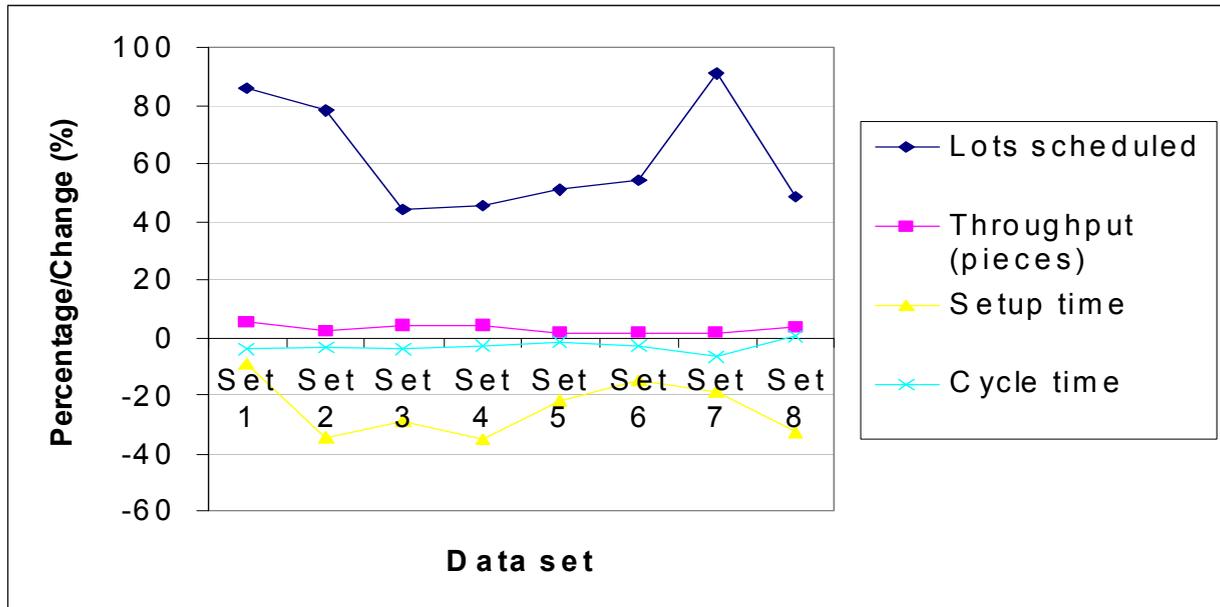


Figure 3: Relationship of Lots Scheduled and Performance Measures Compared to Base Case

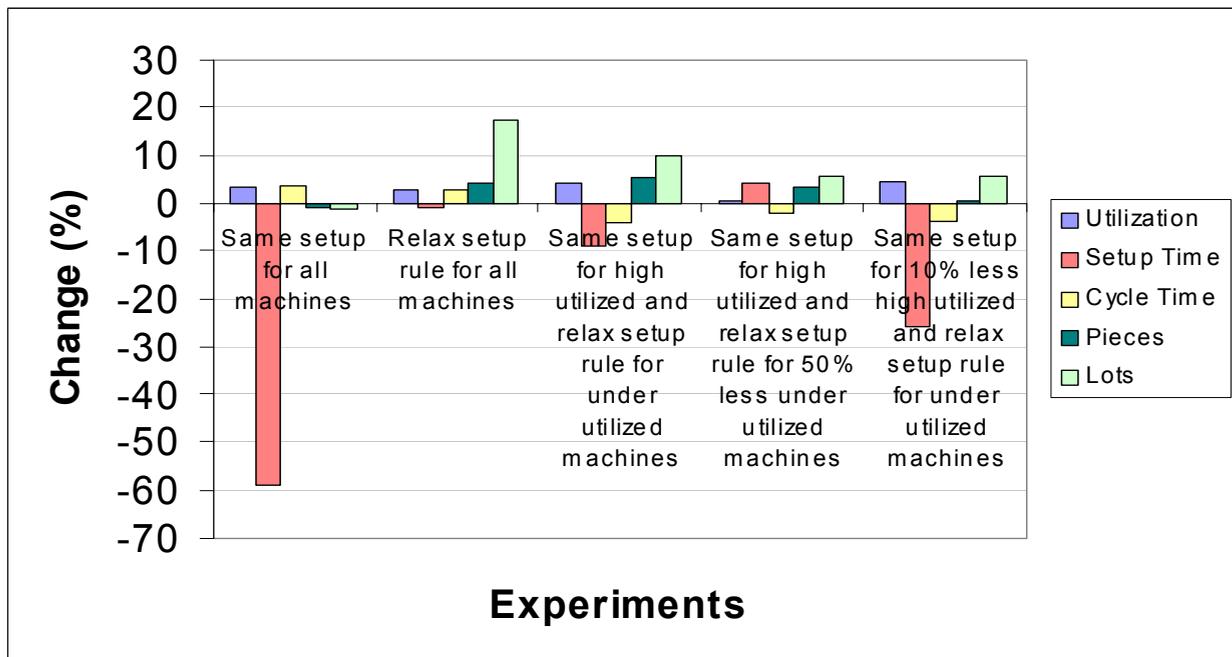


Figure 4: Experiments on Two Pass Approach

5 BENEFITS AND LIMITATIONS OF THE APPROACH

With the proposed two-pass approach, it is demonstrated that bottlenecks can be detected and subsequently taken care of in simulation-based scheduling. Earlier simulation studies had shown that better control of work at bottlenecks could lead to substantial improvements in system performance. In our approach, bottlenecks are handled

specifically in the simulation (i.e. second simulation run). To minimize non-value added times such as setup times at bottlenecks, an obvious dispatching strategy to use is to group lots that need the same setups. This strategy is applied in our work.

Further, less constrained machines are also identified. These machines are allowed to have less strict rules on setups. This permits more lots to be eligible to run on the machines even though it means more frequent setups. The

logic for this is that it is better for the non-bottleneck machines to have more frequent setups to achieve lower cycle times rather than idling.

By adopting this two-pass approach, it is generally possible to use different dispatching strategies for different level of constrained machines. However in the scenario for this project, with a large variety of products and routes, complex product-tester relationships, and many machines with capacity demands close to each other, a minor change in dispatching rules or setups could result in bottlenecks shifting.

Therefore it is believed that a better bottleneck driven approach should incorporate shifting bottleneck procedure. A recent paper (Roser, Nakano, and Tanaka 2002) demonstrated a method for detecting bottlenecks in manufacturing systems and the shifting of these bottlenecks. The method determines the bottleneck based on the duration a machine is active without interruption. This method, in principle, could be implemented in the simulation based scheduling and may thus allow a better control on the bottlenecks. However, the details of this method need to be researched further.

6 CONCLUSIONS

This paper presented a two-pass approach for complex semiconductor test operation scheduling. The approach has successfully considered bottleneck machines, and made use of the bottlenecks in simulation based scheduling. The initial experimentation and evaluation shows improvement in the system performance.

However, this approach can be very specific to the project undertaken, and the results may vary on other scheduling problems. The authors believe that more work can be done to generalize and further refine the approach.

ACKNOWLEDGMENTS

The authors would like to take this opportunity to thank various parties that were involved in the project. The dispatching rule in the first pass was designed by Associate Professor AI Sivakumar with the primary author. A number of people from the semiconductor assembly and test company contributed to this project, but their identities are not disclosed to maintain confidentiality. The authors are also grateful to the management of SIMTech to allow the time for experimenting with the proposed approach.

REFERENCES

- Chong, C. S., A. I. Sivakumar, and R. Gay. 2002. Design, Development and Application of an Object Oriented Simulation Toolkit for Real-time Semiconductor Manufacturing Scheduling. In *Proceedings of the 2002 Winter Simulation Conference*, E. Yücesan, C.-H. Chen, J.L. Snowdon, and J.M. Charnes (eds.), 1849-1856.
- Demirkol, E., S. Mehta, and R. Uzsoy. 1997. A Computational Study of Shifting Bottleneck Procedures for Shop Scheduling Problems. *Journal of Heuristics*, Vol. 3, No. 2: 111-137.
- Fox, R. E. 1984. OPT vs. MRP: Thoughtware vs. Software – Part II, Inventories and Production Magazine, Vol. 4, No. 1: 13-17.
- Gupta, A. K. and A. I. Sivakumar. 2002. Simulation based Multiobjective Schedule Optimization in Semiconductor Manufacturing. In *Proceedings of the 2002 Winter Simulation Conference*, E. Yücesan, C.-H. Chen, J.L. Snowdon, and J.M. Charnes (eds.), 1862-1870.
- Jain, S. and S. Chan. 1997. Experiences with Backward Simulation based Approach for Lot Release Planning. In *Proceedings of the 1997 Winter Simulation Conference*, S. Andradóttir, K.J. Healy, D.H. Withers, and B.L. Nelson (eds.), 773-780.
- Jones, A. and L. C. Rabelo. 1998. Survey of Job Shop Scheduling Techniques, NISTIR, National Institute of Standards and Technology, Gaithersburg, MD, 1998.
- Panwalkar, S. S. and W. Iskander. 1977. A Survey of Scheduling Rules. In *International Journal of Operations Research*, Vol. 25, No. 1: 45-61.
- Perregaard, M. and J. Clausen. 1998. Parallel Branch-and-Bound Methods for the Job Shop Scheduling Problem. *Annals of Operations Research*, Vol. 83, No. 1: 137-160.
- Pinedo, M. 1995. Scheduling: Theory, algorithms, and systems, Prentice Hall International Series in Industrial and Systems Engineering.
- Roser, C., M. Nakano and M. Tanaka. 2002. Shifting Bottleneck Detection. In *Proceedings of the 2002 Winter Simulation Conference*, E. Yücesan, C.-H. Chen, J.L. Snowdon, and J.M. Charnes (eds.), 1079-1086.
- Sivakumar, A. I. 1999. Optimization of Cycle Time and Utilization in Semiconductor Test Manufacturing Using Simulation based, On-line, Near-real-time Scheduling System. In *Proceedings of the 1999 Winter Simulation Conference*, P.A. Farrington, H.B. Nembhard, D.T. Sturrock, and G.W. Evans (eds.), 727-735.
- Sivakumar, A. I. 2000. Simulation based Cause and Effect Analysis of Cycle Time Distribution in Semiconductor Backend. In *Proceedings of the 2000 Winter Simulation Conference*, J.A. Joines, R.R. Barton, K. Kang, and P.A. Fishwick (eds.), 1464-1471.
- Smith, S. F. 1995. Reactive Scheduling Systems. *Intelligent Scheduling Systems*, D. Brown and W. Scherer (eds.), Kluwer Academic Publishing, Boston (Mass.), ISBN/ISSN: 0-7923-9515-8.
- Watt, D. G. 1998. Integrating Simulation based Scheduling with MES in a semiconductor FAB. In *Proceedings of the 1998 Winter Simulation Conference*, D.J. Medeiros, E.F. Watson, J.S. Carson, and M.S. Manivannan (eds.), 1713-1715.
- Yang, Y. and T. S. Chang. 1998. Multiobjective scheduling for IC sort and test with a simulation testbed, IEEE

Transactions Semiconductor Manufacturing, No. 11L 304-315.

AUTHOR BIOGRAPHIES

CHIN SOON CHONG obtained his degree in Electrical and Electronics Engineering from the City University of London, UK. He joined Singapore Institute of Manufacturing Technology (SIMTech), a research institute in Singapore, ten years ago and is currently in the Manufacturing Planning and Scheduling Group of Manufacturing Information Technology Division. He obtained his Master of Engineering in Computer Integrated Manufacturing from Nanyang Technological University, Singapore. During these ten years, he has been involved in simulation, scheduling and optimization related projects in logistic and manufacturing IT domains. The industrial projects include cargo container operation and yard capacity planning simulation for a container port, dairy-printing process simulation for a printing company, manufacturing cycle-time modeling, scheduling and optimization for various MNCs. His current research interest includes intelligent and integrated simulation, planning, scheduling, optimization in the area of manufacturing, logistic, and supply chain. He can be reached via email at [<cschong@simtech.a-star.edu.sg>](mailto:cschong@simtech.a-star.edu.sg).

APPA IYER SIVAKUMAR (Senior member IIE) is an Associate Professor in the School of Mechanical and Production Engineering at Nanyang Technological University (NTU), Singapore and a Fellow of Singapore Massachusetts Institute of Technology (MIT) Alliance (SMA). Prior to this he was at Singapore Institute of Manufacturing Technology, Singapore (SIMTech). His research interests are in the area of simulation-based optimization of manufacturing performance, supply chain, and dynamic schedule optimization. Prior to joining SIMTech in 1993, he held various management positions including technical manager and project manager for nine years at Lucas Systems and Engineering and Lucas Automotive, UK. He received a Bachelors of Engineering from University of Bradford, UK and a PhD in Manufacturing Systems Engineering from University of Bradford, UK. He has been the technical chair and co-edited the proceedings of the 3rd and 4th International Conference on Computer Integrated Manufacturing (ICCIM '95 and ICCIM'97), Singapore. His email and web addresses are [<msiva@ntu.edu.sg>](mailto:msiva@ntu.edu.sg) and [<www.ntu.edu.sg/home/MSiva/>](http://www.ntu.edu.sg/home/MSiva/).

ROBERT GAY obtained his B. Eng, M. Eng and PhD degrees at the University of Sheffield, England, in 1965, 1967 and 1970 respectively. He was awarded the Grouped Scholarship in Engineering and Metallurgy by the University of Sheffield from 1967 to 1970. Since obtaining his PhD, he has been involved in Education and R&D working in institu-

tions such as Singapore University (1972-1979), Rutherford and Appleton Laboratory (England, 1979-1982), NTU (1982-1995 and 1999-present) and Singapore Institute of Manufacturing Technology (1989-1999). He has also been actively involved in promoting innovation in Singapore through work in various committees: Science Quiz (MOE), Science Centre Board, National CAD/CAM (NCB), Tan Kah Kee Young Inventors Award (TKK Foundation & NSTB), National IT Plan (NCB), Technopreneurship incubation center (ITE), Commercenet Singapore, Singapore Computer Society. Currently Professor Gay is in the School of EEE at NTU and also Director and CEO of the ASP Centre. He has more than a hundred publications in journals, conference proceedings and books. His email and web addresses are [<eklgay@ntu.edu.sg>](mailto:eklgay@ntu.edu.sg) and [<www.ntu.edu.sg/eee/icis/cv/robertgay.html>](http://www.ntu.edu.sg/eee/icis/cv/robertgay.html).