Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

PARALLEL SIMULATION-BASED OPTIMIZATION ON SCHEDULING OF A SEMICONDUCTOR MANUFACTURING SYSTEM

Yumin Ma
Fei Qiao
Wei Yu
Jianfeng Lu

School of Electronics and Information Engineering
Tongji University
Shanghai, 201804, P.R.CHINA

ABSTRACT

As an important and challenging problem, the scheduling of semiconductor manufacturing is a hot topic in both engineering and academic fields. Its purpose is to satisfy production constraints on both production process and resources, as well as optimizing some performance indexes like cycle-time, movement, etc. However, due to its complexities, it is hard to describe the scheduling process with a mathematical model, or to use conventional methods to optimize its scheduling problem. A Simulation approach is proposed to optimize the scheduling of a semiconductor manufacturing system, i.e. a simulation-based optimization (SBO) approach. Because the high computational cost of SBO approach could hinder its application in the real production line, a parallel/distributed architecture is discussed to improve its efficiency. Using genetic algorithm (GA) as an optimization algorithm, the proposed parallel-SBO based scheduling approach for semiconductor manufacturing system is tested for its feasibility and effectiveness.

1 INTRODUCTION

Semiconductor manufacturing is a typical complex manufacturing system with reentrancy, uncertainty, high complexity and multi-objectivity. Its scheduling problems have always been researched broadly in both the engineering field and the academic field (Zhang et al, 2009), and have proved to be NP (non-deterministic polynomial)-hard. By now, there are many methods focusing on the optimization scheduling problem, including operational researches, heuristic methods, computational intelligence algorithms, and artificial intelligence algorithms. In addition, simulation techniques are also widely used in manufacturing industry, including the production management. As simulation platform, such as Siemens Plant Simulation, is becoming more mature, the descriptions of scheduling problem have been simplified for complex manufacturing system; the scale of the problem and the description of constraints are easily expressed; and the results are clear and understandable.

978-1-4799-7486-3/14/$31.00 ©2014 IEEE

Although the simulation-based scheduling of semiconductor manufacturing system well integrates heuristic methods into the scheduling model, and nicely explains the influence of the efficiency of the manufacturing system by different scheduling rules, it contributes nothing to optimization. This paper aims to combine the simulation and optimization together, namely simulation-based optimization (SBO) (Zhang et al, 2009; Rose, 2006), to get the optimal scheduling rules for a certain semiconductor manufacturing system. However, one of the main difficulties in the practice of SBO is its massive computational cost, which consumes a lot of time and leads to a poor system efficiency. In order to solve this problem, a parallel/distributed architecture is chosen to implement SBO system, which can deal with optimization scheduling problem and improve the efficiency of performance evaluation at the same time. The rest of the paper is organized as follows. The next section introduces the scheduling simulation system of a semiconductor manufacturing system. Section 3 introduces the concept and difficulties of SBO, while section 4 discusses the parallel SBO-based system, and describes a PSBO-GA (Parallel-SBO-GA) algorithm. It is followed by its application to the semiconductor manufacturing system and the experimental comparison in section 5. The summary of the paper is presented in section 6.

2 THE ARCHITECTURE OF SIMULATION SYSTEM OF SEMICONDUCTOR MANUFACTURING SYSTEM

In practical application, the architecture of a scheduling simulation system of semiconductor manufacturing system is shown in Figure 1, which contains three layers: Data, Model and User Interface.

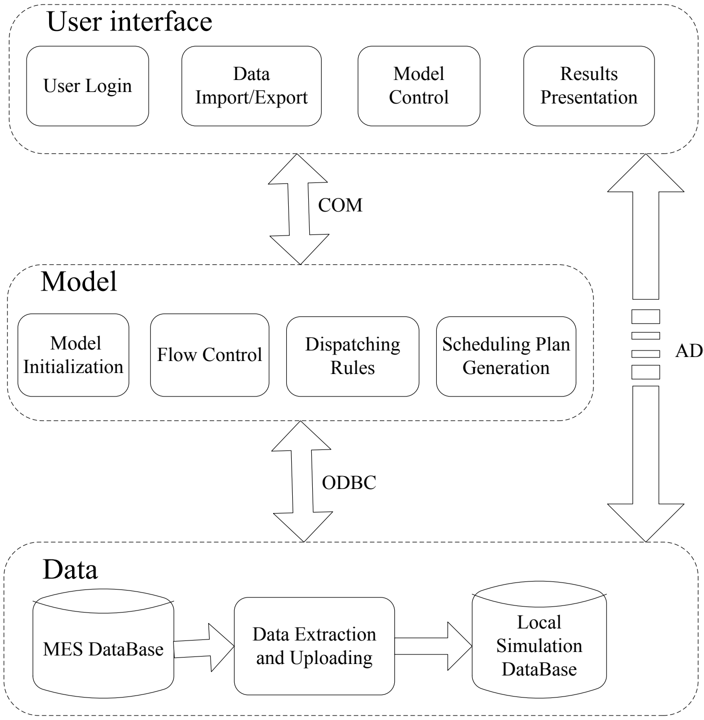

Figure 1: Scheduling Simulation System of Semiconductor Manufacturing System

Data layer uploads enterprise data from MES (Manufacturing Execution System), including both static and dynamic data, to local database of the simulation system in order to decrease the disturbance to the MES system as little as possible;

Model layer is realized by a commercial simulation platform, and is comprised of model initialization, flow control, dispatching rules and scheduling plan generation;

User interface is only needed when manual operation is in demand, including user login, data import/export, model control (run, reset, etc.) and data presentation.

So the complexity of a semiconductor manufacturing model is relevant to not only the number and category of both equipment and products of the manufacturing system, but also the processes and dispatching rules.

3 SIMULATION-BASED OPTIMIZATION

As mentioned above, simulation-based optimization effectively integrates simulation and optimization, so it is essentially consisted of two main steps: the simulation model and optimization algorithm. The simulation model describes the model of a manufacturing system, and outputs the simulation results for evaluation; the optimization algorithm applies evolutionary algorithm, such as generic algorithm (GA), particle swarm optimization algorithm (PSO) and so on, to generate scheme for the simulation model, and then evaluate according to the output of the simulation model. In this way, the whole process is finally updated and iterated. In the following section of this paper, the simulation model of a semiconductor manufacturing system is realized under a discrete-event simulation platform, and genetic algorithm (GA) is used for optimization.

The cycle of SBO can be described as follows. The optimization algorithm changes the control variables x, which is used in a simulation model. The model with its input 'x' runs to get target value c(x) (Klemmt et al, 2008; Gupta et al, 2002). Here, c(x) is the fitness of the algorithm, which is calculated through manufacturing system's performance indexes which can be achieved by the simulation model. Then the value is transferred back to the optimization algorithm. The algorithm evaluates this value and then outputs the next new control variable settings. This cycle is repeated until a termination criterion is satisfied. At last the optimal result is output.

The basic operation cycle of SBO is shown in Figure 2.

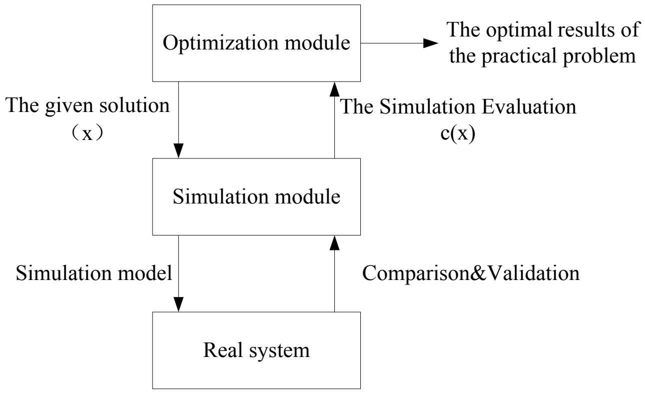

Figure 2: Simulation-based Optimization

As described above, the procedure of finding optimal result is an iterative procedure. In this procedure, simulation model runs again and again, which consumes both massive computational costs and lots of time. Thus a disadvantage of traditional simulation-based optimization is its low efficiency. Therefore, meeting the optimization goal and improving the efficiency are equally important for SBO in practice. In the next section, the idea of parallel/distributed architecture is introduced to solve the problem.

4 PARARREL/ DISTRIBUTED SBO-BASED SYSTEM

One way to solve the large simulation and optimization problems is to split them into smaller sub-problems (Law et al, 2002). Such decomposition may be done in different ways. In this section, a parallel/distributed architecture is applied to the SBO-based simulation and optimization system so that the efficiency of SBO can be improved. System is implemented under .net and a commercial simulation platform. Generally, a parallel/distributed system refers to computer networks at first where individual computers are physically distributed within some geographical areas. Distributed systems are based on a server-client mode and their shared purpose is to solve a large amount of computational problems. This kind of system, however, aims to coordinate the use of shared resources and provide communication services. For the scheduling of a semiconductor manufacturing system, the main purpose of the system is to optimize the scheduling strategies and get optimal performance indexes of the manufacturing system. Meanwhile, the server (main control unit) sends the computational tasks to client nodes and controls these simulation models to run at the same time. So it costs less time than common serial systems do. These features make parallel/distributed SBO-based system become an ideal solution to problems described before.

The architecture of the parallel/ distributed system is shown in Figure 3.

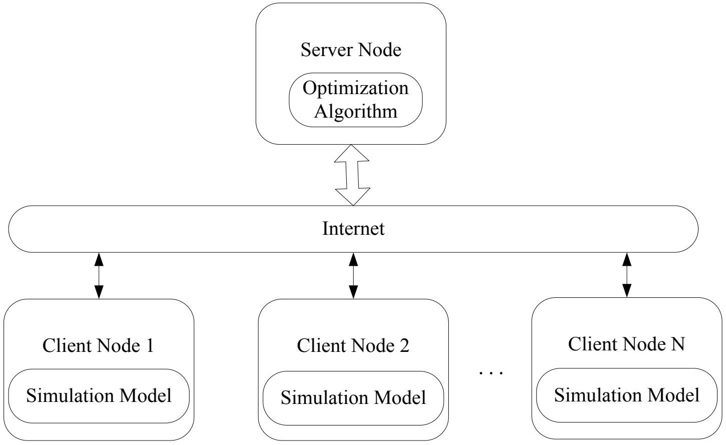

Figure 3: Parallel/Distributed System Architecture

The server node is consisted of a single computer, which controls the behaviors of clients, and runs optimization algorithm. Those client nodes can receive control information from the server, implement simulation models and send simulation result back to the server. Socket is used to realize the communication between server node and client nodes.

4.1 Client Node——SBO Simulation Model

The main function of client node is to receive the commands and control values from server node and to send performance indexes of simulation model back to the server node. Semiconductor manufacturing is a typical discrete-event system. A discrete-event simulation software platform called "Siemens Plant Simulation" is used to complete the modeling (Godding et al, 2007).

4.2 Server Node——Optimization Algorithm

The main function of server node is to send control values to client nodes and control the client nodes' behaviors via optimization algorithm. Genetic algorithm (GA) is a common optimization algorithm which has been proven to be very flexible and reliable in searching for global optimal solutions, and is also capable of solving complex scheduling problems (Chen et al, 2001). Moreover, it can be easily integrated in most discrete-event simulation models. For reasons above, genetic algorithm (GA) is used to implement the optimization algorithm of SBO in the paper, i.e. a genetic algorithm is embedded into the SBO system. Therefore, it is called PSBO-GA (Parallel-SBO-GA). The remaining of this section describes the implementation of PSBO-GA.

4.2.1 Chromosome

For GA, a chromosome/genome is a set of parameters which defines a proposed solution to the problem that needs to be solved. Here the chromosome code is based on 15 regular scheduling rules and work areas, which is shown in Table 1. The scheduling rule is encoded as shown in Table 2.

Table 1: Chromosome Code Design.

Work area 1	Work area 2	…	Work area N
Scheduling Rule 1	Scheduling Rule 2	…	Scheduling Rule N

Table 2: Scheduling Rule Code.

ID	Rule name	Description	Code
1	FIFO	First In First Out	0000/0001
2	EDD	Earliest Due Date	0010
3	EODD	Earliest Operation Due Date	0011
4	LPT	Largest Processing Time	0100
5	SPT	Shortest Processing Time	0101
6	CR	Critical Ratio	0110
7	FSVCT	Fluctuation Smoothing Policy for Variance of Cycle Time	0111
8	LS	Least Slack	1000
9	FIFO+	FIFO+	1001
10	SRPT	Smallest Remaining Processing Time	1010
11	SRPT+	SRPT+	1011
12	SRPT++	SRPT++	1100
13	FSVL	Fluctuation Smoothing Policy for Variance of Lateness	1101
14	LB	Line Balance	1110
15	Pheromone	Pheromone	1111

According to Table 1 and Table 2, if the current wafer fabrication has 3 working areas, then the length of chromosome equals to 12= (3*4). When the chromosome is encoded, the code of chromosome means that these scheduling rules are used for these work areas of a semiconductor manufacturing. For example, the chromosome code "0001 0010 0011" indicates that working area 1 uses the rule of FIFO , working area 2 uses the rule of EDD and working area 3 uses the rule of EODD.

4.2.2 Fitness Function

The fitness function is always problem dependent. Here a weighted sum algorithm is used to solve multi-objective optimization (Klemmt et al, 2007). The formula is shown below.

$$F(x) = w_1 f_1(x) + w_2 f_2(x) + \cdots + w_7 f_7(x). \tag{1}$$

$F(x)$ represents the fitness of an individual x and $f_i \left(i = 1, 2, \cdots 7\right)$ stands for seven performance indexes (1/cycle-time, 1/cycle-time's variance, movement, 1/WIP, overall equipment effectiveness, on time delivery and productivity) of a semiconductor manufacturing system, which are achieved by the simulation model of client nodes. Seven weighted values w_i are model-depended. The purpose of selecting weighted values is to unify the performance indexes, and therefore obtain the fitness that can well evaluate the overall performance of the semiconductor manufacturing system.

4.2.3 Genetic Operators

In the proposed algorithm, a roulette-wheel selection is applied to get better individuals. This selection procedure is repeated until enough individuals are selected. One-point crossover is employed to vary the programming of a chromosome or chromosomes from one generation to the next. And "Bit string mutation" is used as the mutation technique. The mutation of bit strings ensues through bit flips at random positions. For example, "101011" is converted to "101010" after the mutation.

4.2.4 Termination conditions of algorithm

When the algorithm meets one of the conditions below, the iteration stops:

- The algorithm reaches the maximum iteration.
- After several consecutive iterations, the global best fitness does not change.

4.2.5 PSBO-GA Algorithm Procedure

The PBSO-GA algorithm procedure is as following:

Step 1: Server node produces the initial generation;
Step 2: Server node sends individual's chromosome code to each client node;
Step 3: Client nodes receive and decode the codes into scheduling rules;
Step 4: Client nodes run simulation model according to scheduling assigned rules;
Step 5: Client nodes calculate performance indexes and their fitness after finishing model's running;
Step 6: Client nodes send the fitness values back to server node;
Step 7: Server node evaluates the individual's performance according to the fitness values: if the termination condition is satisfied, then the server node terminates algorithm and outputs optimal scheduling; otherwise, go to step 8;
Step 8: Server node executes selection, crossover and mutation operation and produces next generation;
Step 9: Go to step 2.

5 EXPERIMENTS

5.1 System Acceleration efficiency

Due to the purpose of designing the parallel/distributed system, which is to save the consuming time of SBO, the experiment is designed to test the time-saving efficiency. Table 3 exhibits the acceleration efficiency of the distributed system based on three different scales of semiconductor manufacturing systems (MiniFab is a small one, HP24Fab is a medium one, BL is a large one for a practical production line). In the experiment, maximum generation is 50 generations, and four client nodes are used. The size of the initial population is 40, crossover rate is 0.8 and mutation rate is 0.2.

The system gets different acceleration ratios according to different scales of simulation models. The duration of running a large model is much longer than that of synchronization and data communication in the network. As shown in Table 3, BL model and Hp24Fab model get higher acceleration ratios. On the contrary, MiniFab model doesn't get high acceleration ratio because it runs too fast. In a word, a large-scale manufacturing system can obtain better acceleration efficiency in the proposed parallel/distributed architecture than a small-scale one does.

Figure 4 shows the detailed time-cost comparison between models with one client node and that with four by varying different generations for BL Model.

Table 3: System Acceleration Efficiency.

Model	Acceleration Ratio
MiniFab	1.62
HP24Fab	2.96
BL	3.3

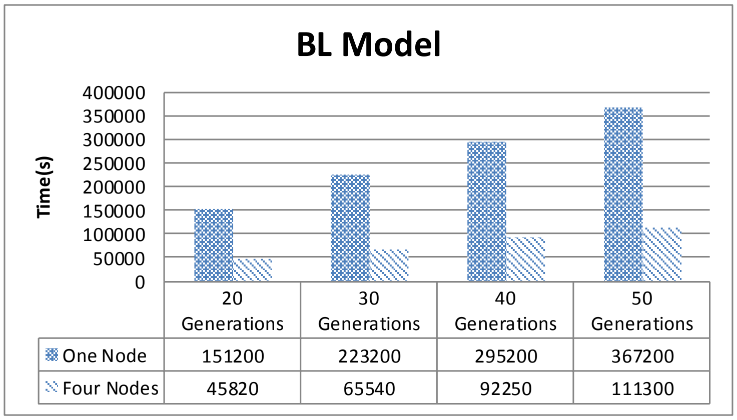

Figure 4: Time-saving efficiency

5.2 Optimal scheduling rules of the wafer fabrication

In this section, BL model is used for illustrating how the SBO system optimizes the selection of scheduling rules. The weight of fitness function is shown as follow to make the seven values between 0 and100 to evaluate the overall performance of the manufacturing system better.

$$F(x) = 10^6 f_1(x) + 10^5 f_2(x) + 10^{-1} f_3(x) + 10^3 f_4(x) + 10^2 f_5(x) + 10 f_6(x) + 10^2 f_7(x) \,. \quad (2)$$

In the experiment, the model runs for 60 days and uses a fixed release plan (4 lots per day). According to experiment result, the best chromosome is as follows:
1010 1111 1001 0001 0111 0110 1101 1001 1111 0111 0011 0111 0011 1010 1101 1100 0000 1011 0101 0011 0010 1110 0101 0101 1101 0110 1000 0110 0100 1101 0101 1111 1110 0001 1010 0010 0110 1000 1110 0110.

Table 4 is the optimal scheduling rule that corresponds to the best chromosome.

Table 4: SBO Optimal Scheduling Rule.

Work area	1	2	3	4	5
Rule	SRPT	Phero	FIFO+	FIFO	FSVCT
Work area	6	7	8	9	10
Rule	CR	FSVL	FIFO2	Phero	FSVCL
Work area	11	12	13	14	15
Rule	EODD	FSVCL	EODD	SRPT	FSVL
Work area	16	17	18	19	20
Rule	SRPT++	FIFO	SRPT+	SPT	EODD
Work area	21	22	23	24	25
Rule	EDD	LB	SPT	SPT	FSVL
Work area	26	27	28	29	30
Rule	CR	LS	CR	LPT	FSVL
Work area	31	32	33	34	35
Rule	SPT	Phero	LB	FIFO	SRPT
Work area	36	37	38	39	40
Rule	EDD	CR	LS	LB	CR

Table 5 is the comparison result of fabrication's seven performance indexes between SBO optimal rule and the three regular scheduling rules, which shows that BL model will get the optimized performance when using SBO optimal scheduling rule.

Table 5: Performance Indexes Comparison.

Performance index	SBO	FIFO	EODD	FSVCT
cycle-time (min)	13251.7	15942.2	16215.6	14984.3
Variance of cycle-time (min)	3425.5	4853.5	4007.2	3123.3
Movement	621.0	608.7	604.6	599.5
WIP (lot/day)	35.7	39.7	40.5	37.8
OEE(%)	15.5	14.5	15.3	15.2
on time delivery (%)	56.6	50	51.5	54.8
productivity (%)	3.27	3.47	3.37	3.28

6 CONCLUSION

An effective approach called simulation-based optimization (SBO) is employed to optimize scheduling problem of semiconductor manufacturing system. A parallel/distributed architecture for SBO is presented to improve the system's efficiency, such as time cost, etc., and some experiments have been done to prove the effectiveness and feasibility of the proposed method.

ACKNOWLEDGMENTS

This paper is supported in part by National Nature Foundation of China (No.61034004, 61273046).

REFERENCES

Zhang, H., Jiang, Z.B., Guo, C.T. 2009. "Simulation-based optimization of dispatching rules for semiconductor wafer fabrication system scheduling by the response surface methodology." *Advanced Manufacturing Technology* 41:110–121.

Rose,O. 2006. "Implementation of a Simulation-Based Optimizer for Semiconductor Wafer Factories." *In IEEE Conference on Emerging Technologies and Factory Automation, ETFA'06*.

Doleschal, D., Klemmt, A. and Weigert G. 2011. "Iterative Simulation-Based Optimization for Parallel Batch Scheduling Problems." In *34th Int. Spring Seminar on Electronics Technology:* 374-379.

Klemmt, A., Horn, S., Beier, E., Weigert, G. 2007. "Investigation of modified heuristic algorithms for simulation-based optimization." In *30th International Spring Seminar on Electronics Technology*: 24 -29.

Klemmt, A., Horn, S., Weigert G. 2008. "Simulation-Based And Solver-Based Optimization Approaches for Batch Processed in Semiconductor Manufacturing." In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler, 2041-2049. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Gupta, A. K., Sivakumar, A. I. 2002. "Simulation Based Multiobjective Schedule Optimization In Semiconductor Manufacturing." In *Proceedings of the 2002 Winter Simulation Conference*, edited by E. Yuecesam, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 1862-1870. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Law, A. M., McComas, M. G. 2002. "Simulation-Based Optimization." In *Proceedings of the 2002 Winter Simulation Conference*, edited by E. Yuecesam, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 41-44. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Godding, G., Sarjoughian, H., Kempf, K. 2007. "Application of combined discrete-event simulation and optimization models in semiconductor enterprise manufacturing systems." In *Proceedings of the 2007 Winter Simulation Conference*, edited by S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, 1729-1736. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Chen, J. H., Fu, L. C., Lin, M. H., Huang, A. C. 2001. "Petri-net and GA based approach to modeling, scheduling, and performance evaluation for wafer fabrication." *IEEE Transaction on Robotics and Automation* 17:619-638.

AUTHOR BIOGRAPHIES

Yumin Ma is an Associate Professor in the School of Electronics and Information Engineering at Tongji University in China. She received a M.S. and Ph.D. in Mechanical Engineering from Tongji University. Her email address is ymma@tongji.edu.cn.

Fei Qiao is a Professor in the School of Electronics and Information Engineering at Tongji University in China. She received a M.S. in Electrical Engineering and a Ph.D. in Economics and Management from Tongji University. Her email address is fqiao@tongji.edu.cn.

Wei Yu is a Master Degree Graduate in the School of Electronics and Information Engineering at Tongji University in China. He received a M.S. in System Engineering from Tongji University. His email address is yuwei06082130@foxmail.com.

Jianfeng Lu is an Associate Professor in the School of Electronics and Information Engineering at Tongji University in China. He received a M.S. in Mechanical Engineering and Ph.D.in Control Science from Tongji University. His email address is lujianfeng@tongji.edu.cn.