DPPS SCHEDULING POLICIES IN SEMICONDUCTOR WAFER FABS

Jim Dai and Steven Neuroth
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332
E-mail: dai@isye.gatech.edu

Key Words: SEMATECH fabs, Distributed Implementation, Cycle-time, Scheduling policies, Simulation

Abstract

The simulation results of a comparative study of scheduling policies using four SEMATECH dataset models are reported. A new family of policies termed Discrete Proportional Processor Sharing (DPPS) is shown to exhibit excellent performance characteristics over a variety of fab situations. The DPPS policies have practical advantages over other scheduling policies including use of only local state information, allowance of flexible decision-making, and no use of job arrival information. The DPPS policies have also been shown mathematically to be throughput optimal.

1 INTRODUCTION

Due to the complex re-entrant nature of semiconductor fabrication process flows, scheduling policies can have a dramatic impact on the performance of a fabrication facility (fab) (Morrison, Janakiram, and Kumar 1999; Schoemig and Mittler 1999). In this paper, a new family of policies termed Discrete Proportional Processor Sharing (DPPS) is introduced and shown to perform exceptionally well in a variety of fab situations. The policies were tested on four SEMATECH datasets in the AutoSched simulation software environment. The four SEMATECH dataset models chosen each have different characteristics with respect to the number of products, number of process flows, overall load on the system, and bottleneck make-up. The First-In-First-Out (FIFO) scheduling policy was chosen as the base for our comparison. The DPPS scheduling policies along with Global FIFO (G-FIFO) scheduling policies are compared against FIFO.

The DPPS family is comprised of three policies. Each policy is designed for either station families that have setup requirements, batching requirements, or neither attribute. (A station family consists of a group of machines having similar capabilities.)

One advantage that DPPS scheduling policies have is that they allow for distributed implementation since they use only local state information such as queue sizes, processing rates, and specific batching or setup information. A practical aspect of distributed implementation is that managers of a particular station family can focus on their area without having to worry too much about other areas. In addition, no lot arrival information is used, since in a real fab lot release rates change from week to week and are thus difficult to collect.

DPPS also allows for flexible decision-making within a station family. Once DPPS policies have been used to determine specific production targets for a certain time period, the operators have complete freedom to do the actual sequencing of the lots for that period. This allows the operators to use other information such as the upstream and downstream station families' information to optimize their operation.

The rest of this paper is organized as follows. Section 2 discusses the performance measures that we used in our studies to evaluate each policy. Also discussed in section 2 are our reasons for choosing FIFO and G-FIFO for comparison to our DPPS policies. Section 3 discusses modeling issues for the SEMATECH models that we chose. In Sections 4 through 7, each of the models is discussed and the results for each model are presented. Section 8 provides an overview of the results.

2 PERFORMANCE MEASURES

Performance of a fab is commonly measured using cost, throughput, WIP, and cycle-time.

Generally, scheduling policies do not have significant implementation costs associated with them when the overall cost of a fab is considered. A scheduling policy can have an effect on inventory that directly relates to cost. We know from Little's Law that WIP and cycle-time are related by the following formula:

$$\text{Mean WIP} = \text{Throughput} \times \text{Mean Cycle-time}$$

In the SEMATECH datasets, lot releases are modeled following a certain input rate, which is not dependent on the state of the fab. For each scheduling policy under study, the SEMATECH fabs are found to be stable with a given lot release pattern. Thus, the throughput is equal to the input rate into the system.

For such a system with consistent throughput, reduction in WIP is obtained by reducing the cycle-time. Thus, the important measure to study is cycle-time.

Two methods of measuring cycle-time are to look at the mean cycle-time of a fab and the standard deviation of the cycle-time. The mean cycle-time is used to determine how quickly a lot can be produced. The smaller the standard deviation of the cycle-time, the more reliably one can predict the finishing time of a lot.

The cycle-time for a lot is made up of two additive parts: raw processing time (RPT) and queue time. The raw processing time consists of the time when a lot is actually being processed. The remainder of time is the time that a lot spends in queue. The queue-time is the time that scheduling policies can have an impact on. Thus, one of our performance measures is the mean queue-time.

One of the objectives of this study is to show that DPPS policies improve mean cycle-time across *all* products. Our simulations also show that G-FIFO often improves the mean cycle time for some products at the expense of other products. The performance of each scheduling policy is shown as the percentage improvement of both mean queue-time and standard deviation of the cycle-time for each product in a fab. The improvements are relative to the results of the FIFO scheduling policy. For each policy, the ratio of mean cycle-time over raw processing time is also given for each product. To give a sense of the overall performance of the fab, the weighted average of all product ratios is given as well.

Many policies, including FIFO, are not throughput optimal (Bramson 1994; Seidman 1994). Even though the FIFO policy may not be throughput optimal in a wafer fab, we choose it as a nominal policy to compare with because it is used in many simulated or actual manufacturing environments, although not necessarily wafer fabs. We include G-FIFO in our study because it has been proven mathematically throughput optimal (Bramson 2001) and some wafer fabs have been using the policy. Global FIFO is also known as First-In-System-First-Out, or FISFO (Hopp and Spearman, 1996). As mentioned earlier, our DPPS policies have been proven to be throughput optimal. This study shows they also give good cycle-time performances, much better than G-FIFO.

3 SEMATECH MODELS

The simulation models were created using four datasets provided by SEMATECH. (SEMATECH datasets can be found at ftp://ftp.eas.asu.edu/pub/centers/masmlab/factory_datasets/.) The models contain all of the information within the original datasets with one exception. Operators were not modeled for the simulations. It has been shown the number of operators in the original datasets cannot handle the workload (Richardson 1996). In order to model operators, it would be necessary to add operators until there were enough to handle the work load requirements. It was determined that the performance should be measured based on the machine capabilities and scheduling policies alone.

Other modifications made to the original datasets are those that are documented within the comment files associated with each dataset including setup avoidance at specific station families. Minimum batch sizes were ignored when running DPPS simulations since the DPPS policies make all batching decisions that do not violate maximum batch size constraints.

Ten simulation runs were done with each scheduling policy for each fab. The lengths of the runs depended on how long the fabs took to reach steady state.

4 SEMATECH FAB ONE

Fab one represents a non-volatile memory plant. The fab includes two products each with a unique process flow. The station families considered bottlenecks were diverse in this fab. There was one severe bottleneck, with over 98% utilization, that was neither a batching or setup station family. There were five other station families with utilizations exceeding 90%: two set-up station families, one batching station family, and two station families with no batching or setup requirements.

For both the FIFO and G-FIFO policies, setup avoidance was implemented in both implanter station families. The DPPS policies have their own internal setup policies. Ten simulations were run lasting for 30,000 hours each. The first 10,000 hours of each run was truncated to allow the fab to reach steady state.

The raw processing times (RPT), in hours, along with the ratio of mean cycle-time to raw processing time for each product is shown in table 1. Table 1 also gives the weighted average of both the RPT and the performance ratios.

Product	RPT (hr)	FIFO	Global FIFO	DPPS
1	314.4	2.03	2.07	1.8
2	358.6	2.28	1.99	2
Overall	**328.5**	**2.11**	**2.04**	**1.87**

Table 1. RPT and the ratio of mean cycle-time to RPT for fab one

Figures 1 and 2 show the percentage improvements in mean queue-time and standard deviation of the cycle-time for DPPS and G-FIFO compared to the results obtained using the FIFO scheduling policy.

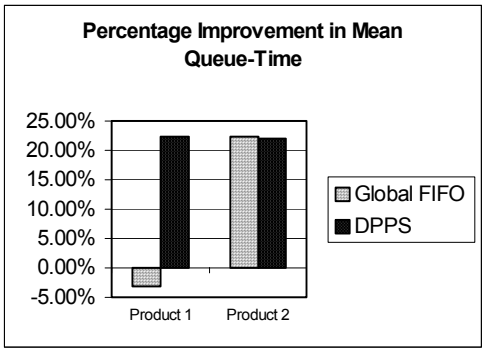

Figure 1. Percent improvement in mean queue-time over FIFO

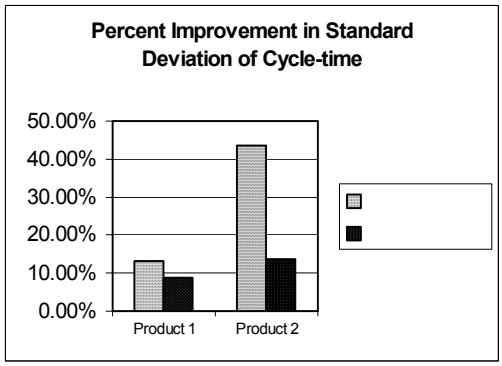

Figure 2. Percent improvement in standard deviation of cycle-time over FIFO

5 SEMATECH FAB THREE

Fab three is made up of 11 different types of memory products. The eleven products each have different process flows, however, some products have similar flows. The products can be broken down into four product groups with very different flows. There were two bottleneck stations in fab three. Two of the five furnace stations, both requiring batching, had over 99% utilization. However, most of the other stations were not heavily utilized.

Setup avoidance was necessary at the stepper machine for both the FIFO and G-FIFO simulation runs. The release rate for fab three was decreased in compliance with the associated comment file. Furthermore, the comment file specified that the lot inter-release times should be exponentially distributed. (For the other datasets, the inter-release times were deterministic.) This was done to measure performance of a normal rather than a saturated system. Ten simulations for each policy were run for 20,000 hours. The first 10,000 hours of each run was truncated to allow the simulation to reach steady state.

The raw processing time for each product, as well as the weighted average of all product raw processing times, are given in Table 2. Table 2 also gives the ratios of mean cycle-time to raw processing time for each of the 11 products in fab three along with the weighted average of each ratio.

Product	RPT (hr)	FIFO	Global FIFO	DPPS
1	296.6	1.27	1.36	1.21
2	296.8	1.26	1.34	1.20
3	289.2	1.24	1.33	1.19
4	354.9	1.25	1.30	1.19
5	359.1	1.24	1.29	1.19
6	175.5	1.25	1.37	1.20
7	354.8	1.25	1.29	1.20
8	359.2	1.23	1.28	1.19
9	173.4	1.25	1.39	1.19
10	203.5	1.22	1.33	1.18
11	188.4	1.23	1.33	1.20
Overall	**275.6**	**1.25**	**1.34**	**1.20**

Table 2. RPT and the ratio of mean cycle-time to RPT for fab three

The percentage improvements in mean queue-time and standard deviation of the cycle-time with respect to FIFO for fab three are given in figures 3 and 4, respectively.

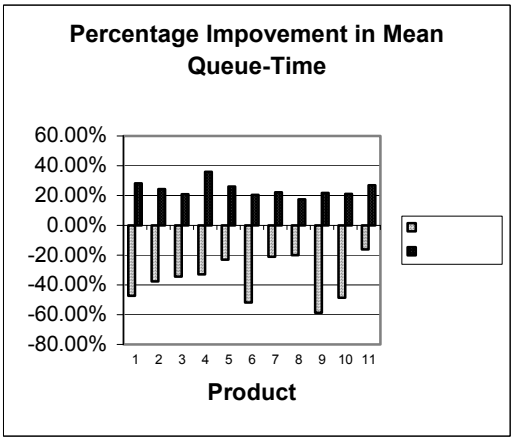

Figure 3. Percent improvement in mean queue-time over FIFO

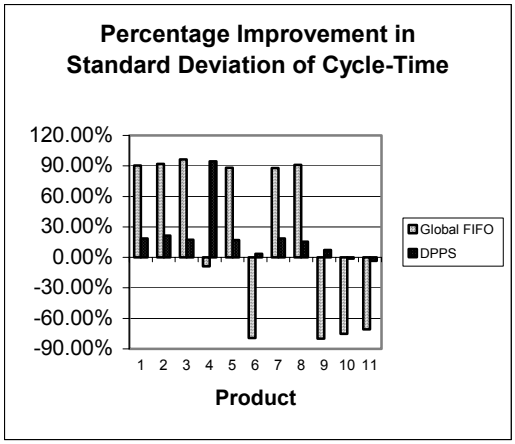

Figure 4. Percent improvement in standard deviation of cycle-time over FIFO

6 SEMATECH FAB FOUR

Fab four models a microprocessor fab and has a total of seven products. There are two unique process flows. Three products follow one full-loop process flow, and the other four products follow a short-loop process flow. Two of the five furnace stations, both requiring batching, were considered bottlenecks with over 98% utilization.

None of the machines in fab four require lot-specific setup; so only two of the DPPS policies were actually in use. The simulations were run for 30,000 hours with 10,000 hours truncated to allow the simulation to reach steady state.

Table 3 gives the raw processing time for each product and the weighted average of all product raw processing times. Table 3 also gives the ratios of mean cycle-time to raw processing time for each product in fab four as well as the weighted average of all performance ratios.

Product	RPT (hr)	FIFO	Global FIFO	DPPS
1	119.1	1.76	1.75	1.75
2	119.1	1.80	1.78	1.79
3	19.6	1.68	1.77	1.65
4	19.6	1.81	1.97	1.76
5	119.1	1.86	1.82	1.85
6	19.6	1.90	2.07	1.84
7	19.6	2.01	2.20	1.92
Overall	**109.3**	**1.81**	**1.81**	**1.79**

Table 3. RPT and the ratio of mean cycle-time to RPT for fab four

Figures 5 and 6 show the percentage improvement in mean queue-time and standard deviation of cycle-time for each product in fab four.

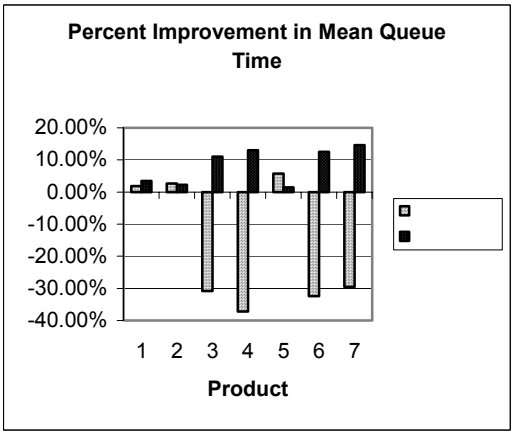

Figure 5. Percent improvement in mean queue-time over FIFO

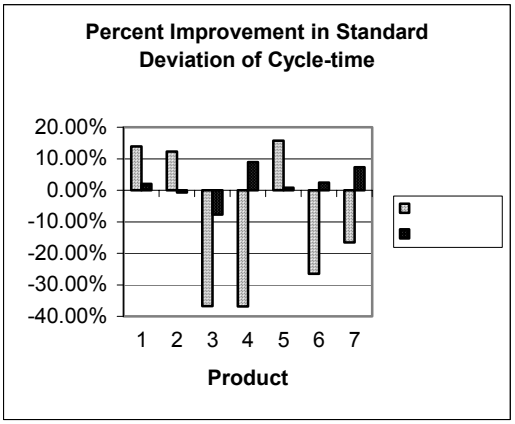

Figure 6. Percent improvement in standard deviation of cycle-time over FIFO

7 SEMATECH FAB SIX

The SEMATECH dataset modeling fab six was the most complex and heavily loaded of the four studied. The dataset models a fab that produces nine ASICS products. Each product has its own unique process flow. The bottleneck make-up for SEMATECH fab six was quite extensive. There were six station families that were utilized more than 98% of the time: four batching station families and two station families that required neither batching nor setup. Two of the station families requiring setup avoidance were utilized more than 94% of the time. The bottlenecks were not uncharacteristically high when compared to other station families, as was the case in some of the other fabs. There were ten more station families utilized more than 80% of the time.

Setup avoidance was required for four machine families. G-FIFO could not handle the workload with the minimum batch sizes provided in the dataset. For G-FIFO to keep the system from overloading, the minimum batch size for a bottleneck machine family called ASMB2 was increased from one to three in all of its batching groups except for the most common, whose minimum batch size remained at one. The simulations were run for a total of 50,000 hours, each run truncated by 20,000 hours.

Table 4 gives the raw processing time and the ratio of mean cycle-time to raw processing time for each of the products in fab six. The weighted average of both measures is also given.

Product	RPT (hr)	FIFO	Global FIFO	DPPS
1	421.6	2.82	2.40	2.24
2	398.8	2.28	2.54	2.17
3	260.4	2.54	3.36	2.77
4	361.6	2.85	2.59	2.42
5	283.9	2.89	3.07	2.18
6	323.6	2.46	2.83	2.16
7	356.5	2.28	2.80	1.99
8	316.2	2.32	2.96	1.98
9	307.9	2.19	2.94	2.32
Overall	**324.0**	**2.61**	**2.88**	**2.23**

Table 4. Ratio of mean cycle-time over raw processing time for fab six

Percent improvement in mean queue-time and the percent improvement in standard deviation of cycle-time are given in figures 7 and 8.

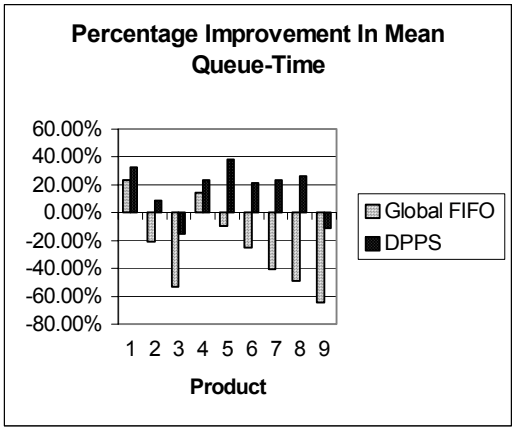

Figure 7. Percent improvement in mean queue-time over FIFO

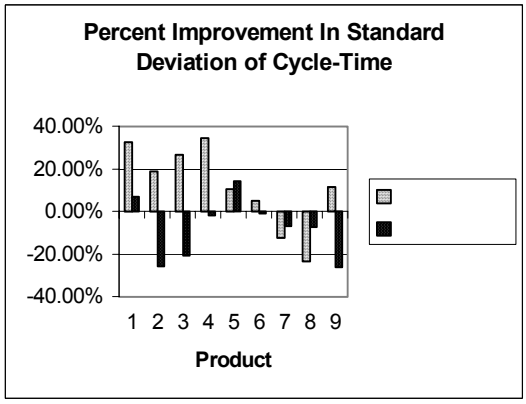

Figure 8. Percent improvement in standard deviation of cycle-time over FIFO in fab six

8 CONCLUDING REMARKS

The DPPS policies are shown to perform extremely well in a number of different situations. The policies perform particularly well, comparatively, in more complex and heavily loaded fab environments. The DPPS policies also exhibited a great deal of consistency over the products in a fab with respect to mean queue-times.

The DPPS policies were shown to perform better in all of the SEMATECH models, with respect to queue-time, when compared with both FIFO and G-FIFO policies. The improvement is most evident in the more complex fabs. In fab three, G-FIFO did not affect the mean queue-time compared to FIFO, while DPPS had over 15% improvement in mean queue-time for each product. In fab six, G-FIFO had worse performance in mean queue-time when compared to FIFO even with adjusted minimum batch sizes. DPPS again exhibited excellent performance in mean queue-time. Fabs one and four demonstrate both the consistency of the DPPS policies and the contrasting inconsistency of the G-FIFO policies in less complex environments.

Although DPPS does not necessarily perform as well as G-FIFO with respect to the standard deviation of cycle-time overall, it does perform more consistently over the products in a fab. DPPS typically reduced the standard deviation of cycle-time in all of the fabs except fab six where the weighted average gives only a 1.46% increase in the standard deviation of cycle-time. G-FIFO shows vast improvement in standard deviation of cycle-time with respect to some products; however, it often receives this improvement by increasing the standard deviation for other products in the same fab.

9 ACKNOWLEDGEMENT

This research was supported in part by National Science Foundation grants DMI-9457336 and DMI-9813345, and by The Logistic Institute-Asia Pacific, a partnership between National University of Singapore and Georgia Institute of Technology.

We would like to thank John Everton of Brooks Automation for helping us convert the datasets to the AutoSched AP format. We would like to thank Caiwei Li and Josh Reed for implementing an earlier version of the DPPS policies in the AutoSched AP environment. We would also like to thank John Fowler for taking time to answer our numerous questions regarding the SEMATECH datasets.

10 AUTHORS

Jim Dai joined the Georgia Institute of Technology in 1990. Since 1998 he has been a professor of Industrial & Systems Engineering and of Mathematics. He received the B.A. and the M.S. degrees from Nanjing University, in 1982 and 1985, respectively, and the Ph.D. degree from Stanford University in 1990.

Jim Dai's primary research interest is in the performance analysis and control of stochastic networks, with applications in semiconductor manufacturing. Over the past 10 years, his research group has worked with a number of semiconductor companies including Texas Instruments and Harris Semiconductor to devise efficient scheduling policies to improve fab performance. In recent years, he has also been interested in supply chain management research.

Jim Dai is a member of Institute for Operations Research and the Management Sciences (INFORMS). He is also a member of Institute of Mathematical Statistics. He has received a number of awards from professional societies including The Best Publication Award in 1997 and The Erlang Prize in 1998, both awarded by the Applied Probability Society of INFORMS. Jim Dai served as an associate editor for *Mathematics of Operations Research*. He is currently an associate editor for *Operations Research, Management of Science*, and *Queueing Systems*.

Steven Neuroth is an undergraduate student at Georgia Tech graduating in May 2002. His work on this project has included assisting in development of the policies, simulation, and programming. He has been working with Jim Dai since April of 2001. Steven plans on beginning his career in the semiconductor industry once he graduates. For more information, he can be contacted at steve.neuroth@lycos.com or (678) 467-8333.

REFERENCES

Bramson, M. 1994. Instability of FIFO Queueing Networks. *Annals of Applied Probability*, 4, 414-431.

Bramson, M. 2001. Stability of Earliest-Due-Date, First-Served Queueing Networks. *Queueing Systems*, 39(1). 79-102.

Hopp, W.J. and M.L. Spearman 1996. *Factory Physics*. Richard D. Irwin. Boston, MA. 437-8.

Morrison, J., M. Janakiram, and P.R. Kumar. 1999. A Comparative Study of Scheduling Policies at Motorola Fabs. In *Proceedings of the 1999 International Conference on Semiconductor Manufacturing Operational Modeling and Simulation (SMOMS'99).* San Fransisco, CA. 51-56.

Richardson, R.M. 1996. *Adaptive normalized slack policies for multiple process flow wafer fabrication plants*. Master of Science Thesis. Department of Electrical Engineering, University of Illinois at Urbana-Champaign.

Schoemig, A.K. and M. Mittler 1999. Comparison of Dispatching Rules for Semicondutor Manufacturing Using Large Facility Models. In Proceedings of the 1999 Winter Simulation Conference (WSC'99). Phoenix, AZ.

Seidman, T.I. 1994. 'First come, first served' can be unstable! *IEEE Transactions on Automatic Control*, 39. 2166-2171.