FabTime Cycle Time Management Newsletter

Volume 5, No. 6 **July 2004**

Information

Mission: To discuss issues relating to proactive wafer fab cycle time management

Publisher: FabTime Inc. FabTime sells cycle time management software for wafer fab managers. New features in the current version (6.0) include support for very long filters (e.g. a list of 75 or more operations, or dozens of tools) and automatic identification and warning of filter entries that do not match any database contents

Editor: Jennifer Robinson

Contributors: Douwe van Engen (Philips Semiconductors); James Morrison (IBM); Derek Watson (Micron)

Table of Contents

Welcome

Welcome to Volume 5, Number 6 of the FabTime Cycle Time Management Newsletter! We hope that you are enjoying your summer, and enjoying the industry upturn (if in fact there is an upturn, and if it lasts long enough to be appreciated). This month we have subscriber discussion concerning several previously introduced topics: WIP Utilization Percentage, Dynamic X-Factor, and the Closest-to-Completion-Time Dispatch Rule. Our FabTime user tip of the month concerns setting filter defaults.

For our new topic of the month, we discuss increasing fab throughput through improvements in cycle time constrained capacity. The idea is that fabs always have a buffer of planned idle time on tools, designed to keep cycle times from getting out of control. Through variability reduction, fabs can sometimes squeeze this buffer, without increasing cycle time. In an up market, this can lead to increased sales, from the same equipment set. The financial benefit from this can be substantial, and provides a clear justification for variability reduction / cycle time improvement efforts.

Thanks for reading!—Jennifer

FabTime

Tel: (408) 549-9932
Fax: (408) 549-9941
www.FabTime.com
Sales@FabTime.com

Community News/Announcements

1st International SEMATECH Manufacturing Initiative Symposium on Manufacturing Effectiveness

25-28 OCTOBER 2004
AUSTIN, TX

International SEMATECH Manufacturing Initiative (ISMI) is holding its first Symposium on Manufacturing Effectiveness to address key manufacturing issues through three parallel tracks dealing with productivity, fab design, and statistical methods.

The focus of this symposium is on reducing manufacturing costs through advances in equipment, process, resources, fab design, and manufacturing methods both in existing and next-generation factories.

Papers from selected ISMI projects as well as leading semiconductor device and equipment manufacturers will be presented. A panel of industry executives and experts will discuss trends and strategic thrusts to sustain productivity improvement and profitability

FabTime welcomes the opportunity to publish community announcements. Send them to newsletter@FabTime.com.

FabTime User Tip of the Month

Setting Default Chart Filters

Do you find yourself always using the same set of filters on most of your charts? For example, do you always filter by owner, to only view MFG lots? Or, do you look mainly at one particular product type? If so, you might save time by setting default values for your filters. To do this, follow these steps:

1. Display the chart list: Click on the Charts link in the FabTime toolbar.

2. Enter your preferred filter values in the filter controls to the left-hand side of the screen, and click GO. FabTime displays the message “Defaults have been updated.” Click on the Continue link to return to the Chart list.

3. Create any chart using the FabTime Charts list. Your filters will already be included on the resulting chart page. If you wish to temporarily clear the filters (e.g. for a particular chart), just delete the filter values on the detailed page for that chart, and replace them with whatever values you prefer.

4. To clear the default filter values: Select the text in the each filter and press your Delete key or your Backspace key to clear the text. Click GO, and then click on the Continue link to return to the Chart list.

If you have any questions about this feature (or any other software-related issues), just use the Feedback form in the software.

Subscriber Discussion Forum

WIP Utilization %

Douwe van Engen (Philips Semiconductors) wrote: "As you know we have something similar in our heartbeat tool (I showed it to you during our CT Summit in the Netherlands). Instead of WIP utilization % we called it "Efficiency loss", but the purpose is exactly the same. Dividing the Idle time (with WIP) by the productive time is indeed an improvement (relative measure). Because of transport times, reactions times of operators etc. sometimes as much as half of the standby time on certain tools might be time when WIP is available to be processed.

We use the heartbeat tool in our litho area. The difficulty is that the WIP is not always in front of the tool. Sometimes is stopped one or more operations in front of it (same problem with furnace operations, the WIP is always standing for the cleaning). Another attention point is that only WIP should be measured which can be processed by the operator. We call it "active WIP", and the opposite "Blocked WIP".

Besides this a similar instrument can be used for equipments, which are down. Down and WIP deserved more attention than Down without WIP.

To make it clear where to work on, we developed a nice instrument in our shopfloor control system. It visualizes all the equipments in our fab. The equipment number in this screen is blinking when there is active WIP in front of it, and the equipment state is "in repair" or "standby". The simple message with this is for everyone "Nothing may blink".

Anyway, I strongly believe in this approach."

FabTime Response:

I'm glad to hear that this metric is so consistent with what you're doing at Philips. And I think that you make some excellent points! I'm not sure what to do about the problem of WIP being stopped at other operations (such as in front of a clean step instead of a furnace), but I have already heard other companies mention this issue, too. One company that I talked with used a custom script to project the WIP forward and report it as being in front of the true constraint (e.g. the furnace). In any case, you're certainly right about only including "active WIP."' In our software, we only include WIP that is not on hold, and is qualified to be run on the particular tool. This probably still leaves some issues with reticles, but should capture most of this problem. I especially like your extension of this approach to think about down tools with WIP waiting, and the "nothing may blink"' idea. It sounds like your graphical display is an excellent way to communicate this on the shop floor.

Closest to Completion Time / Shortest-Remaining-Processing-Time Dispatch Rule

James Morrison of IBM wrote in response last month's question about research related to closest-to-completion-time dispatching, submitting the following document:

"An incomplete survey of the SRPT policy, extensions to re-entrant semiconductor manufacturing and recent scheduling directions."

The closest-to-completion dispatching policy is often referred to as the shortest remaining processing time (SRPT) policy and has been extensively studied in the queueing and related literature. It has been shown for M/G/1 queues that the SRPT policy is optimal in the sense of the mean number of lots in the system, though there has been some concern that this optimality is at the expense of lots with longer process times. For re-entrant queueing models (used to study the behavior of semiconductor manufacturing lines), the

SRPT policy is akin to the LBFS (last buffer first served) policy which is provably able to attain the bottleneck throughput. Yet, despite the fact that the SRPT policy is optimal for the M/G/1 queue, simulation studies have shown that other policies outperform the SRPT for models of semiconductor manufacturing lines. In particular, variation reduction policies such as the FSMCT (discussed below) policy can outperform the SRPT and there are other promising policies based on fluid network models which promise to perform as well (and perhaps better) than variation reduction heuristics.

The shortest remaining processing time (SRPT) policy has been studied since before 1966 when the waiting time and in-system time distributions for M/G/1 queues operating under the SRPT policy were demonstrated in [1]. Later in [2], the optimality of the SRPT policy in an M/G/1 system with respect to the number of lots was proven (under certain assumptions). The form of the policy studied assumes that the processing time of each lot is known when it arrives to the queue and that preemption of service is allowed in favor of an incoming lot with shorter processing time. There are variations of the SRPT policy for the single queue in which different assumptions are placed on the information available regarding the processing time (perhaps only the processing time distribution is known until service is complete) and the form of preemption (perhaps preemption is not allowed).

There has been interest in using the SRPT policy for practical applications, such as the scheduling of web servers, and this interest has led to further investigations into the policy's properties. Though the SRPT is known to optimize the mean time lots spend in an M/G/1 queue ([2]), there has been a feeling that large processing time jobs are unfairly penalized in favor of smaller jobs. To understand this phenomenon, [3] and [4] investigated the properties of the policy from the perspective of unfairness using a metric termed slowdown (cycle time of a lot divided by its process time). One approach to addressing the unfairness concerns, akin to the recommendation proposed by FabTime [5] for those considering an SRPT policy, is to increase the priority of lots with waiting time greater than a control limit.

The SRPT policy has been considered for use in semiconductor fabricators, no doubt in part due to the promise of obtaining optimal average cycle times (another reason the policy might be considered is the ease of implementation). A policy analogous to the SRPT policy in the wafer fab is last buffer first served, or LBFS (within each product flow), because as a lot approaches processing steps (or buffers) closer to the end of processing, there is less remaining processing time. Hence, SRPT is a myopic scheduling policy in that it emphasizes those stages of processing closer to the end of the manufacturing line (and thus supports a strategy of attempting to complete as much as possible in the near future – neglecting the fact that later stages of processing must be fed from the early stages of processing). Across multiple product flows, one would have to decide whether to implement the policy globally (so that a flow with very little total process time would have priority over a longer flow except in the latest stages of the longer flow's processing) or to weight in some fashion the priorities of the various product flows.

One theoretical property of the LBFS scheduling policy (akin to SRPT as mentioned above) is that it is provably stable. That is, under certain fairly relaxed assumptions, a re-entrant line operating under an LBFS policy can achieve a throughput as close as desired to the bottleneck limited throughput ([6]). This a nontrivial result as there are scheduling policies which can restrict throughput to below the bottleneck limited throughput (even in the absence of setup and batching losses), see [7].

Given the optimality of the SRPT policy in the M/G/1 queue with respect to the mean number of lots in the system (and hence mean cycle time), and in light of the promising stability properties of the analogous LBFS scheduling policy for re-entrant models of semiconductor manufacturing lines, how does this policy (or policies like SRPT or LBFS) perform? In [8], various scheduling policies are studied via simulation. It is demonstrated that, for the models under consideration, other scheduling heuristics (in particular, the FSMCT – Fluctuation Smoothing for the Mean of Cycle Time – policy) outperform the SRPT policy in terms of mean cycle time and variation of cycle time. More recently, scheduling policies have been developed (based on intimate stability and performance connections between queueing networks and their fluid models) which promise to compete with – and possibly outperform – the reduction variation heuristics of policies such as FSMCT. A recent class of policies derived from fluid model concepts and intended for use by industry is presented in [9]. Thus, while SRPT remains a simple policy to implement, other classes of scheduling policies promise to achieve the goal of cycle time optimization via a careful and calculated selection of the next lot to process.

Though the SRPT policy has a rather august history and has been carefully studied for single queue environments (possibly with multiple servers), there are other less myopic scheduling policies which are more promising for semiconductor manufacturing lines.

References

[1] L. E. Schrage and L. W. Miller, "The queue M/G/1 with the shortest remaining processing time discipline", *Operations Research*, 14:670-684, 1966.

[2] L. E. Schrage, "A proof of the optimality of the shortest-remaining-processing time discipline", *Operations Research*, 16:687-690, 1968.

[3] N. Bansal and M. Harchol-Balter, "Analysis of SRPT Scheduling: Investigating Unfairness", *Proceedings of the 2001 ACM SIGMETRICS international conference on measurement and modeling of computer systems*, 2001, Cambridge, Massachusetts, United States.

[4] M. Harchol-Balter, K. Sigman and A. Wierman, "Asymptotic convergence of scheduling policies with respect to slowdown", *Performance Evaluation*, 49:241-256, 2002.

[5] J. Robinson, editor, FabTime Cycle Time Management Newsletter, 5(5):4, 2004.

[6] J. G. Dai and G. Weiss, "Stability and Instability of Fluid Models for re-entrant lines", *Mathematics of Operations Research*, 21:115-135, 1996.

[7] P. R. Kumar and T. I. Seidman, "Dynamic instabilities and stabilization methods in distributed real-time scheduling of manufacturing systems," *IEEE Transactions on Automatic Control*, 35(3):289-298, 1990.

[8] S. C. H. Lu, D. Ramaswamy, and P. R. Kumar, "Efficient scheduling policies to reduce mean and variance of cycle-time in semiconductor manufacturing plants," *IEEE Transactions on Semiconductor Manufacturing*, 7(3):374-385, 1994.

[9] M. Chen, R. Dubrawski and S. P. Meyn, "Management of demand-driven production systems," *IEEE Transactions on Automatic Control*, 49(5):686-698, 2004.

Dynamic X-Factor

Derek Watson (Micron) wrote: My company has been trying to apply the dynamic x-factor to see if it is a good indicator of cycle time for us. After reading the FabTime newsletters from March 24, 2004 and August 19, 2003 (volumes 4.8

and 5.3) I have begun gathering data for the DXF for individual workstations and correlate them to the cycle time. I was wondering if you could give me more insight into what others have done to use this metric to drive cycle time improvements. Have any of them tried to calculate a correlation to cycle time and calculate what cycle time will be x-time periods later because the DXF is at a certain level right now? If so, how have they determined the lag between the DXF and the resulting cycle time? Have they primarily used this metric fabwide or for individual workstations in order to cut out some of the noise and specifically highlight the problem stations? Any input that you have would be greatly appreciated.

FabTime Response:

While we do know of some companies that have at least experimented with using this metric, we don't have any information at the level of detail of your questions, so we will pose this question for our readers.

Increasing Fab Cycle Time Constrained Capacity

Introduction

The paper today repeated a Merrill Lynch conclusion that semiconductor fab revenues have already peaked for this business cycle. We hope that this isn't the case. But in any event, this seems like a reasonable time to talk about one of the benefits of cycle time reduction: squeezing more capacity out of the same equipment set. Clearly, this benefit is only relevant during a time of high fab loadings. During a downturn, the problem is getting business to fill the fab, rather than squeezing out every last drop of capacity. However, right now, there appear to be fabs out there that are nearing the capacity of their toolsets. This article is about undertaking variability reduction efforts, and using the improvement to increase capacity of a given toolset.

Background

When estimating the capacity of a wafer fab, people typically plan for a 10-15% buffer on all of the tools. This buffer (called by various names, such as "catch-up capacity") exists in reality to ensure that the fab can achieve a reasonable cycle time. If we were to plan to operate all of the tool groups at 100% of capacity, cycle times would rapidly rise out of control. This is due to the behavior that we have discussed many times in this newsletter: in the presence of variability, cycle time generally increases with equipment loading, increasing without bound at 100% utilization. Therefore, people plan to run fabs at some percentage of the maximum theoretical capacity, and expect to achieve a certain cycle time.

This concept was formalized during a project that we worked on 10-12 years ago as "cycle time constrained capacity." Cycle time constrained capacity is the throughput rate at which some target cycle time can be achieved. Cycle time constrained capacity is expressed as a multiple of theoretical cycle time (e.g. 3X-capacity is the throughput rate at which average cycle time is three times raw process time). The cycle time constrained capacity of a fab depends on the shape of its operating curve (the graph of cycle time vs. utilization). The shape of the operating curve depends, in turn, on the amount of variability in the fab. Reducing variability will tend to pull the operating curve downward. This means that for the same start rate, the fab can achieve a lower cycle time. Alternatively, it also means that for a given cycle time target, the fab can choose to increase the cycle time constrained capacity. An example is given below.

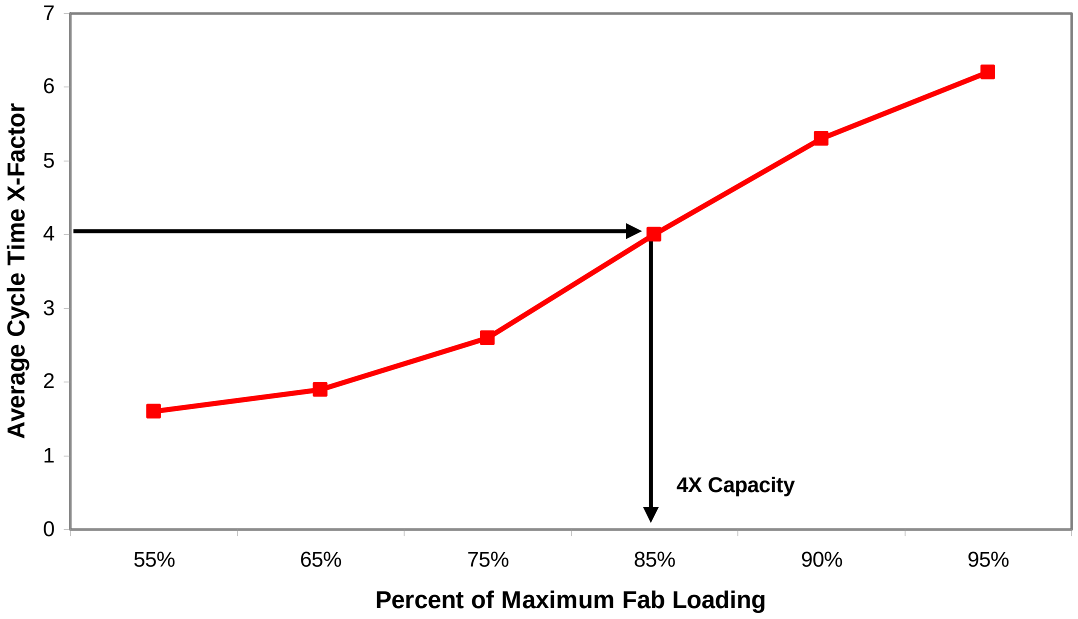
Example

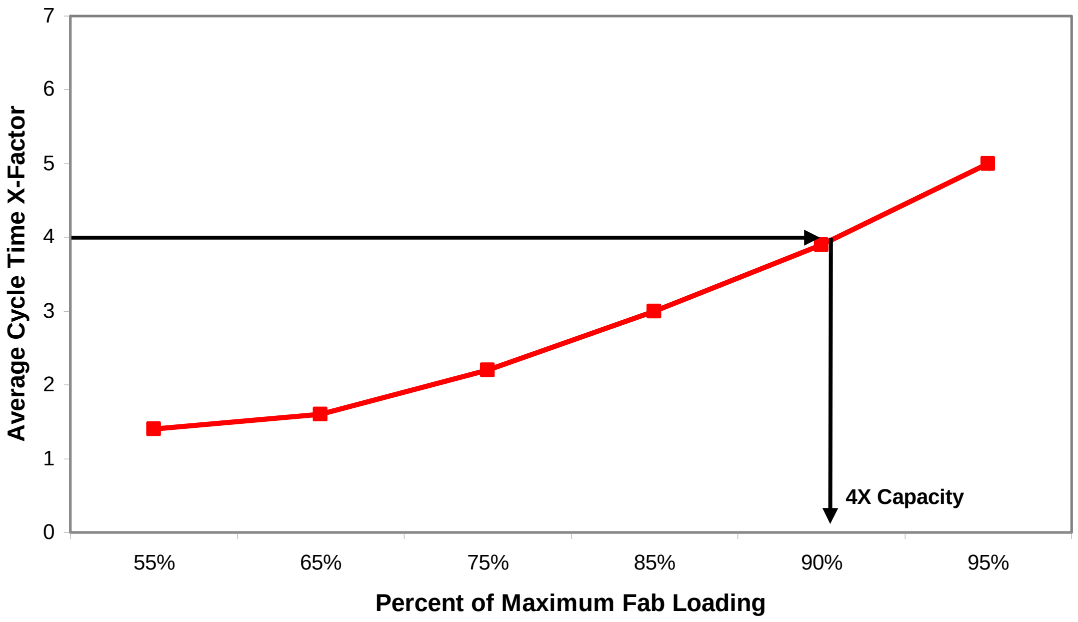
Suppose that based on a simulation model, your fab IE team generates an operating curve for your fab. That is, they run the model at intervals ranging from 55% of maximum loading to 95% of maximum loading, and record the average cycle time / raw process time at each start rate (with the same toolset and relative product mix for all runs). An example of the results might look like the following chart. (This data was generated using the Factory Explorer(r) capacity analysis and simulation tool, distributed by Wright Williams & Kelly, used with their permission. The data is based on a sample fab model with five products, averaging 140 steps per process flow, with a maximum theoretical capacity of 3000 wafer starts per week. The average line yield for the fab is 70%, so the maximum throughput is 2100 wafers per week.)

For this example, the 4X capacity is at 85% of the maximum loading for the fab. That is, at a start rate of 0.85*3000 = 2550 wafer starts per week, the average cycle time is 4 times theoretical. The corresponding throughput rate (after accounting for line yield) is 0.7*2550 = 1785 wafer outs per week.

Now suppose that the fab undertakes some cycle time reduction activities, represented by reduced variability in the simulation model. For this example, we reduced the variability of arrivals into the fab, and also made downtimes shorter and more frequent (without varying the percentage of time that tools spent down). We did not make any changes that would alter the theoretical maximum capacity of the fab (3000 wafer starts per week). The resulting cycle times from the simulation are shown at the top of the next page.

Sample Operating Curve for a 5-Product Wafer Fab, Default Variability





Sample Operating Curve for a 5-Product Wafer Fab, Reduced Variability

For this example, the 4X capacity is slightly above 90% of the maximum loading for the fab. Conservatively, we can use the 90% value. The corresponding start rate = .90*3000=2700 wafer starts per week. The corresponding throughput rate is .7*2700 = 1890 wafer outs per week. This means that if we could reduce variability in this fab, as in the example, we could maintain a 4X cycle time target, but increase outs from 1785 wafers per week to 1890 wafers per week, a more than 5% increase. And, as we know from many of the past articles in this newsletter, there are a variety of ways to reduce variability in a fab. Several are listed in the next section.

Some Ideas for Improving Cycle Time Constrained Capacity

This list is was first included back in Volume 1, Number 7 of the newsletter. It is focused on operational changes to reduce variability, and hence improve cycle time constrained capacity. It is not meant to be a comprehensive listing, but rather, to suggest some places to start.

- Eliminate large minimum batch size requirements for all but very highly loaded tools.
- Cross-train equipment maintenance personnel, to reduce long delays waiting for the right repair person.
- Reduce tool dedication.
- Explore process changes to alleviate single-pass operations, e.g. operations that can only be performed on a single piece of equipment.
- Cross-train regular operators to handle more types of equipment, and to balance schedules.
- Change preventive maintenance schedules to minimize variability.
- Modify setup avoidance policies to ensure that low-volume products are not excessively delayed.
- Reduce transfer lot batch sizes.
- Modify lot release policies to smooth flow through the early steps of the process (lower variability).
- Explore batching rules, to make sure that all lots that can be batched together are batched together (eliminate unnecessary waiting to form batches)
- Check batching and setup assumptions for rework wafers. The entire parent lot is usually delayed whenever the rework

wafers are waiting for processing. Also make sure all operations within the rework loop are necessary.

Financial Implications

Suppose that the fab in the above example has a capital equipment base of $100 million. Assuming five-year straight-line depreciation, we know the fab must generate at least $20 million annually in profit simply to cover depreciation. If we can increase daily wafer ships by 5%, and these wafers can be sold at the same profit margin, then we have an increase in annual profit of $1,000,000. This can also be expressed as a monthly net cash benefit of $83,333. Obviously, the assumption that the increased wafers can be sold, and can be sold at the same profit margin, will not always hold. But, as this analysis shows, for times when it does hold, any efforts to increase daily wafer shipments, while maintaining the same toolset and the same cycle time, can have a significant payoff.

Conclusions

The relationship between cycle time, capacity, and variability makes up the operating curve for a fab. For a fixed toolset and a given amount of variability, cycle time is a function of start rate (or throughput rate). Reducing variability for the fab will pull the operating curve downward and towards the right. Assuming no significant changes in the fab yield rates, this gives the fab a choice: a) keep the same start rate, and achieve lower cycle times, or b) keep the same cycle time, and start more wafers. During times when capacity is a premium, this second approach can generate a nice dollar benefit.

Closing Questions for FabTime Subscribers

Do you think that the industry downturn has ended? Do you think that the upturn has already peaked? If you could reduce variability, would you use it to improve cycle time, increase starts, or both?

Further Reading

■ S. Brown, F. Chance, J. W. Fowler, and J. K. Robinson, "A Centralized Approach to Factory Simulation", *Future Fab International*, Vol. 3, 83-86, 1997. This paper describes the application of cycle time constrained capacity as a performance metric at Siemens Semiconductor Division (now Infineon Technologies). The paper is available for download from www.fabtime.com/abs_FutureFab.shtml.

■ D. Y. Burman, F. J. Gurrola-Gal, A. Nozari, S. Sathaye, and J. P. Sitarik, "Performance Analysis Techniques for IC Manufacturing Lines," *AT&T Technical Journal*, Vol. 65, No. 4, 46-57, 1986. This paper defined fab capacity as the start rate that gives a "reasonable" WIP level, though without explicitly defining "reasonable."

■ F. Chance, J. K. Robinson, J. Fowler, O. Gihr, B. Rodriguez, and L. W. Schruben, "A Design of Experiments Methodology for Semiconductor Wafer Fab Capacity Planning" *SEMATECH Technology Transfer #95062860A-TR*, 1995. This paper used cycle time constrained capacity as a formal performance metric in evaluating the impact of several variables on wafer fab capacity.

■ J. W. Fowler, S. Brown, H. Gold, and A. Schoemig, "Measurable Improvements in Cycle-Time-Constrained Capacity," *Proceedings of the 6th IEEE/UCS/SEMI International Symposium on Semiconductor Manufacturing (ISSM)*, October 6-8, 1997, San Francisco, A21-A24. This paper describes a simulation-based study performed to evaluate the operating practices of a high-volume, multiple-product semiconductor fab. The paper is available for download from www.fabtime.com/abs_SiemFab.shtml.

■ J. Robinson and F. Chance, "Improving Factory Cycle Time Through Changes at Non-Bottleneck Tools," *FabTime Cycle Time Management Newsletter*, Vol. 1, No. 7, 2000.

Subscriber List

Total number of subscribers: 1653, from 394 companies and universities. 25 consultants.

Top 10 subscribing companies:

- Analog Devices (80)
- Intel Corporation (78)
- Motorola Corporation (57)
- Infineon Technologies (51)
- STMicroelectronics (49)
- Philips (45)
- Seagate Technology (41)
- Micron Technology, Inc. (40)
- Texas Instruments (40)
- AMD/ Spansion (35)

Top 3 subscribing universities:

- Arizona State University (10)
- Virginia Tech (10)
- Nanyang Technological University (6)
- University of California – Berkeley (6)

New companies and universities this month:

- 40-30
- ACS
- Advanced MicroSensors Inc.
- Ben-Gurion University of the Negev
- DotChain Consultant, Inc.
- Eyelit Inc.
- First Technologies
- General Motors
- PennWell Corporation
- Robert Bosch GmbH
- Tru-Tech Electronics
- UMCi

Note: Inclusion in the subscriber profile for this newsletter indicates an interest, on the part of individual subscribers, in cycle time management. It does not imply any endorsement of FabTime or its products by any individual or his or her company.

There is no charge to subscribe and receive the current issue of the newsletter each month. Past issues of the newsletter are available for a small fee from FabTime's Amazon zShop, at www.amazon.com/shops/fabtime.

To subscribe to the newsletter, send email to newsletter@FabTime.com, or use the form at www.FabTime.com/newsletter.htm. To unsubscribe, send email to newsletter@FabTime.com with "Unsubscribe" in the subject. FabTime will not, under any circumstances, give your email address or other contact information to anyone outside of FabTime without your permission.

FabTime® Cycle Time Management Training

"It was helpful to see best-in-class methods for wafer fab cycle time management. Discussing these matters in-depth with you was quite valuable, as we could ask questions specific to our fab and processes."

Shinya Morishita
Manager, Wafer Engineering
TDK Corporation

Course Code: FT105

This course provides production personnel with the tools needed to manage cycle times. It covers:

- Cycle time relationships
- Metrics and goals
- Cycle time intuition

Price

$4950 plus travel expenses. On-site delivery for up to 15 participants, each additional participant $195. Discounts available for multiple sessions.

Interested?

Contact FabTime for a quote.

FabTime Inc.
325M Sharon Park Drive #219
Menlo Park, California 94025
Phone: +1 (408) 549-9932
Fax: +1 (408) 549-9941
Email: Sales@FabTime.com
Web: www.FabTime.com

Do you make the best possible decisions?

- Do your supervisors possess good cycle time intuition?
- Are you using metrics that identify cycle time problems early?
- Can you make operational changes to improve cycle time?

FabTime's Cycle Time Management Training is a one-day course designed to provide production personnel with an in-depth understanding of the issues that cause cycle time problems in a fab, and to suggest approaches for improving cycle times. A two-day version is also available upon request.

Prerequisites

Basic Excel skills for samples and exercises.

Who Can Benefit

This course is designed for production personnel such as production managers, module managers, shift supervisors, hot lot coordinators, and production control.

Skills Gained

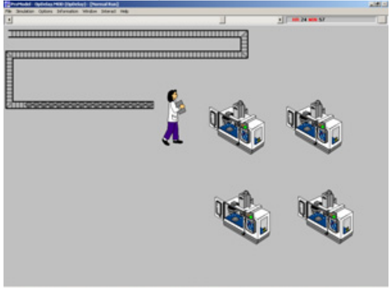
Upon completion of this course, you will be able to:

- Identify appropriate cycle time management styles.
- Teach others about utilization and cycle time relationships.
- Define and calculate relevant metrics for cycle time.
- Teach others about Little's law and variability.
- Quantify the impact of single-path tools and hot lots.
- Apply cycle time intuition to operational decisions.

Sample Course Tools

Excel Cycle Time Simulator

Staffing Delay Simulator



Additional Half-Day Modules

- Executive Management Session.
- Site-Specific Metrics Review.
- Capacity Planning Review and Benchmark.