Fab Agility: Quantifying the Economic Impact of Cycle Time Waste

The need to focus the fab productivity discussion on waste reduction has been widely discussed and recognized, [1-3] yet, the industry still lacks basic tools to comprehend the complete productivity picture. For example, cycle time is relatively intractable and its value is difficult to quantify because it consists of missed opportunity rather than incurred costs. In this study we propose a basic toolkit for definition and valuation of cycle time, an important first step towards focusing attention and action on improving fab agility.

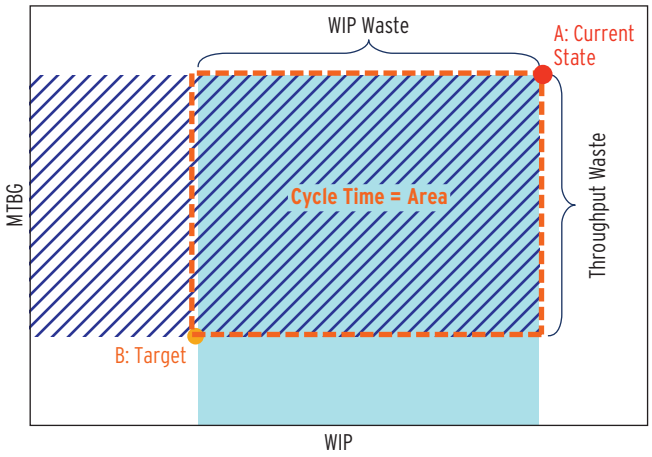
Interest in reducing cycle time is growing across business models and market segments, from manufacturers of microprocessors (Intel and AMD), memory (Samsung, Inotera, Spansion), foundry (TSMC), and even development fabs.[4-17] However, since there is no unified definition of cycle time, we cannot calculate its value.

To remedy this, we are working with ISMI and SEMI to formalize a simple set of definitions, building on Little's Law and existing SEMI Standards (primarily E124-1103). Little's Law states that average work-in-process (WIP) equals the product of arrival rate and average cycle time, as follows:

(Eq.1)

$$Average\ WIP = Arrival\ Rate \cdot Average\ Cycle\ Time$$

Simple manipulation and reflecting the likelihood of loss during fab processing (substituting output rate, like finished units out, for input rate, like arrival rate) yields



▲ **FIGURE 1:** Visualizing cycle time and waste. MTBG is the inverse of throughput, run rate of successful completion.

the definition of cycle time as the ratio of average WIP to output rate (finished units out):

(Eq.2) $$Cycle\ time = \frac{1}{WIP\ turnover} = \frac{average\ WIP}{finished\ units\ out}$$

Note that this definition is based on a top-down view, as opposed to classical definitions based on the sum of cycle times of individual process steps. Also, this definition is not limited to fab-wide operations, but can be applied to any productive part of it. Cycle time is typically calculated per wafer, and possibly per die.

To enable graphic representation, we have adopted a metric called "mean time between good units out," or MTBG. MTBG is simply the inverse of the throughput rate of the fab (or some element of it). In that case, cycle time is the product of WIP and MTBG:

(Eq.3) $$Cycle\ Time = average\ WIP \times MTBG$$

This relationship is shown in Figure 1, for example, where two states are compared: A, or current state, vs. B, or target (ideal) case. The rectangle in orange, with corners A and B, can be viewed as the cycle time waste, or "waste" generated in the fab.

The relative size of these rectangles can indicate the overall level of waste. For example, the ratio of rectangle B (area defined from origin to B) to rectangle A (area defined by origin to A) is equivalent to the primary indicator advocated by Hyder, namely, load-adjusted cycle time equivalent or LACTE. [5]

DRIVERS OF CYCLE TIME

Analysts and practitioners have traditionally quantified the economic impact of cycle time waste using a "bottom-up" approach, typically focusing on modeling the impact of longer time to market and reduced differentiation,[18-21] slower yield learning, or simply inventory carrying costs. [14, 22-24]

Our approach was instead to go to "first principles" based on the assumption that a fab's objective is to maximize profitability. Profitability is generally increased with utilization, so with a simple representation like Eq. 4, we can see a relationship as expressed in Figure 2 for any given time period.

(Eq.4)

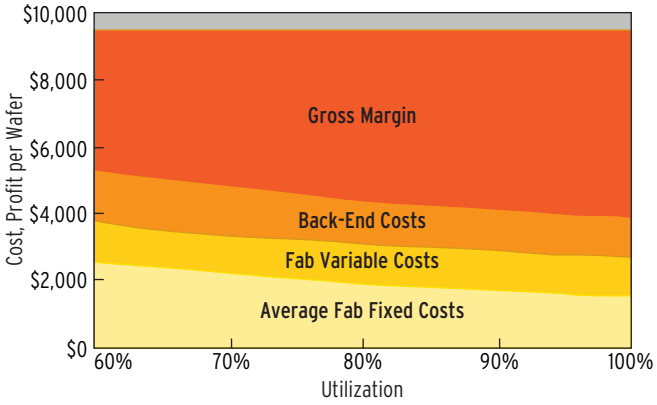
$$GM = u \cdot Capacity \cdot Availability \cdot (yield \cdot ASP - UnitVarCost) - TotalFixedCost$$

where GM is gross margin, u is utilization of available capacity, ASP is average selling price, $UnitVarCost$ is variable cost per unit.

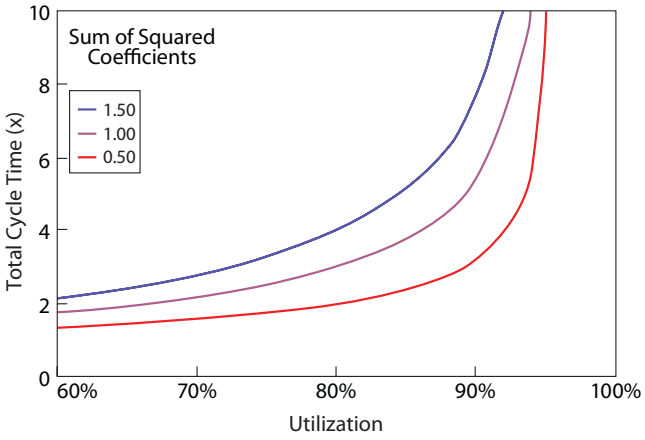
As a result, fabs may be tempted to push utilization to very high levels; however, extremely high utilization is rare due to the conflict between agility and utilization. This relationship is known as the "fab operating curve," illustrated in Figure 3. The fab operating curve has been extensively analyzed in queuing theory[4] and one model is shown in Eq 5.

(Eq.5) $CT_q = \frac{(C_a^2 + C_e^2)}{2} \cdot \frac{u}{(1-u)} \cdot T_e$ (G/G/1): one processing path

In this equation, C_e is coefficient of process time variation, C_a is coefficient of arrival time variation, u stands for utilization of available capacity, and T_e is average processing time.



▲ **FIGURE 2:** Economics driven by utilization, given the following assumptions: $9,500 revenue per wafer out, gross margin is 55%, wafer processing costs are 70% of total costs, and fixed costs are 55% of total costs at target utilization of 85%.



▲ **FIGURE 3:** Fab operating curve. Total cycle time is shown for as a multiple (X) of average processing time for different levels of variation (represented by the sum of squared coefficients) and utilization; more variability (higher coefficients) causes longer cycle time at any level of utilization.

VALUE OF CYCLE TIME: FRAMING THE MODEL

As is evident from the fab operating curve shown in Figure 3, the cycle time penalty tends to stand in the way of increased fab utilization. As utilization increases, cycle time increases and reaches unacceptable levels, forcing fabs to limit fab loading. This relationship between increasing cycle time and increasing fab utilization provides an important hint about how to quantify the cost of cycle time waste. Instead of divining the way fabs should run their business, it is more straightforward to observe how they do it, then impute from their actions and priorities the relative value of cycle time.

Fabs use extensive analytical and planning tools to strike an optimal trade-off. The rationale is to extend cycle time to increase utilization up to the point where additional increases would make cycle time unacceptably long (resulting in lost business, worse operational performance). Alternatively, they cut cycle time up to the point where further reductions would erode utilization and profitability more than the contribution of shorter cycle time.

Based on this logic, we can identify a "sweet spot" where the value of incremental profit (from higher utilization) roughly equals the incremental economic cost (from longer cycle time). This allows us to find the "shadow price" of cycle time by observing the fab's actual performance. From this we can attribute what we call the Lost Opportunity due to waste, or LOW function, to each specific user situation.

CALCULATING THE VALUE OF CYCLE TIME

The mathematical approach uses the concept of "elasticity" (η), or the ratio of the percent change in one variable to the percent change in another variable.

Step 1: Determine the sensitivity of fab owner's profitability to fab utilization, based on Eq.4:

$$\text{(Eq.6)}\quad \eta_u^{GM} = \frac{\partial GM/GM}{\partial u/u} = \frac{\partial GM}{\partial u}\cdot\frac{u}{GM} = FC\%\cdot\left(\frac{1}{GM\%}-1\right)+1$$

In this equation, *GM* is gross margin, *GM%* is gross margin % of revenue; u is utilization of available capacity; *FC%* represents fixed costs as a % of total manufacturing costs.

Step 2: Determine the sensitivity of cycle time (in queue) to fab utilization using Eq.5:

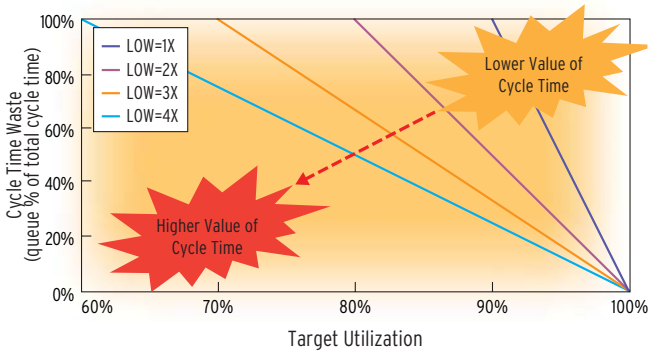
$$\text{(Eq.7)}\quad \eta_u^{CT_q} = \frac{\partial CT_q/CT_q}{\partial u/u} = \frac{\partial CT_q}{\partial u}\cdot\frac{u}{CT_q}$$
$$= \frac{(C_a^2+C_e^2)}{2}\cdot\frac{T_e}{(1-u)^2}\cdot\frac{u\cdot(1-u)\cdot(C_a^2+C_e^2)}{2\cdot u\cdot T_e} = \frac{1}{(1-u)}$$

where CT_q is cycle time in queue, C_e is coefficient of process time variation, C_a is coefficient of arrival time variation, and T_e is average processing time.

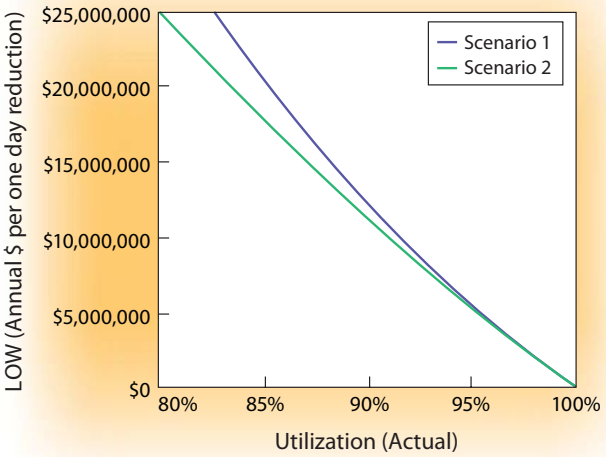
Step 3: Divide the sensitivities to determine overall response of profitability to cycle time (in queue):

$$\text{(Eq.8)}\quad \eta_{CT_q}^{GM} = \frac{\eta_u^{GM}}{\eta_u^{CT_q}} = \left[FC\%\cdot\left(\frac{1}{GM\%}-1\right)+1\right]\cdot(1-u)$$

Step 4: Adjust the calculation to represent the response of profits to changes in total cycle time and to convert this profit impact to "equivalent costs" (shown as a % of wafer processing costs):



▲ **FIGURE 4:** Fab operational state–indifference curves. The fab operational state drives the opportunity cost; fabs move to indifference curves with lower LOW multiple as their target utilization increases (moving to the right in the above graphic) and/or as their cycle time lengthens (moving up in the above graphic).



▶ **FIGURE 5:** Cost of lost opportunity due to waste (LOW).

	Scenario 1	Scenario 2
Wafer per day	3,000	1,000
Cycle time, days	40	60
Time in queue	80%	90%
Target utilization	90%	90%
At targeted utilization:		
Cost per wafer	$2,000	$4,500
GM%	35%	55%
Fixed Costs %	70%	60%
Front End %	70%	50%

$$(\text{Eq.9})\quad \eta_{CT}^{EqCost} = \eta_{CT}^{GM} \cdot \frac{GM}{FE\% \cdot (Rev - GM)}$$

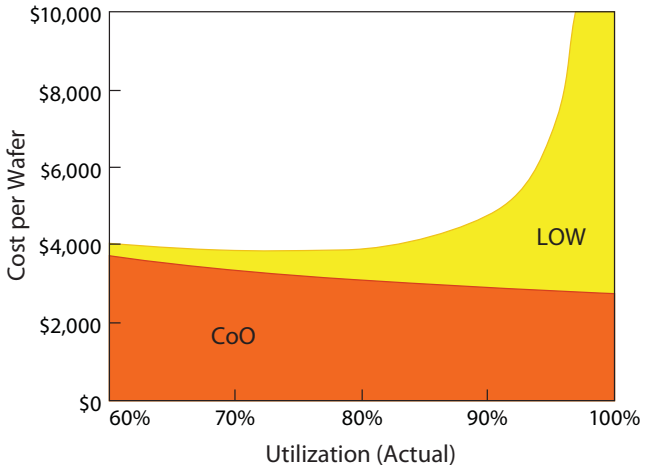
$$= \frac{\eta_{CT_q}^{GM} / \eta_{CT_q}^{CT}}{FE\% \cdot (1/GM\% - 1)} = \frac{(1-u)}{r} \cdot \frac{FC\% + \frac{GM\%}{1-GM\%}}{FE\%}$$

where Rev is Revenues, $r = \frac{CT_q}{TotalCycleTime}$ and $FE\%$ represents the share of front-end (fab processing) costs of total chip costs.

RESULTS AND INTERPRETATION

Equation 9 provides a simple model to evaluate the LOW (or the economic value of cycle time) based on information about a fab's current operational characteristics and business model.

It is instructive to consider the equation in some detail. The first argument $(1-u)/r$, reflects the operational characteristics of the fab. The behavior of this term is consistent with our intuition. Mathematically, it drives the opportunity cost of cycle time (LOW) to be lower, the higher the currently targeted utilization (u) is and the longer the relative queue time (r) is. This mathematical behavior and its graphical representation (Figure 4) reflects the real-world situation. Fab managers who choose to run their factories at high utilization/high queue time regimes (upper right-hand corner of Figure 4) are responding to the expectations of the overall enterprise by putting throughput ahead of time to market. Their actions reveal the relatively low value that they assign to cycle time. The reverse would be true in situations of lower utilization and queue time (lower left-hand corner of Figure 4).



▲ **FIGURE 6:** Wafer CoO vs. cost of LOW using the following assumptions similar to Figure 2. At target utilization of 85%, wafer processing costs are $3,000, comprising 70% of total production costs, 55% of which are fixed, and gross margin is 55%.

The second argument in Eq 9 reflects the basic financial model of the particular fab. Figure 5 illustrates the calculation of the LOW function for two fabs with different business models. Even for high-utilization fabs like these, an extra day of cycle time costs many millions in missed economic opportunity (this is consistent with other observations).[5, 18]

More importantly, this model allows us to contrast the cost of lost opportunity due to waste with the direct fabrication costs (CoO). There is always some element of waste, and this adds to the total wafer cost. As utilization rates are pushed higher, this waste component grows, and can grow to the point where it more than doubles the total economic cost per wafer. Figure 6 shows this comparison for a given scenario (fab-wide situation). LOW clearly becomes comparable to and even overwhelms CoO in typical situations.

Interestingly, total economic costs can actually be minimized at lower utilization (77%), saving 4% vs. target utilization.

CONCLUSIONS

We have shown that cycle time can be clearly defined and that the lost opportunity incurred by fab owners due to cycle time waste may be on the same order of magnitude as the direct fabrication costs.

Can this lost opportunity be addressed? A detailed approach is beyond the scope of this paper, but we have discussed elsewhere [1, 3] that the key levers for cycle time waste reduction consist of:

a. Predictable, low-variability, "smart" tools
b. High-capacity AMHS and factory systems

c. Universal, high-productivity, single-wafer proces tools

This report lays out a straightforward toolkit for the definition and measurement of cycle time, waste, and its economic impact. It is our hope that with the greater clarity and visibility into cycle time waste and its economic implications that these tools provide, agility will receive the recognition it deserves within the fab productivity agenda.

ACKNOWLEDGMENT

The authors wish to thank Applied's Dr. Dajiang Xu for his critical contributions in preparing this report.

REFERENCES

(1) Singer, P., "Redefining fab productivity from a waste perspective," Solid State Technology, May 2008.

(2) Singer, P., "New roadmap targets cycle time, waste," Solid State Technology, June 2008.

(3) Peters, L., "Technology Is No Longer King," Semiconductor International, June 20, 2007.

(4) Shanthikumar, et al., "Queueing Theory for Semiconductor Manufacturing Systems: A Survey and Open Problems," IEEE Transactions on Automation Science and Engineering, vol.4, no.4, pp.513-522, Oct. 2007.

(5) Hyder, J., "A Primer: Lean Manufacturing Applications for the Semiconductor Industry," presented at SAME-TEC 2005, Maricopa Advanced Technological Education Center (MATEC) and Technician Performance Improvement Council (TMIC).

(6) Woolverton, A. et al., "Fast cycle time in high-mix technology development and manufacturing," IEEE International Symposium on Semiconductor Manufacturing Conference Proceedings, pp.57-60, 1999.

(7) Govind, N., et al., "Operations Management in Automated Semiconductor Manufacturing With Integrated Targeting, Near Real-Time Scheduling, and Dispatching,", IEEE Transactions on Semiconductor Manufacturing, vol.21, no.3, pp.363-370, Aug. 2008.

(8) Ward, M., "Fab cycle time improvement through inventory control: a wafer starts approach," Thesis (M.B.A.)--Massachusetts Institute of Technology, Sloan School of Management; and, (S.M.)--Massachusetts Institute of Technology, Engineering Systems Division; in conjunction with the Leaders for Manufacturing Program at MIT, 2007.

(9) Sadjadi, F., Baker, T., "Comprehensive cycle time reduction program at AMD's fab25," 2001 IEEE International Semiconductor Manufacturing Symposium, pp.95-98, 2001.

(10) Leachman, R., Kang, J., Lin, V., "SLIM: Short Cycle Time and Low Inventory in Manufacturing at Samsung Electronics," Interfaces, 2002 INFORMS, Vol. 32, No. 1, pp. 61–77, January–February 2002.

(11) Kan Wu, "An examination of variability and its basic properties for a factory," IEEE Transactions on Semiconductor Manufacturing, vol.18, no.1, pp. 214-221, Feb. 2005.

(12) Inoue, T. et al., "Study of cycle time caused by lot arrival distribution in a semiconductor manufacturing line," IEEE International Symposium on Semiconductor Manufacturing, pp. 115-118, 13-15 Sept. 2005.

(13) Hsu, Jung Pin, et al, "Managing Process Constrain Effectively to Enable Fab Cycle Time Reduction," IEEE International Symposium on Semiconductor Manufacturing, pp.293-295, 25-27 Sept. 2006.

(14) Wen-Chi Chang, et al., "Yield improvement through cycle time and process fluctuation analyses,", 2001 IEEE International Semiconductor Manufacturing Symposium, pp.267-270, 2001.

(15) Cheng Chung Chien, et al., "Cycle time learning curve in semiconductor foundry industry," IEEE International Symposium on Semiconductor Manufacturing, pp. 359-360, 13-15 Sept. 2005.

(16) Yon-Chun Chou, Chuan-Shun Wu, "Economic analysis and optimization of tool portfolio in semiconductor manufacturing," Semiconductor Manufacturing, IEEE Transactions on , vol.15, no.4, pp. 447-453, Nov 2002.

(17) Janakiram, M., "Cycle time reduction at Motorola's ACT fab," Advanced Semiconductor Manufacturing Conference and Workshop Proceedings, pp.465-469, 12-14 Nov 1996.

(18) Leachman, R., Lin, V., Palezzato, P., "Scheduling Dedicated Lithography Equipment," presented at The Role of Optimization in Supply Chain Management Symposium, University of Minnesota's Institute for Mathematics and Its Applications, September 2002.

(19) Leachman, R., Ding, S., "Integration of Speed Economics into Decision-Making for Manufacturing Management," International Journal of Production Economics, Special Section on Building Core-Competence through Operational Excellence, Volume 107, Issue 1, pp.39-55, May 2007.

(20) Leachman, R.C. et al, "Economic Efficiency Analysis of Wafer Fabrication," IEEE Transactions on Automation Science and Engineering, vol.4, no.4, pp.501-512, Oct. 2007.

(21) Page, M., "The free factory: cutting cycle time and gaining output," IEEE/SEMI 1996 Advanced Semiconductor Manufacturing Conference and Workshop, Proceedings., vol., no., pp.146-150, 12-14 Nov 1996.

[22] Chen, K.C. et al., "Cycle time and process improvement by single wafer thermal processing in production environment," 10th IEEE International Conference of Advanced Thermal Processing of Semiconductors, 2002, pp. 171-176, 2002.

[23] Cunningham, S.P.; Shanthikumar, J.G., "Empirical results on the relationship between die yield and cycle time in semiconductor wafer fabrication," IEEE Transactions on Semiconductor Manufacturing, vol.9, no.2, pp.273-277, May 1996.

[24] Nemoto, K.; Akcali, E.; Uzsoy, R.M., "Quantifying the benefits of cycle time reduction in semiconductor wafer fabrication," IEEE Transactions on Electronics Packaging Manufacturing, vol.23, no.1, pp.39-47, Jan 2000.

Authors: Iddo Hadar and Eric Englhardt.
For additional information, please contact
Iddo_Hadar@amat.com