Proceedings of the 2011 Winter Simulation Conference
S. Jain, R.R. Creasey, J. Himmelspach, K.P. White, and M. Fu, eds.

USING STATIC CAPACITY MODELING AND QUEUING THEORY EQUATIONS TO PREDICT FACTORY CYCLE TIME PERFORMANCE IN SEMICONDUCTOR MANUFACTURING

Roland Schelasin

National Semiconductor Corporation
Industrial Engineering
5 Foden Road
South Portland, ME 04106, USA

ABSTRACT

In order to maximize asset utilization and meet customer delivery requirements manufacturing facilities are driven by two key metrics: utilization of production capacity and cycle time. The science of factory physics indicates that queuing theory algorithms relying on an understanding of factory variability at the equipment level can make it possible to use static calculations to estimate factory cycle times. This approach has been frequently dismissed as insufficiently accurate due to the difficulty associated with determining the required variability factors. This paper outlines a method using queuing theory equations together with targeted historical data to estimate total cycle times. Initial validation results indicate that the approach can provide sufficiently accurate results to be useful in manufacturing decision making. Equations, data requirements, and validation results are presented. Opportunities for improvement of the methodology as well as further refinement of the equations for calculating equipment specific variability factors are also discussed.

1 INTRODUCTION

With the ever present focus on maximizing asset utilization as well as meeting customer on-time delivery requirements manufacturing facilities are driven by two key metrics: utilization of production capacity and cycle time. Companies in the analog semiconductor business face a unique challenge due to a high number of different products produced simultaneously and shrinking product life cycles in markets such as consumer products. The increased level of complexity together with downward price pressure necessitates a scientific approach with pragmatic results toward balancing these key performance metrics.

Wafer start requirements in combination with required process step routings and planned equipment performance parameters provide the basis for typical factory capacity modeling used to predict factory utilization. Given the increased computing power in today's desktop computers this can be accomplished with a sufficiently high degree of accuracy through relatively straight forward spreadsheet calculations. While these spreadsheet models have been and continue to be used extensively to estimate factory utilization, in the past it has been mostly left up to dynamic factory simulation models to provide cycle time predictions.

The science of manufacturing as put forth by Hopp and Spearman (2001) indicates that queuing theory algorithms relying on an understanding of factory variability at the equipment set level can make it possible to use static calculations to estimate factory cycle times. This approach to estimating cycle time has been frequently dismissed as insufficiently accurate due to the difficulty associated with determining the required variability factors. This paper outlines a historical data driven approach utilizing the general multi-machine Kingman equation to calculate the requisite operation specific queue times. The G/G/m model is used as a starting point to evaluate the feasibility of the approach. Work is under way to investigate whether other more detailed models such as those proposed by Whitt (1993) will produce even better results.

While the use of historical data to estimate factory variability was initially chosen merely as a starting point, it was found to be an appropriate choice since inter-arrival variability will always have to be an approximation that is best founded in historical data as long as it is available.

Comparison of initial validation results against actual factory cycle time performance of a complex wafer fabrication facility indicate that the current approach can provide sufficiently accurate results to be useful in manufacturing decision making.

Equations, associated data requirements, and validation results are presented. Opportunities for improvement of the current methodology employed as well as plans for further refinement of the equations used to calculate equipment specific variability factors, as also suggested by previous research, are discussed.

2 METHODS FOR CALCULATING CYCLE TIME

Several different approaches for calculating estimated cycle times in manufacturing systems have been developed over the years. It is beyond the scope of this paper to present a comprehensive list; however, a brief review of the key methodologies is appropriate.

Chuang and Huang (2002) classify the approaches to cycle time estimation into four major categories: analytical, simulation, statistical analysis, and hybrid methods. The first two have historically been those mostly used to help predict cycle time in semiconductor manufacturing. While the latter two appear to provide viable alternative solutions they were only briefly reviewed as part of this work and are therefore not discussed here. For a more thorough review of all four cycle time estimating methods including statistical analysis and hybrid methods see Pearn, Tai, and Lee (2009) as well as Shanthikumar, Ding, and Zhang (2007).

Simulation has been the de facto standard for cycle time estimation for many years. It has been used primarily in the area of dispatch and scheduling research. Due to the inherent ability in the simulation approach to build a detailed model that very closely represents the complexity of an actual factory, simulation models can produce very accurate cycle time predictions. Simulation success though can only be guaranteed as long as the proper level of resources is applied to ensure model validity. The proper resource commitment has to be made both to build the model as well as to maintain the model at the proper level of detail required to deliver the desired level of accuracy. There appears to be general consensus within the modeling and simulation community that for truly meaningful results the level of effort required is non-trivial.

Analytical methods developed around queuing theory equations have been widely discussed and expanded over the last few years. As pointed out by Shanthikumar, Ding, and Zhang (2007), while analytical methods are significantly faster than simulation models, their accuracy has been found to be lacking when compared to that of simulation models. The cause for this is stated to be in part due to the fact that semiconductor factory behavior is too complex to be represented by one single queuing model (FabTime, 2011). The potential of using several different queuing models to improve model accuracy will be further elaborated later in this paper. It has also been frequently suggested that queuing theory equations can be useful but only at the individual workstation and work cell level. The work presented below has shown that combining basic queuing theory equations with targeted historical data can provide a way to estimate factory complexity and provide cycle time results at a useful level of accuracy.

3 THEORY OVERVIEW

3.1 Method

In their discussion of the roots of randomness Hopp and Spearman (2001) suggest that the nature of randomness can be viewed one of two ways. The first view is that everything is deterministic, the second that everything actually behaves randomly. Regardless of the view adopted, whether due to lack of knowledge or an inability to ever get that knowledge, there is agreement that we do not have a complete scientific understanding of randomness.

Hopp and Spearman (2001) conclude that the goal therefore is to find sufficiently robust policies and methods that will work *most of the time*. Accordingly the objective of this work was to arrive at a methodology to predict cycle time that would work *most of the time*. Additionally the methodology needed to require a reasonable level of resources to maintain and use on an ongoing basis commensurate with the benefits it provides. The method found that appears to meet both of these objectives is the analytical approach of determining wait time using equations based on queuing theory. The requisite factory variability is derived using historical factory performance data representative of the anticipated factory state.

3.2 Equation

The fundamental equation used is the well known Kingman equation with the underlying assumptions that both inter-arrival times and process times conform to a general distribution, that the machines are all parallel processing machines, and that the maximum number of jobs in the system is infinite. Using *Kendall's notation* this is also known as the G/G/m queuing system.

The equation for the wait time in queue for this system is defined as:

$$\mathrm{CT}_q = \left(\frac{c_a^2 + c_e^2}{2}\right)\left(\frac{u^{\sqrt{2(m+1)}-1}}{m(1-u)}\right)t_e \tag{1}$$

where:

c_a = the coefficient of variation of inter-arrival times to the equipment set
c_e = the coefficient of variation of effective process times at an equipment set
m = the number of machines in the equipment set
u = utilization of an equipment set
t_e = effective process time.

4 SOLVING THE VARIABILITY DILEMMA

4.1 The Variability Challenge

In its simplified form, the Kingman equation (1), also called the VUT equation, reads:

$$\mathrm{CT}_q = VUt_e \tag{2}$$

In a predictive modeling environment the utilization component U can be calculated from modeled demand OEE, planned availability, planned efficiency, and planned equipment count m.

The effective process time t_e by operation can be obtained from either planned or historical manufacturing data and is the combination of raw process time, equipment specific overhead, and move time.

The variability component V however is more difficult to derive. The two possible approaches that can be used to estimate V are either a forward calculation using process times and planned or historical inter-arrival times, or a backward calculation using historical queue time performance representative of specific factory states. Li et al (2005) suggest using the backward calculation and effectively apply it to generate operation curves for an assembly and test facility. Since this approach requires that sufficient history is available to provide data representative of the desired or anticipated factory state, it is not viable for green field factory planning or for factories that are only in the start-up phase. However for more mature factories it provides a practical and effective way to derive factory variability.

4.2 Backward Calculation of V

As suggested by Li et al (2005) the VUT equation (2) can be solved for V as follows:

$$V = \frac{\mathrm{CT}_q}{Ut_e}$$

Combined with equation (1) the resulting equation for the backward calculated variability component becomes:

$$V_B = \frac{\mathrm{CT}_q}{\left(\frac{u^{\sqrt{2(m+1)}-1}}{m(1-u)} \right) t_e} \tag{3}$$

We can now calculate the factory variability V since the variables in equation (3) are all known as long as historical data is used for CT_q. To make this value useful the following consideration need to be observed for CT_q and the other variables:

- CT_q is the historical average queue time for each equipment set during the time period chosen that is representative of either the desired or anticipated factory state.
- u is the utilization for each equipment set calculated using the following:
  - demand OEE modeled using the known historical factory loading mix and volume.
  - actual availability and efficiency for each equipment set during the factory state that the variability is being calculated for.
- m is the number of machines in each equipment set that was active during the time period that is representative of either the desired or anticipated factory state.
- t_e is the mean effective process time for each equipment set during the time period that is representative of either the desired or anticipated factory state. The t_e values typically do not vary much over time unless the factory experiences significant changes in loading mix.

4.3 Resulting Equation for Calculating CT_q

The equation for time waiting in queue using the backward calculated variability component V_B can now be written as:

$$\mathrm{CT}_q = V_B \left(\frac{u^{\sqrt{2(m+1)}-1}}{m(1-u)} \right) t_e$$

where:

- V_B is the variability for each equipment set backward calculated for the time period that was chosen to be representative of the factory state to be modeled
- u is the modeled utilization for each equipment set calculated using the following:

  - demand OEE as modeled using the planned/expected loading mix for the scenario that cycle time is being calculated for.
  - either planned availability and efficiency for each toolset or actual availability and efficiency representative of the factory state that cycle time is being calculated for.
- m is the planned/expected number of tools in each toolset for the scenario being modeled.
- t_e is the mean effective process time for each equipment set during the time period that is representative of either the desired or anticipated factory state.

5 IMPLEMENTATION AND CHALLENGES

The implementation process follows the steps shown in Figure 1.

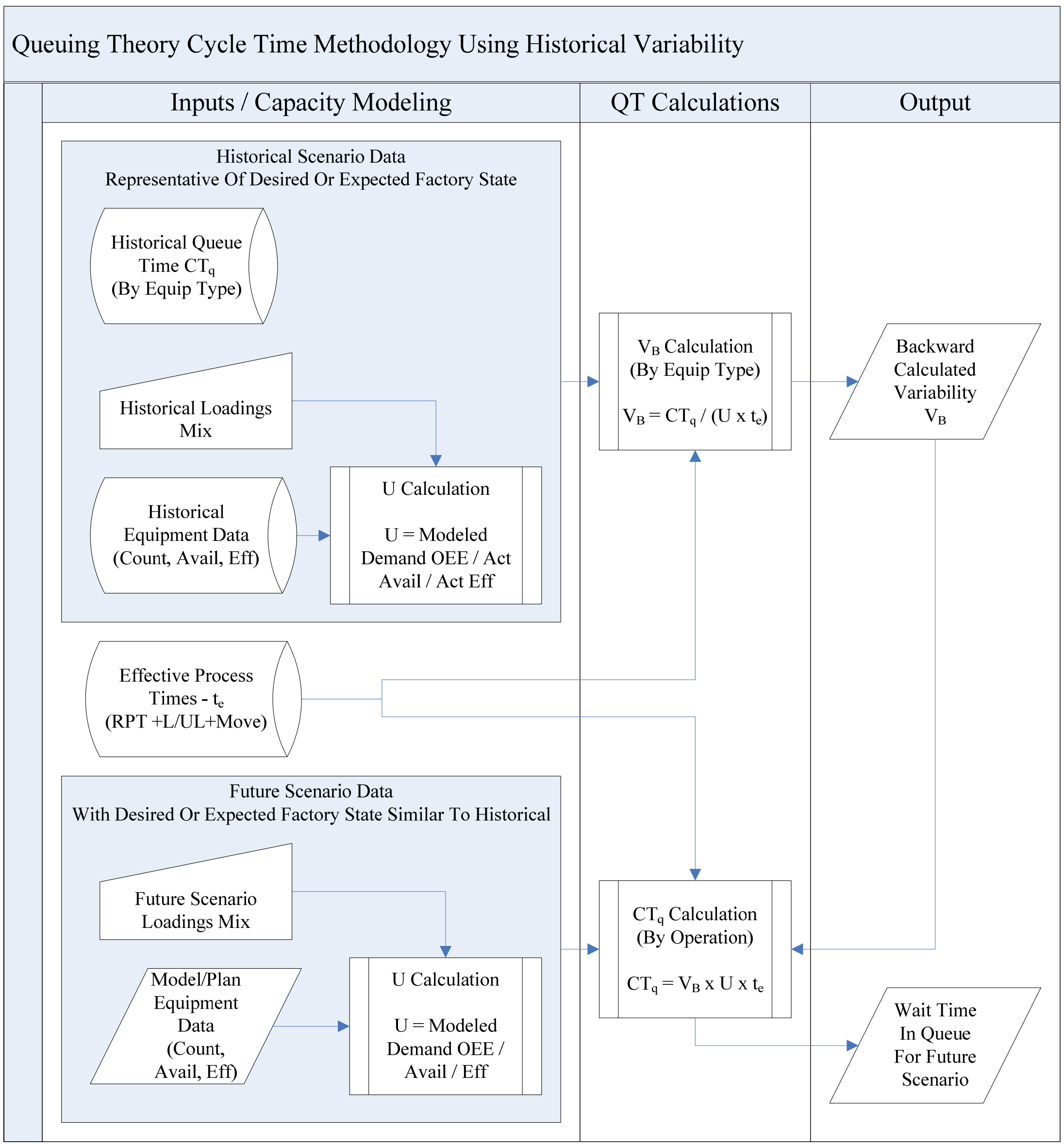

Figure 1: Queuing theory cycle time implementation process flow

First a time period is chosen that is representative of the desired or anticipated factory state to be modeled. Then historical loading mix and equipment data for that time period are extracted and fed into the model to calculate *U*. Historical queue time data are extracted for the same historical time period and together with effective process times used to calculate the variability V_B for each equipment set. These variability values can now be used to calculate the wait times in queue and subsequently estimate overall cycle times for future scenarios assuming a similar state of the factory.

Key challenges associated with the implementation process include the following:

The most significant element for success to make this approach work at all is access to extensive historical data. Fortunately, many facilities have information and automation systems that can provide the relevant data to allow the necessary analyses to occur. In these factories historical queue times, historical equipment data, as well as factory wide effective process time components are readily available for any desired time period.

The next challenge is correlating historical time periods with specific factory performance states. A knowledge of historical factory performance is paramount to choose the right time period in order to ensure that the variability calculated is as close as possible to the factory state being modeled.

An underlying requirement is the accuracy of the capacity model which generates the utilization for all of the individual equipment types. Many capacity models in the industry have been built without consideration for cycle time calculations and frequently do not include all or some of the metrology and electrical test steps. This was indeed the case when the cycle time modeling methodology presented here was added to an already existing factory capacity model. Thus far the current methodology estimates the impact of metrology and electrical test on cycle time as a percentage of the overall cycle time based on historical data. While this has been found to be reasonably accurate it is one of several opportunities for future refinement as discussed later.

6 VALIDATION PROCESS AND RESULTS

Figure 2 below outlines the validation process used.

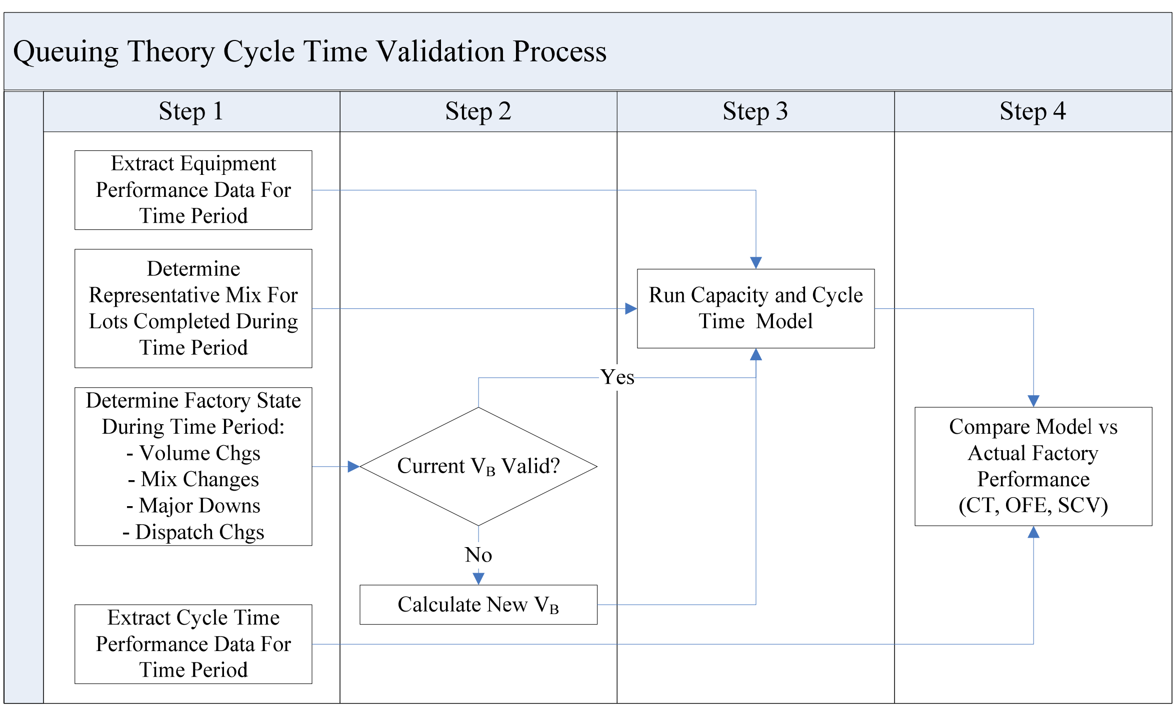

Figure 2: Queuing theory cycle time validation process flow

The first step in the validation process matches the implementation flow and consists of extensive data extraction. The second step requires the evaluation and possible recalculation of the backward calculated variability for each equipment type. This should not be required unless the state of the factory has been affected by significant changes in volume, mix, or dispatch rules, or major equipment availability issues. Once a variability assumed to be sufficiently representative of the time period for which the cycle time is being validated has been calculated, the model can be run to determine cycle times that can be compared against actual factory performance.

The validation chart below shows the results for four major technology flows running through a factory with a complex mix of over fifteen different technologies. The technology flows selected are representative of the process families that drive the majority of the factory volume and are therefore used as marker flows to model cycle time. In addition to the cycle times by technology a volume weighted average of these cycle times is calculated to arrive at an overall factory cycle time which is not shown below but correlates similarly with actual factory performance.

As is evident from Figure 3 below all four technology cycle times have been validating well for most time periods.

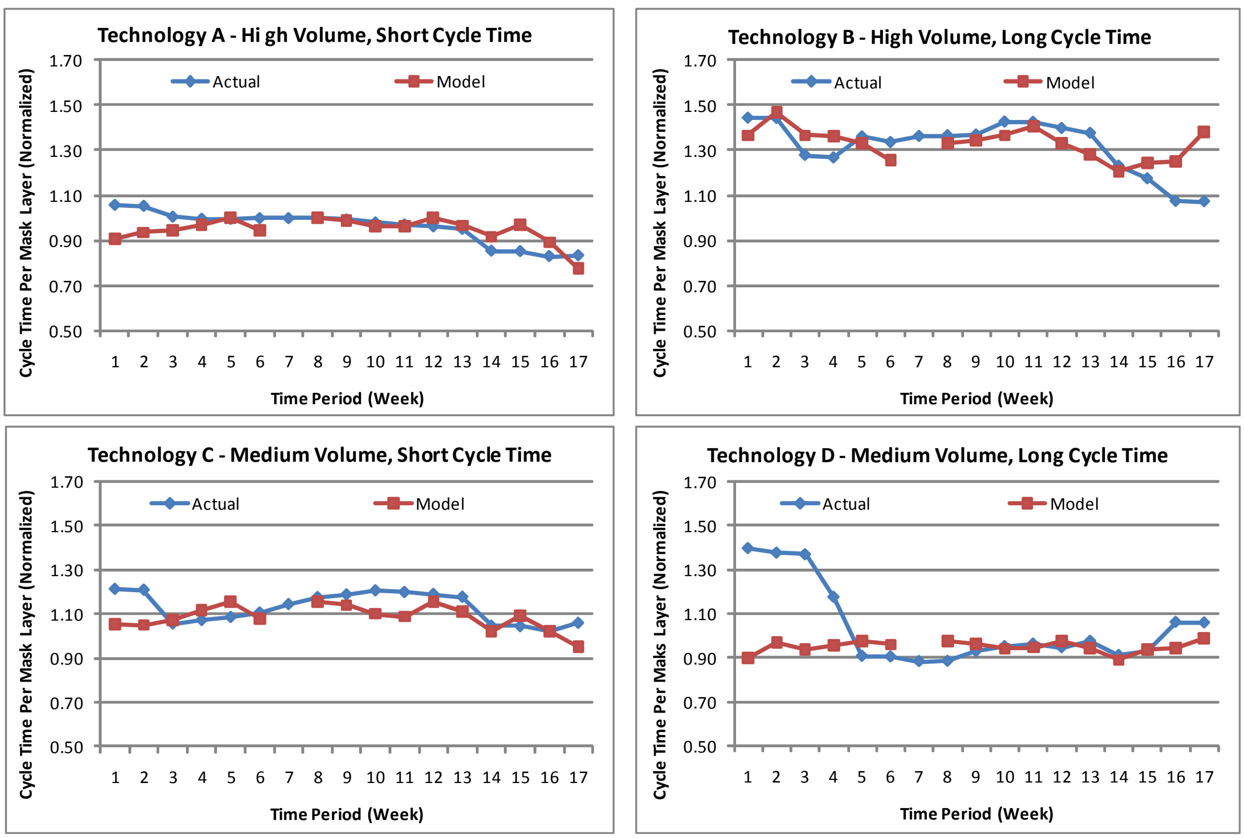

Figure 3: Queuing theory cycle time validation results

The above charts by technology show three instances where the modeled cycle time did not validate well against actual factory cycle time performance. During those time periods the factory was temporarily underutilized. This produced highly irregular fab behavior during and right after those time periods which violated steady-state model assumptions.

It is interesting to note that Technology B shows a divergence between modeled and actual cycle time over the last three time periods. This is a result of deliberate actions taken by the manufacturing floor to manage around constraints unique to this particular technology thereby again violating model assumptions.

Overall validation results were highly encouraging especially given the additional opportunities for further refinement of the methodology employed.

7 APPLICATIONS

Using the above outlined methodology the calculated cycle times have been used to aid in ongoing factory performance monitoring and analysis as well as future factory performance estimation.

7.1 Tactical Factory Performance Monitoring

The model has been used to evaluate tactical loading mix changes both for volume increases as well as volume decreases and their respective impacts on cycle times. The model correctly predicted technology specific cycle time fluctuations resulting from specific equipment set utilization changes due to shifts in loading mix.

Similarly the model was used to identify what tools were no longer needed when loading temporarily declined. While the capacity portion of the model on its own predicted certain tool count reductions based purely on utilization, the cycle time portion was used to identify the resulting cycle time impacts. Based on this additional analytical capability it was possible to adjust tool counts to ensure the factory would continue to meet the desired cycle time targets.

7.2 Strategic Factory Performance Estimation

By using a variability that reflects optimum factory performance observed in the past it is possible to estimate factory performance for projected strategic factory loading or intended mix changes. The model has been used to gain a relative understanding of the factory performance for both short term and long term loading volume and mix scenarios. While the accuracy, as is the case with all models, is not expected to be very high and declines with the degree and number of differences between the modeled scenario and loading producing the known historical factory states, the model cycle time results can serve as a good indication of what kind of factory performance can be expected based upon the variability, equipment count, and other model assumptions used.

8 OPPORTUNITIES

While initial validation results have been very encouraging, we acknowledge that the modeling method as presented still has several opportunities for improvement.

The key aspects of our current approach that are recognized to require additional work can be divided into three areas: Factory Definition, Forward Calculated Variability, and Enhanced Variability.

The current model was originally designed to only address capacity calculations and as such was built around the premise that a small but representative number of process flows would be sufficient to properly model the factory. An increasing portfolio of process technologies has led to an attending proliferation of products using multiple process variants. This has produced a situation where modeling only representative process flows can fail to capture equipment demands due to not modeling certain process flow variants. The missed demands may be minor or major depending on loading mix. As mix changes are highly unpredictable it has become necessary to increase the number of process flows being modeled to include as many flows as possible and ideally all those present in the factory. An initial evaluation of the feasibility to increase the number of process flows being modeled from dozens to hundreds has been performed resulting in tens of thousands of modeled process flow operations requiring increased computing power. The model also did not include metrology or electrical test operations. This further exacerbates the problem of needing to model an increased number of process steps. Both metrology and electrical test have recently been included in the model and are currently undergoing validation.

As outlined in this paper the model currently derives the variability component of the Kingman equation by backward calculating the variability component for each equipment type from historical queue time data. As pointed out this is valid only as long as the key factors affecting variability in the factory are assumed to remain unchanged. As such there is a danger that major swings in loading and/or mix will render the results less accurate. A forward calculation of variability using process variability as well as other variability subcomponents should help mitigate this danger; however, the need to estimate the inter-arrival time variability remains a challenge that will likely always be solved using historical data. Work is under way to test refinements to the G/G/m model such as those proposed by Whitt (1993). Initial results show that further improvement can be gained by using Erlang distributions with 2 phases for process and inter-arrival times.

The third area requiring additional work that should further improve robustness of this approach is a refinement of the elements within the variability component equation itself. We concur here with Shantikumar, Ding, and Zhang (2007) that the accuracy of analytical models can be further improved by incorporating key characteristics of wafer fabrication facilities. Future efforts are intended to focus on testing and developing equations that address unique equipment behavior for the following major equipment types: single wafer serial processing, batch serial processing, single wafer parallel/re-entrant processing, batch parallel/re-entrant processing. Each one of these equipment types will likely require additional unique variability factors to properly represent their respective operational behavior. We expect to draw on previous work done in this area by Akhavan-Tabatabaei and Ding (2009) and Morrison and Martin (2007).

9 CONCLUSIONS

While the methodology presented in this paper to use historical data to calculate factory variability and then use that variability to predict cycle time clearly has its limitations, ongoing validation and successful applications are proving that model results correlate well with actual performance in a factory running a complex mix of multiple process technology flows. Key to a successful implementation are extensive historical data access and the ability to properly correlate factory performances states with historical time frames. Additionally the underlying capacity model that predicts equipment utilization has to be accurate and sufficiently detailed to allow for cycle time calculations at all requisite operational steps.

The model in its current state has been used to predict cycle time for both tactical as well as strategic scenarios and has been found to predict factory behavior sufficiently well to aid significantly in operational and capital planning decision making.

ACKNOWLDEGEMENTS

The author expresses appreciation to the following individuals for their insight and assistance during the development of this methodology, its implementation, validation, application, and refinement:

Abolfazl Taghizadeh, National Semiconductor, Santa Clara

Paul Fearon, National Semiconductor, South Portland

Scott Hopkins, National Semiconductor, South Portland

REFERENCES

Akhavan-Tabatabaei, R., and S. Ding. 2009. "A Method For Cycle Time Estimation Of Semiconductor Manufacturing Toolsets With Correlations." In *Proceedings of the 2009 Winter Simulation Conference*, edited by M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin and R. G. Ingalls, 1719-1729. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

FabTime. 2011. "Cycle Time Estimation Formulas." Accessed March 18, 2011. http://www.fabtime.com/formula.shtml.

Chung, S. H., and H. W. Huang. 2002. "Cycle Time Estimation for Wafer Fab with Engineering Lots." *IIE Transactions* 34: 105-118.

Hopp, W. J., and M. L. Spearman. 2001. *Factory Physics: Foundations Of Manufacturing Management*. 2nd ed. New York: McGraw-Hill, Inc.

Li, N., L. Zhang, M. Zhang, and L. Zheng. 2005. "Applied Factory Physics Study on Semiconductor Assembly and Test Manufacturing." *IEEE International Symposium on Semiconductor Manufacturing,* 307-310.

Morrison, J. R., and D. Martin. 2007. "Practical Extensions to Cycle Time Approximations for the G/G/m-Queue With Applications." *IEEE Transactions On Automation Science and Engineering* 4(4):523-532.

Pearn, W. L., Y. Tai, and J. H. Lee. 2009. "Statistical Approach for Cycle Time Estimation in Semiconductor Packaging Factories." *IEEE Transactions On Electronics Packaging Manufacturing* 32(3):198-205.

Shanthikumar, J. G., S. Ding, and M. T. Zhang. 2007. "Queuing Theory for Semiconductor Manufacturing Systems: A Survey and Open Problems." *IEEE Transactions On Automation Science and Engineering* 4(4):513-522.

Whitt, W. 1993. "Approximations For The GI/G/m Queue." *Production and Operations Management* 2(2):114-161.

AUTHOR BIOGRAPHY

ROLAND SCHELASIN is the site modeling and simulation engineer at National Semiconductor Corporation in South Portland, Maine (National Semiconductor Corporation was acquired by Texas Instruments Incorporated in September 2011). He started working in the semiconductor industry in 1988 at IBM in Essex Junction, Vermont. Collaborating with IBM Research in Yorktown Heights, New York, he pioneered the development and use of simulation models for complex equipment within the semiconductor industry and has authored or co-authored several papers on this subject. In 1996 he joined National Semiconductor Corporation (NSC) to manage the modeling and simulation efforts for National's first 200MM wafer fabrication facility. As part of the Industrial Engineering organization at NSC he continues to provide modeling and simulation expertise to develop decision making tools that are used to manage and improve equipment and overall factory effectiveness. Email: Roland.Schelasin@ti.com.