Proceedings of the 2008 Winter Simulation Conference
S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler eds.

USING LITTLE'S LAW TO ESTIMATE CYCLE TIME AND COST

Kristin Rust

Advanced Micro Devices
Center for Operations Research
5204 East Ben White Blvd MS 563
Austin, TX 78741, USA

ABSTRACT

Well designed models can provide timely answers and summarize understanding of complex systems. Simple use cases can illustrate the applicability and inferences possible with even the most general models. This paper uses Little's Law to demonstrate the prediction of fab operating conditions and associated cost implications.

1 INTRODUCTION

"Essentially, all models are wrong, but some are useful" (Box 1987.) This quote from famed industrial statistician George Box highlights the risks and rewards of modeling. Models are used by practitioners to quantify the effects of decision making in complicated systems. However, if not carefully designed or understood, models and simulations often draw criticism for their abstraction of reality. This can be particularly true when practitioners seek complete autonomy in building a model; the design and interpretation of the model is just as important.

The first step in model building is understanding the purpose of the model and its usage. The alternative is an increased risk of failure in practice either because the model use is ambiguous, and does not fit well with the use cases, or is not robust enough to evolve with changes in the modeled system. Equally important, analysts of the results must understand the assumptions present in the data and abstractions made during model construction. Failure to understand how model abstractions reflect the applicability of model results can dampen the validity of conclusions drawn from the model. Therefore, careful validation of a model (input data, constructs, and reports) is necessary to ensure model results can withstand the scrutiny given during an analysis of conclusions. Properly designed models account for data exceptions and document data understanding as assumptions in the input and output of a model, increasing the validity of the model and its interpretation. In cases in which emulation is used, the model is designed to have complete comprehension of system details and exception cases. Generally, experimental use of emulation is more difficult due to the detailed exception handling required during development; it is also less robust in its use, and use incurs significantly greater costs in terms of time and money.

So, although a model is an abstraction of reality, its usefulness in decision making should not be marginalized. Carefully constructed models and simulation experiments can provide quantified estimates of data and summarize large amounts of data in ways that would otherwise be impossible to understand. This paper seeks to demonstrate the usefulness and pitfalls of utilizing general models through described use cases. Additionally, the paper discusses a methodology for constructing the cost impact associated with complex systems as fabs.

2 USING LITTLE'S LAW TO PREDICT AND UNDERSTAND FAB METRICS

Industrial engineers use Little's Law to predict system behavior and validate operational improvements in a variety of factory settings. Measuring current performance metrics against values predicted by Little's Law also provides a measure of the stability in the current system. This measure can quantify the maximum achievable operational improvements. In addition, Little's Law can be used to build intuition about expected factory performance when implementing change given the current factory state. The law states that, for a system in steady state, the long-term average throughput equals the ratio of the total average work in process (WIP) over the average cycle time (CT) (Little 1961.) The assumption of steady state is made for large systems when analyzing factory-wide behavior and localized system performance. The law can be proven mathematically and is often used in a factory setting to intuitively predict changes in CT and WIP. Although it is used widely in practice, interpretation of the applicability of Little's Law in a fab is disputable; often-asked questions include "How good is the assumption of steady-state behavior?" and "How close to reality are the long-term results?"

Arguably a fab is not in steady state. Shifting bottlenecks and product mix and prioritization changes due to uncertainty in tool availability and demand, in addition to the availability of material transport, create a dynamic environment. This also impacts the capacity of the factory. Fab engineers constantly struggle to manage these changes with local system improvements and employ company-wide initiatives to reduce variability (such as Lean or Six Sigma.) The objective is improved tool utilization or improved cycle times. Often, while these targets provide benefit to many cost measures, other company goals and metrics can hinder WIP controls. Additional WIP provides a quantified measure of productivity by utilizing the additional capacity gains in the factory and generating more impressive tool utilization figures, daily WIP turns, and higher reporting of activities. Since the results of such policies in practice is higher WIP and CT, or both, the expected lower WIP or improved CT predicted when applying Little's Law do not materialize. So, instead of seeing a declining WIP level due to improvements in CT, the WIP level often will increase or remain constant. This is possible if throughput is increased as a result of the CT improvements.

Therefore, two key assumptions must be made when interpreting and applying the results of Little's Law: first, the throughput associated with CT and WIP used in the calculation is constant for that specific use case; and, second, the results indicate the long-term, steady-state behavior under the conditions reported. Understanding these two assumptions is key to applying the results of Little's Law for business decisions and identifying exception cases that may distort the results in practice. So, although a fab may not operate in steady state over the long term, using Little's Law to predict the influence that changes in CT, throughput, or WIP may have on each other is valid as long as the results are interpreted for specific use.

The applicability of Little's Law is often accepted without first validating these assumptions. The following case studies provide a mechanism for categorizing states of factory performance such that predictions can be made for transition states of the factory. The relationship between parameters defining these states adheres to the constructs of Little's Law. The paper then demonstrates that the formulation holds under mathematical proof.

3 CASE STUDIES

It is proposed that the changing rates of cycle time (CT), work in process (WIP), and throughput characterize the different states of a factory. If this is true, the relationship between them as defined by Little's Law allows factory states to be categorized into four distinct cases:

1. Throughput decreases and WIP increases
2. Throughput increases and WIP increases
3. Throughput increases and WIP decreases
4. Throughput decreases and WIP decreases

In practice, the controls available that impact CT, WIP, and throughput are changes in capacity, changes in operational policies, and changes in the start rate. For each of the following cases, changes in WIP and CT are a result of these controls. Changes in throughput are made by adjusting the start rate of the factory.

Table 1: Definition of terms

$w\ (t_x)$	WIP level for period *x*
$c_x(t_x)$	CT value for given WIP level in period *x*
t_x	System throughput for given WIP level in period *x*

3.1 Case (1): Throughput decreases and WIP increases

Although in practice management does not need mathematical proof that implementation of suboptimal operation policies leads to higher cycle time, Case 1 is a demonstration. The poor operation policies would result either from loss of capacity at bottleneck tools or from implementing a change that creates a shift of bottleneck to a tool with lower throughput than the current bottleneck. Examples are suboptimal recipe/setup management policies, worse wafer cascading rules, or poor batch rules.

$$
\begin{aligned}
&\forall t_1, t_2 \\
&\mathit{if} \\
&t_1 \geq t_2, \\
&w\ (t_1) < w\ (t_2) \\
&\rightarrow w(t_1)/t_1 < w(t_2)/t_2 \\
&\mathit{Little's}\quad \mathit{Law}: \\
&c(t_1) = w(t_1)/t_1 \\
&c(t_2) = w(t_2)/t_2 \\
&\therefore \\
&c(t_1) < c(t_2)
\end{aligned}
$$

3.2 Case (2): Throughput increases and WIP increases

Improving the capacity limits or throughput at bottleneck tools provides the opportunity for an increased start rate. WIP may accumulate because of additional capacity. Throughput may increase because of operation improvements or tool improvements. Two sub-scenarios exist to define the resulting CT performance change; these scenarios (2A and 2B) are defined by the ratio of the rate of change of WIP and throughput.

$$
\begin{aligned}
&\forall t_1, t_2 \\
&if \\
&t_1 \le t_2, \\
&w\ (t_1) < w\ (t_2) \\
&\rightarrow w(t_1) = \omega\, w(t_2) \quad where \quad \omega > 1 \\
&\quad t_1 = \tau\, t_2 \quad where \quad \tau \ge 1 \\
&Little's \quad Law: \\
&c(t_1) = w(t_1)/t_1 = \omega\, w(t_2)/\tau\, t_2 \\
&c(t_2) = w(t_2)/t_2 \\
&\therefore \\
&A: \quad \omega \ge \tau \quad \Leftrightarrow \quad c(t_1) \ge c(t_2) \\
&B: \quad \omega < \tau \quad \Leftrightarrow \quad c(t_1) < c(t_2)
\end{aligned}
$$

3.2.1 Scenario 2A

If the system throughput increases at a faster rate than WIP, cycle time decreases. This case illustrates examples when material is held in the line, processing when capacity has been added, or at non-bottleneck operations. Although it may seem counterintuitive to some that CT could decrease after an increase in WIP, it is important to note that the WIP is added in areas in which it does not introduce additional queue time. This scenario demonstrates the behavior of a system when adding bottleneck capacity or improving operating policies to increase the operating throughput of bottleneck operations. With this increased capacity, queue levels improve and, while there is queuing of the additional starts, the overall queue time and/or non-queue time improves. Despite the CT benefits, this scenario indicates increased costs associated with the changed system. Additional storage and material handling costs and increased capital equipment cost at the bottleneck operations is likely. The new system described could also exhibit additional queueing at non-bottleneck operations due to the increase in starts. Systems would be required to manage the additional WIP and additional controls put in place to limit increases in variability in material handling/leveling across the factory due to the higher WIP. Apart from the changes in factory systems associated with scenario 2A, it can be argued that better accuracy of product demand would be needed to manage the risk of over/under production of a particular product. This risk of starting too many products of a particular type is countered by shrinking the time window between product start and customer delivery. It therefore becomes unclear what the cumulative risk/benefits are from a CT reduction perspective. A more optimal operating scenario may be to increase throughput at a rate consistent with the rate increase of system WIP.

3.2.2 Scenario 2B

If throughput increases faster than WIP, Little's Law predicts what fab managers already know: pushing additional starts into a production line already at capacity does not provide benefits. Not only does WIP increase, CT performance deteriorates. Clearly this case illustrates the result of suboptimal operating policies and the transition of a factory out of steady-state operations. Operating practices are in place to guard against this effect. Examples include determining in advance the threshold capacity of the factory's bottleneck tools for capacity planning, implementing queue limits as part of any push dispatching policies, and balancing factory measures that solely stress tool utilization metrics with on-time delivery or queue time metrics.

3.3 Case (3): Throughput increases and WIP decreases

If throughput increases and WIP decreases, Little's Law predicts that CT will decrease. This indicates excess capacity was available at bottleneck tools or improved operating policies were implemented at bottleneck tools to generate additional working capacity. This additional capacity provides the opportunity to increase the start rate. In addition, WIP controls are put in place to maintain or lower the queue levels and non-productive time of lots. Such examples are next-generation scheduling systems that reduce non-productive lot time, improvements in automated material handling systems to reduce variability and transport times, improvements to tool operating parameters (cascade lengths, chamber scheduling, setup management, preventative maintenance policies), and improvement in tool reliability. At AMD, these types of system improvements have historically had the largest focus when seeking productivity improvements, and the company has had many successes improving the productivity of tools, automation systems, and dispatching rules. Most recently, Lean concepts are providing measured success in WIP management and variability reduction when applied to operating policies in dispatching and production planning.

$$\begin{aligned}
&\forall t_1, t_2 \\
&if \\
&t_1 < t_2, \\
&w\ (t_1) \geq w\ (t_2) \\
&\rightarrow w(t_1)/t_1 > w(t_2)/t_2 \\
&Little's \quad Law: \\
&c(t_1) = w(t_1)/t_1 \\
&c(t_2) = w(t_2)/t_2 \\
&\therefore \\
&c(t_1) > c(t_2)
\end{aligned}$$

3.4 Case (4): Throughput decreases and WIP decreases

Depending on the rate of change, reduced start rate and lower WIP can lead to lower CT or higher CT. CT is lower if WIP decreases at a faster rate than throughput. Conversely, when throughput decreases faster than WIP, CT is higher. Lowering the throughput and WIP can occur if the number of bottleneck tools is not optimal (too few) for minimizing CT. This creates the opportunity to improve queue times at bottleneck tools and reduce WIP in the system by reducing starts. This scenario is valid when discussing factory systems that may be ramping to full production or in transition state; its use in practice is thus limited. As described in Cases 2A and 2B, the increase or decrease in CT depends on the rate of WIP change as compared to the throughput change. WIP must decrease at a faster rate than throughput to realize any CT benefits.

$$\begin{aligned}
&\forall t_1, t_2 \\
&if \\
&t_1 > t_2, \\
&w\ (t_1) > w\ (t_2) \\
&\rightarrow w(t_1) = \omega\, w(t_2) \quad where \quad \omega < 1 \\
&\quad t_1 = \tau\, t_2 \quad where \quad \tau \leq 1 \\
&Little's \quad Law: \\
&c(t_1) = w(t_1)/t_1 = \omega\, w(t_2)/\tau\, t_2 \\
&c(t_2) = w(t_2)/t_2 \\
&\therefore \\
&A: \quad \omega \geq \tau \quad \Leftrightarrow \quad c(t_1) \geq c(t_2) \\
&B: \quad \omega < \tau \quad \Leftrightarrow \quad c(t_1) < c(t_2)
\end{aligned}$$

3.4.1 Scenario 4A

Conditions in which CT may be improved by lowering the WIP and throughput exist in factories with great variability in the WIP distribution across stations, and in factories not constrained to produce at a pre-defined throughput rate. Operating policies that would redistribute WIP to better balance station queues or policies to batch differently are examples. A more difficult gauge is that of altering the production start rate; it is often difficult to determine if improvements to CT result from lowering the start rate. In addition, if the incremental benefit in CT is small, it is not supported by ROI analysis.

3.4.2 Scenario 4B

The change in CT resulting from decreasing throughput at a faster rate than WIP is not often considered. Increased CT in this case occurs as factories do not operate in practice. Therefore, discussion of this case is not practically relevant.

3.5 Summary of Cases

Each case study presented uses the relationship between throughput, CT, and WIP to identify distinct factory states and uses these states to categorize management decisions and factory behavior. Little’s Law provides a mechanism for quantifying the factory performance that intuition predicts. In enumerating the different states possible using Little’s Law in which throughput is the changing factor, each defined state could be proven mathematically. In addition, using the case studies to define use cases associated with these states demonstrates how, in practice, values may be quantified using this method. Figure 1 shows a graphical representation of the case studies.

The cases presented are characterized graphically by differentiating the WIP movements along the associated throughput curve. The impact to cycle time is also displayed as a function of WIP. Changes in throughput may be made by moving up or down the throughput curve. Movements to the left or right of the inflection point of the throughput curve separate the sub-cases (A and B) for Cases 2 and 4. Cases 1 and 3 occur as upward and downward movements on the throughput curve after the stationary point. The range of CT values defined by WIP levels where the throughput curve is at its stationary point defines the range in which a factory typically operates. Values outside of this range may be thought of as states where atypical operation decisions exist for the factory. The areas for CT improvement that are investigated are described by Cases 4A and 3. It is in these cases -- reducing variability in WIP levels across the fab, optimally determining the factory start rate, and improving tool and lot scheduling and predictability -- that operational improvements define improvements in CT.

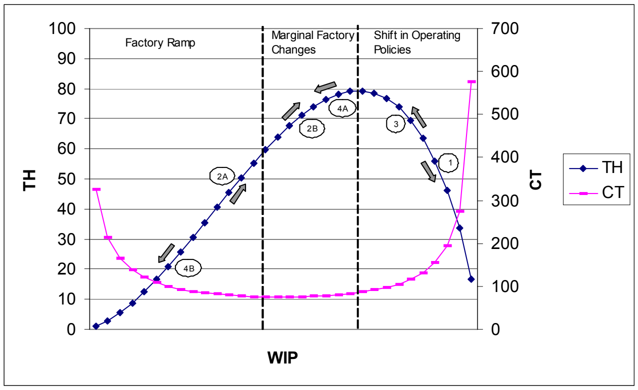

Figure 1: Operating curves for throughput (TH) and cycle time (CT).

4 ASSOCIATING COST WITH CYCLE TIME

A question often asked in practice is, “What is the cost associated with a change in cycle time?” This is a difficult question to answer. Although there is undisputedly an intrinsic value to cycle time in terms of customer responsiveness and manufacturing simplicity (WebProNews 2006), associating a dollar value to cycle time often trivializes the true impact of manufacturing time on total cost to the company. Not only is there the capital cost of equipment, material handling, and storage required to manufacture products with large lead times, but there is also an opportunity cost associated with CT in terms of lost sales and higher finished goods inventory. Specifically, orders placed for products with lower CT have a greater probability of being filled before the orders expire or change. This may be estimated by deriving a sensitivity curve representing the relationship between the probabilities of order fulfillment given the amount of time since an order was placed. The greater the required order lead time, the less the probability that the order will be confirmed and the greater the probability of moving the product to finished goods inventory. However, this cost is rarely considered when evaluating CT impacts to cost.

What typically is considered is cost of equipment, factory infrastructure installation and construction, consumable goods used as part of the manufacturing process given assumed technology nodes, and wafer size (Semiconductor International 2005). Additional costs are estimated for storage of finished goods inventory and WIP. While these cost are the largest components in defining the product cost, using these measures when evaluating the cost of differing scenarios can be misleading. The cost of improved CT is not solely a function of the capital and variable cost of the factory. This is especially true when determining the relative change in CT across alternative factory scenarios. Other costs directly related to the WIP level in a factory are increased cost of scrap, variability in material movement from additional traffic congestion, and increased development cost from longer CT of non-product material produced in the factory. Increased CT results in higher cost due to scrap in WIP and finished goods inventory. Higher scrap costs result not only from a short product obsolescence lifecycle but also from WIP loss because of longer time to defect detection. In addition to costs such as increased number of loadports and local material storage solutions, longer CT and increased automated material handling system (AMHS) delivery times increase factory costs by adversely affecting product yield. Yield loss can be attributed to additional handling and process controls sensitive to time.

These costs can be quantified using the relationship between the length of time in production and the overhead cost of the factory and/or capital cost of equipment supporting higher WIP levels. Additional, more subjective measures are typically more difficult to estimate. For example, estimating revenue loss as a function of the time to product introduction may depend on an estimate of market responsiveness and competitive behavior. Methods for obtaining estimates of these factors are not well defined and may differ widely depending on market characteristics.

Another example of cost measures that may be difficult to estimate are those associated with quantifying cost of additional traffic and increased risk of scrapped material. Both measures are a result of increased variability and CT length and can be minimized by well designed factory systems. The risk of increased scrap material can be managed by good control systems in the factory and traffic can be managed by carefully scheduling material movements or designing an AMHS with low variability in material movement times. However, assessing the cost of more sophisticated systems cannot be tied to a single factor of these measurements, nor are they directly given in terms of CT. Cost measures for additional traffic would include the cost of additional vehicle hardware and support systems to achieve a target value for AMHS statistics. However, the cost of traffic reduction should also include the opportunity cost associated with idle time due to late delivery for which there is no defined cost measure. Additionally, scrap cost would include the material costs and estimated labor for production of the scrapped WIP; however, no direct translation to CT is made without further measures of how much time is saved from not having started the scrapped material.

5 CONCLUSION

While all modeling is subject to interpretation, the benefit of well designed and executed experiments include better defined components of factory CT and cost, predicted areas for process improvements, and estimates that can quantify benefits or drawbacks of differing operating alternatives. The first step for successful modeling is understanding the model use and experiment objectives.

Clearly stating assumptions associated with the model data, model development, and output statistics is essential to interpreting model results. The assumptions associated with Little's Law hold in a variety of practical cases.

As shown in Section 3, the characterization of discrete factory states provides insight into the types of operational improvements that can provide cycle time benefits. Furthermore, the use of Little's Law provides a mechanism to prove the validity of these state representations and the predicted changes in cycle time. It also provides a structure to categorize intuitive factory states recognizable in practice.

The importance of changes in cycle time relates directly and indirectly to cost. Section 4 provides insight into methods for quantifying cost and how it can be associated with change in cycle time. It also demonstrates the importance of understanding the method and models used when given cost values. As with any model, the key to results analysis and use is understanding the assumptions present in the input data, model mechanics, and output data.

ACKNOWLEDGEMENT

The author thanks Detlef Pabst for his assistance in the formulation of the concepts and proofs presented in this paper.

REFERENCES

Box, G., and N. Draper. 1987. *Empirical model-building and response surfaces*. New York: John Wiley.

Little, J. 1961. A proof of the queueing formula L=λW. *Operations Research* 9:383-387.

Semiconductor International. 2005. A simulation study of the cost and economics of 450 mm Wafers. Available via <http://www.semiconductor.net/article/CA632496.html> [accessed March, 2008].

WebProNews. 2006. Creating supply chain value with cycle time and inventory yield. Available via <http://www.webpronews.com/expertarticles/2006/09/06/creating-supply-chain-value-w-cycle-time-inventory-yield> [accessed March, 2008].

AUTHOR BIOGRAPHY

KRISTIN RUST is a senior industrial engineer in the Center for Operations Research at Advanced Micro Devices. She has enjoyed a wide-ranging career in project management and industrial engineering across multiple industries including aerospace, transportation, general manufacturing, and semiconductor. She received an undergraduate degree in finance and engineering from the University of Texas at Austin, where she went on to complete her Master of Mechanical Engineering in Operations Research. She began working at AMD in 2004. While there, she has pioneered the development of next-generation fab modeling techniques and proving new manufacturing concepts. She has additional expertise in systems deployment, simulation design and development, and statistical analysis.