Proceedings of the 2010 Winter Simulation Conference
B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, eds.

APPLICATION OF ERLANG DISTRIBUTION IN CYCLE TIME ESTIMATION OF TOOLSETS WITH WIP-DEPENDENT ARRIVAL AND SERVICE IN A SINGLE PRODUCT-TYPE SINGLE FAILURE-TYPE ENVIRONMENT

Raha Akhavan-Tabatabaei
Juan José Ucrós

Universidad de los Andes
Cra. 1 No. 18A-10
Bogotá D.C., COLOMBIA

J George Shanthikumar

Purdue University
47906
West Lafayette, IN, USA

ABSTRACT

This paper proposes a methodology based on phase type distributions and a state dependent Markov chain model to estimate the cycle time of workstations (toolsets) in semiconductor manufacturing. Due to implicit operational policies adopted by the line managers the performance of the existing queueing models for toolsets is not satisfactory. On the other hand developing accurate simulation models for toolsets can be very time consuming and hard to maintain. In this paper we propose a Markov chain model with the ability to include implicit operational rules on dispatching and maintenance. We verify the performance of this model via simulation and present the results for a variety of arrival and service distribution shapes.

1 INTRODUCTION

Reducing cycle time (CT) and improving delivery performance has long been a key focus area for semiconductor manufacturers. Accurate cycle time estimation can play a major role in production planning and scheduling of semiconductor fabrication facilities (fab) which are among the most capital intensive industries. The key to efficiently design a fab is to have proper capacity predictions based on accurate cycle time estimations. Understanding the main factors that contribute to high cycle time will greatly help identify the most effective levers to reduce it. Due to the complicated nature of semiconductor manufacturing systems (SMS) it is not easy to accurately estimate the cycle time for fab toolsets. Most of the SMS toolsets are very complicated in design and their operations are closely governed by line managers. Scheduled and non-scheduled tool downtime add to the uncertainty of the service provided by these toolsets and the line managers constantly try to synchronize the operations to meet cycle time goals.

Shanthikumar *et al.* (Shanthikumar, Ding, and Zhang 2007) discuss the common approaches to cycle time estimation in SMS. They point out that one of the main reasons that the classical queueing models are inaccurate for SMS is the assumption of independent relationships. In a previous work Akhavan-Tabatabaei (Akhavan-Tabatabaei, Ding, and Shanthikumar 2009) perform a case study on several $G/G/m$ approximations including those that are specifically developed for SMS. They show that such formulas can estimate the cycle time with very large margins of error and hence can be misleading in many cases.

Classical queueing models assume that the arrival process and the service process are independent. However in many cases in SMS the line managers interfere with the random process of arrival and service. They make operational decisions and adopt certain policies to meet the cycle time goal or work in process (WIP) goals. These operational rules mainly target the adjustment of arrival

978-1-4244-9864-2/10/$26.00 ©2010 IEEE

process based on the WIP level or the adjustment of failure rates based on the toolset availability and through pulling-in or pushing out the maintenance activities. Therefore such operational rules create dependency between the arrival and service process through correlating them to the WIP level or toolset availability.

In this paper we propose a state-dependent Markov chain model that is capable of reflecting these operational rules in cycle time estimation. Unlike classical $G/G/m$ formulas that are blind to the implicit operational rules, this model has the flexibility to adjust the arrival rate or the failure rate based on the WIP level or toolset availability. We verify the results of this model by a simulation model that closely mimics a toolset with operational rules. Our comparison shows that the estimated cycle time by the Markov chain model is within the 95% confidence interval of the simulated toolset's mean cycle time.

The remaining of the paper is structured as follows. The detailed formulation of the state-dependent Markov chain model is discussed in Section II. Section III presents three case studies that verify the accuracy of the proposed model under different operational rules and various shapes of arrival and processing distributions. Section IV concludes the paper and discusses the next steps to extend the model for more general cases.

2 MODEL FORMULATION OF THE STATE-DEPENDENT MARKOV CHAIN

We consider a toolset with k^{max} parallel tools and buffer size of w^{max} lots. The inter-arrival time of lots to this toolset follows exponential distribution with mean $1/\lambda$ and their processing time also follow exponential distribution with mean $1/\mu$, the time to fail and time to repair for each tool follow exponential distributions as well, with means $1/d$ and $1/u$, respectively. The toolset also follows one or more operational rules. We propose a continuous-time Markov chain model for this toolset with the general state space of $\{[K(t),W(t)],t=0\}$, where $K(t)$ represents the number of available tools at time t, and $W(t)$ denotes the number of lots in process or in the queue at time t. Variables $K(t)$ and $W(t)$ take on values in the set of nonnegative integers $K(t)=0,\cdots,k^{max}$ and $W(t)=0,\cdots,w^{max}$. We also define $[k,w]$ as the state descriptor for this stochastic process at an arbitrary time, t_0, such that $K(t_0)=k$ and $W(t_0)=w$. At any time epoch a one-step transition from the current state $[k,w]$ is triggered by the occurrence of one of the following events:

- A new lot arrives to the toolset, which takes the model from state $[k,w]$ to $[k,w+1]$.
- A lot finishes its processing and departs the toolset. The transition is from $[k,w]$ to $[k,w-1]$.
- One of the tools fails and becomes unavailable for production. In this case the model transits from state $[k,w]$ to state $[k-1,w]$.
- One of the currently failed tools is repaired and made available for production. In this case the model transits to $[k+1,w]$ from $[k,w]$.

The presence of operational rules in SMS makes these transition rates dependent on k,w or both. For example if the operational rule dictates the adjustment of arrival rate based on the WIP level then when the toolset is in state $[k,w]$ the arrival rate is a function of w. Similarly if the operational rule calls for the adjustment of failure rate in low availability then the failure rate depends on k, hence making the transition rates dependent on the state variables. For such a state-dependent Markov chain model we denote the transition rates as Arrival rate: $\lambda^{k,w}$, Processing rate: $\mu^{k,w}$, Failure rate: $d^{k,w}$, Repair rate: $u^{k,w}$ Figure 1 shows a partial view of the state transition diagram for the Markov chain model.

Given the values of k,w, and the state-dependent transition rates for a toolset of interest, the steady-state probabilities can be calculated through solving the balance equations of the Markov chain, assuming that the inter-arrival, processing, time to fail and time to repair distributions follow the exponential distribution with corresponding rates of $1/\lambda$, $1/\mu$,$1/d$ and $1/u$. The steady-state

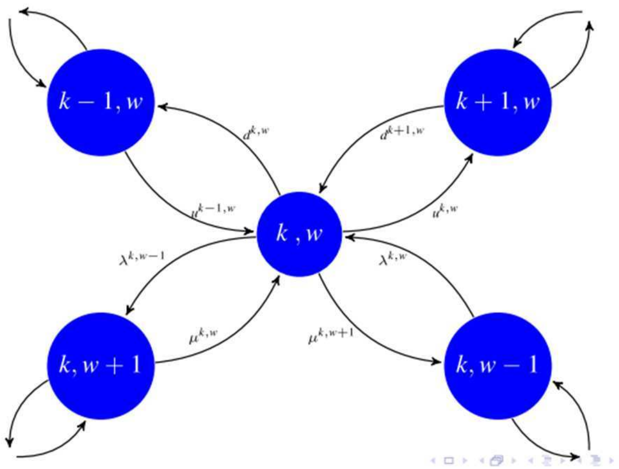

Figure 1: State transition diagram of Markov Model

probability for state $[k,w]$ is denoted by $\pi_{(k,w)}$ and the average number of lots at the toolset, WIP , can be found through

$$\overline{WIP} = \sum_{k,w} w \times \pi_{(k,w)}. \tag{1}$$

We define $\bar{\lambda}$ as the long-run average arrival rate of the lots to the toolset where

$$\bar{\lambda} = \sum_{k,w} \lambda^{k,w} \times \pi_{(k,w)}. \tag{2}$$

Applying Little's formula, finds the long-run average cycle time of the proposed state-dependent Markov chain as

$$\overline{CT} = \frac{\overline{WIP}}{\bar{\lambda}} = \frac{\sum_{k,w} w \times \pi_{(k,w)}}{\sum_{k,w} \lambda^{k,w} \times \pi_{k,w}}. \tag{3}$$

2.1 Restrictions of The Current Model

Since we use a Markov chain framework for this model one necessary assumption is that all the transition probabilities follow exponential distribution. This can be limiting since in reality of SMS toolsets there are many instances with non-exponential arrival or service distributions. To address this restriction in the next section we introduce a method to estimate a continuous distribution with squared coefficient of variation (SCV) less than 1 with an Erlang distribution. Erlang distribution is a sequence of exponential phases in series and hence it retains the memoryless property of exponential while being able to model various shapes of continuous distributions. The current model structure also requires two assumptions regarding the failure of the toolset. The first is that the tools do not fail when they are idle and the second is that when a busy machine fails the failure preempts the job in process and the job goes back to the queue and restarts its processing time, in the same machine, as if it is an unprocessed job. These assumptions keep the state-space smaller but on the other hand sacrifice the accuracy of the model compare to a real system. For example the state-space can be expanded to keep track of the phase at which the processing of a lot gets interrupted due to a failure. Then after the failed tool is repaired the job can continue its processing from where it was interrupted. This modification adds k^{max} variables to the state-space to keep track of the lot on each failed machine.

2.2 Model Verification Through Simulation

In order to verify the results by the proposed Markov chain model we develop a simulation model of the toolset with the same characteristics and restrictions as mentioned in subsections 2.1 and 2.2. For a variety of inter-arrival and processing time distributions we compare the cycle time estimation of the proposed model with the cycle time of the simulated toolset. We make this comparison when the system in under heavy traffic since in light traffic the operational rules do not apply. We estimate the cycle time via both models for different utilization levels above 70% and calculate the percent difference at each point. We incrementally change the utilization by increasing the arrival rate to the system. We consider the average percent difference between simulation and the proposed model in high utilization as a quantitative measure of model accuracy.

2.3 Application of Erlang Distribution for Non-exponential Arrival and Service processes with Low Coefficient of Variation

Like many other processes, SMS toolsets often deviate from the exponential distribution in their inter-arrival, processing, time to fail and time to repair distributions. However, when solving complex models in queuing theory involvement of any non-exponential probability distribution complicates the task of finding the steady-state probabilities with analytical solutions (Cox 1955). One approximate method to overcome this issue is to match the first moment of a continuous non-exponential distribution with its exponential equivalent and use the latter in the model. However, this approximation adds to the inaccuracy of the cycle time estimation by the proposed model.

A different approach to this problem is application of phase-type distributions (Neuts 1994) to approximate any positive valued distribution. Phase-type distributions are dense in $[0,\infty)$ and can provide a close approximation to any positive continuous distribution. In the past twenty years the problem of fitting the parameters of a Phase-Type distribution has been studied extensively in the applied probability community and different approached have been proposed. These approaches can be classified in two categories of maximum likelihood estimates (MLE) and moment matching techniques (Dempster, Laird, and Rubin 1977), (Asmussen, Nerman, and Olsson 1996), (Lang and Arthur 1996), (Riska, Diev, and Smirni 2004), and (Horváth and Telek 2000). However, application of such complex distributions to the Markov chain model adds to the complexity of the model and the time to solve it.

To model the inter-arrival, processing, failure or repair times in a toolset by phase-type distribution we need to introduce new variables in the stochastic process underlying the Markov chain to indicate the phase in which each of these variables reside at any time t. In order to keep a balance between accuracy and complexity we use three categories of Erlang distributions with distinct SCV's and replace any non-exponential distribution with one of them that has the closest SCV to the original distribution.

More specifically we introduce the following Erlang categories of $Erlang(1,x)$,$Erlang(2,x)$ and $Erlang(10,x)$ where the rate parameter, x is to be adjusted to match the first moment of the original distribution. $Erlang(1,x)$ distribution is essentially the exponential distribution with rate x and Squared Coefficient of Variation (SCV, from now) equal to 1, $Erlang(2,x)$ is a right-skewed distribution with $SCV = 0.5$ and $Erlang(10,x)$ is a symmetric distribution with $SCV = 0.1$.

Therefore for any non-exponential distribution of inter-arrival or service we can pick the Erlang distribution from this set that has the closest SCV to the original distribution. Since any $Erlang(k,x)$

distribution consists of k phases of exponential distributions each with rate x, it can be easily integrated in the proposed Markov chain model to represent the original distribution.
In the next section we present numerical results for a number of cases with and without operational rules and with the application of the Erlang distribution.

3 NUMERICAL RESULTS

In this section we present the numerical results of applying the proposed model to a toolset with two parallel servers. We first develop a simulation model of this toolset as described in 2.3 and then build a Markov chain model that represents the same system as discussed in 2.1. We compare the results of the proposed model with the cycle time of the simulated toolset in the absence of any operational rule and also in presence of two types of operational rules, for adjusting the arrival rate and the failure rate according to the WIP level.
We also make the comparison with two $G/G/m$ models that are commonly used in manufacturing. In this case we do not apply any operational rule on the toolset since the queueing formulas are not capable of modeling those. However we show that even in the absence of the operational rules the proposed model gives more accurate estimation In both cases we measure the accuracy by average percent difference with simulation in heavy traffic as discussed in 2.3. We try different distributions for inter-arrival and service times to see the effect of approximation with Erlang distribution, as discussed in 2.4, on the accuracy of cycle time estimation by the proposed model.

3.1 Toolsets with Operational Rules

3.1.1 Case with Erlang Distribution

First we use three different Erlang distributions for the inter-arrival times and also for the processing times, namely exponential with $SCV = 1$, skewed with $SCV = 0.5$ and symmetric with $SCV = 0.1$. The distribution of time to failure and time to repair are both exponential with rate $d = 4$ and $u = 1$ in all cases. We apply three conditions regarding the operational rules to this system. The first case depicts the toolset in the absence of operational rules, in the second case (Rule I) the operational rule adjusts the arrival rate based on the WIP level and the third case (Rule II) adjusts the failure rate based on the WIP level. For each case all the 9 combinations of inter-arrival and processing time distributions are applied and the resutls are obtained. In the case of Rule I we decrease the arrival rate by 50% compared to the case with no rule, whenever the two servers are busy and increase the arrival rate when at least one server is idle. For Rule II the failure rate is decreased by 50% when the two servers are busy and is increased when there is at least one idle server. We calculate the average percent difference between the two models at utilization levels of 70%, 80% and 90%. We verify the performance of the model using simulation and present the results in Table 1.

Table 1: Average Percent Difference of Cycle Time Estimation by Simulation and The Proposed Model in Heavy Traffic

High Utilization ($\geq$ 70%)	Processing Time								
InterarrivalTime	Erlang(1, 1), SCV=1			Erlang(1, 2), SCV=0.5			Erlang(1, 10), SCV=0.1		
	NO Rule	*Rule I*	*Rule II*	*NO Rule*	*Rule I*	*Rule II*	*NO Rule*	*Rule I*	*Rule II*
Erlang-(1,X) SCV =1	-0.5%	-4.5%	-3.5%	-0.5%	-4.9%	-6.1%	-0.7%	-5.0%	-7.2%
Erlang-(2,X) SCV =0.5	0.8%	-4.1%	-4.7%	-0.5%	-4.5%	-6.6%	-0.5%	-4.5%	-9.6%
Erlang(10,X) SCV =0.1	-0.3%	-3.4%	-5.1%	-0.5%	-3.5%	-7.9%	-0.1%	-3.7%	-11.2%

As it is observed in this table that percent difference between simulation and Markov chain prediction of cycle time does not exceed 12% in any of the cases. This level of accuracy is higher than the other models in the literature, such as (Morrison and Martin 2007). The complex simulation models that are used in the semiconductor companies can predict the cycle time with at most 90% accuracy. For example Chen (Chen, Harrison, Mandelbaum, van Ackere, and Wein 1988)

state that their proposed queueing model predicts the cycle time with 12% accuracy.
However, since Erlang distribution is a special case of Gamma distribution and also all the failure and repair distributions are exponential, this level of accuracy is not surprising.

3.1.2 Case with Lognormal Distribution

We repeat the same experiment of 3.1.1 but with longnormal distributions for the inter-arrival and service time. We apply Rule I and Rule II respectively and measure the percent difference between the cycle time of simulation and the Markov model in high utilization. For this case we try two different methods of approximation for the sake of comparison.
First we directly match the first moment of the original distribution with that of exponential and use the exponential equivalent in the Markov model. The Results for this approximation method are shown in Table 2. Then we match an Erlang distribution from the chosen set to each distribution of arrival and service and show the results in Table 3
.

Table 2: Average Percent Difference of Cycle Time Estimation by Simulation and The Proposed Model in Heavy Traffic, with exponential approximation

High Utilization ($\geq$ 70%)	Processing Time								
InterarrivalTime	LogN(1, 1), SCV=1			LogN(1, 0.71), SCV=0.5			LogN(1, 0.32), SCV=0.1		
	NO Rule	*Rule I*	*Rule II*	*NO Rule*	*Rule I*	*Rule II*	*NO Rule*	*Rule I*	*Rule II*
LogN-SCV=1	-3.0%	-14.4%	-5.4%	9.2%	-9.8%	-4.9%	22.0%	-6.0%	-7.5%
LogN- CV=0.5	-21.1%	-17.2%	-19.1%	-9.2%	-12.6%	-20.4%	3.3%	-8.8%	-24.9%
LogN-SCV=0.1	-41.0%	-19.7%	-34.1%	-30.6%	-15.2%	-37.5%	-20.9%	-11.5%	-43.6%

Table 3: Average Percent Difference of Cycle Time Estimation by Simulation and The Proposed Model in Heavy Traffic, with Erlang approximation

High Utilization ($\geq$ 70%)	Processing Time								
InterarrivalTime	LogN(1, 1), SCV=1			LogN(1, 0.71), SCV=0.5			LogN(1, 0.32), SCV=0.1		
	NO Rule	*Rule I*	*Rule II*	*NO Rule*	*Rule I*	*Rule II*	*NO Rule*	*Rule I*	*Rule II*
LogN-SCV=1	-3.0%	-14.4%	-5.4%	-5.0%	-15.9%	-6.2%	-5.4%	-16.8%	-9.2%
LogN- CV=0.5	-1.4%	-8.2%	-3.2%	-2.4%	-9.1%	-4.6%	-2.0%	-9.5%	-7.7%
LogN-SCV=0.1	-0.1%	-3.6%	-2.7%	-0.4%	-4.4%	-4.9%	0.0%	-4.4%	-8.5%

It is observed that in the majority of cases using an Erlang distribution improves the results. This proves the fitting of a phase-type distribution a promising method to enhance the accuracy.

3.2 Classical $G/G/m$ Models

In this section we make a comparison between the classical $G/G/m$ models and the proposed Markov chain model in the absence of any operational rule. The classical queueing models all have the inherent assumption of independence between the arrival and service processes and hence are incapable of modeling operational rules. Hence in this section the comparison is only made in the absence of operational rules. We consider two $G/G/m$ approximations that are commonly used in manufacturing. Hopp and Spearman (Hopp and Spearman 2002) present the approximation proposed by Kingmanm (Equation 4) for the cycle time of a $G/G/m$ queue with failure prone servers and with the notion of the effective processing time, t_e. In this approximation ρ is the effective utilization of the system, m is the number of parallel servers and C_a^2 and C_e^2 present the squared coefficient of variation of effective processing time and inter-arrival time respectively.

$$CT \approx \left(\frac{C_a^2+C_e^2}{2}\right)\left(\frac{\rho^{\sqrt{2(m+1)}-1}}{m(1-\rho)}\right)t_e + t_e. \tag{4}$$

Buzacott and Shanthikumar also propose Equation 5 to approximate the cycle time of the $G/G/m$ queue based on the cycle time of the $M/M/m$ queue.

$$CT^{G/G/m} \approx \frac{C_a^2(1-(1-\rho)C_a^2)/\rho + C_e^2}{2} t_q^{M/M/m} + t_e, \tag{5}$$

where $t_q^{M/M/m}$ denotes the queue time of the $M/M/m$ system. We have examined the system under the same combinations of inter-arrival and processing time distributions as in 3.1 and with the same parameters for time to fail and time to repair. The comparison is done with a base simulation model as described in 2.3. The results are presented in Table 4.

Table 4: Average Percent Difference of Cycle Time Estimation by the Proposed Model and Two Classical $G/G/m$ Formulas Under Heavy Traffic and No Operational Rule

High Utilization (≥ 70%)	Ek/Ek/2											
Interarrival Time	Processing Time											
	Erlang(1, 1), SCV=1				Erlang(1, 2), SCV=0.5				Erlang(1, 10), SCV=0.1			
	Expo	*Erlang*	*H&S*	*B&S*	*Expo*	*Erlang*	*H&S*	*B&S*	*Expo*	*Erlang*	*H&S*	*B&S*
Erlang-(1,X) SCV =1	-0.5%	-0.5%	-44.9%	-51.3%	13.1%	-0.5%	-3.4%	-7.6%	25.7%	-0.7%	26.1%	23.5%
Erlang-(2,X) SCV =0.5	-18.6%	0.8%	-42.4%	-51.5%	-7.1%	-0.5%	-2.4%	-8.9%	5.0%	-0.5%	27.2%	22.6%
Erlang(10,X) SCV =0.1	-41.3%	-0.3%	-42.6%	-49.0%	-30.8%	-0.5%	0.0%	-3.8%	-21.0%	-0.1%	30.2%	28.2%

As we observe in Table 4 even in the case with no operational rule the proposed Markov chain model performs with higher accuracy than the two $G/G/m$ approximations of (Hopp and Spearman 2002) and (Buzaccott and Shantikumar 1993).

The results of average percent difference of cycle time (for utilization ≥ 70%) in the model without operational rules are shown in Table 5

Table 5: Average Percent Difference of Cycle Time Estimation by The Proposed Model and Two Classical $G/G/m$ Formulas Under Heavy Traffic and No Operational Rule

High Utilization (≥ 70%)	LogN/LogN/2											
Interarrival Time	Processing Time											
	LogN(1, 1), SCV=1				LogN(1, 0.71), SCV=0.5				LogN(1, 0.32), SCV=0.1			
	Expo	*Erlang*	*H&S*	*B&S*	*Expo*	*Erlang*	*H&S*	*B&S*	*Expo*	*Erlang*	*H&S*	*B&S*
LogN- SCV =1	-3.0%	-3.0%	-47.8%	-54.4%	9.2%	-5.0%	-8.1%	-12.4%	22.0%	-5.4%	22.5%	19.8%
LogN- SCV =0.5	-21.1%	-1.4%	-45.4%	-54.6%	-9.2%	-2.4%	-4.3%	-11.0%	3.3%	-2.0%	25.9%	21.2%
LogN- SCV =0.1	-41.0%	-0.1%	-42.1%	-48.5%	-30.6%	-0.4%	0.1%	-3.7%	-20.9%	0.0%	30.2%	28.2%

In Table 5 we observe that, we can see that for the base case, without operational rules, the approximation always perform better (with regard to the classical queueing models). In Figure 2 we can see that for the base case with inter-arrival and service time $SCV = 1$, the result for both models (Markov-Expo and Markov-Erlang), are exactly the same as we expect. For another case whit inter-arrival and service time $SCV = 0.71$, the results with the Markov-Expo are better than the classical queueing theory, but the Markov-Erlang gives the closest results to simulation, as it can be observed in Table 5

For the model with Rule I, the average percent differences of the cycle time approximations are presented in the Table 6; and for model with rule II is presented in Table 7.

Table 6: Average Percent Difference of Cycle Time Estimation by The Proposed Model and Two Classical $G/G/m$ Formulas Under Heavy Traffic and Operational Rule1

High Utilization (≥ 70%)	LogN/LogN/2 (RULE 1)					
Interarrival Time	Processing Time					
	LogN(1, 1), SCV=1		LogN(1, 0.71), SCV=0.5		LogN(1, 0.32), SCV=0.1	
	Expo	*Erlang*	*Expo*	*Erlang*	*Expo*	*Erlang*
LogN- SCV =1	-14.4%	-14.4%	-9.8%	-15.9%	-6.0%	-16.8%
LogN- SCV =0.5	-17.2%	-8.2%	-12.6%	-9.1%	-8.8%	-9.5%
LogN- SCV =0.1	-19.7%	-3.6%	-15.2%	-4.4%	-11.5%	-4.4%

As can be see in Table 6, using the operational rule I (described in section 3.1), the results of the Markov model for service time with SCV of 1 or 0.5 (all cases of interarrival times SCV) are almost always better than the exponential approximation. Also, the worst error is around 17%. One examples of these results are also presented in Figure 3

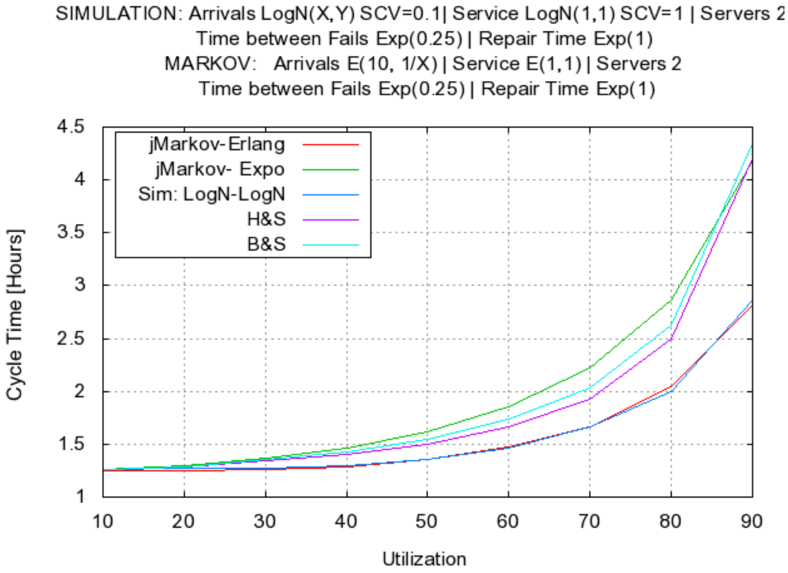

Figure 2: Results for LogN-LogN without rules with service time SCV=1 and arrival time SCV=0.1

Table 7: Average Percent Difference of Cycle Time Estimation by The Proposed Model vs 4 and 5 Under Heavy Traffic and Operational Rule2

High Utilization ($\geq$ 70%)	LogN/LogN/2 (RULE 2)					
InterarrivalTime	Processing Time					
	LogN(1, 1), SCV=1		LogN(1, 0.71), SCV=0.5		LogN(1, 0.32), SCV=0.1	
	Expo	*Erlang*	*Expo*	*Erlang*	*Expo*	*Erlang*
LogN-SCV=1	-5.4%	-5.4%	-4.9%	-6.2%	-7.5%	-9.2%
LogN- CV=0.5	-19.1%	-3.2%	-20.4%	-4.6%	-24.9%	-7.7%
LogN-SCV=0.1	-34.1%	-2.7%	-37.5%	-4.9%	-43.6%	-8.5%

The result for the model with operational Rule II (Table 7) show that the Markov with Erlang approximation is always better than the Exponnetial approximation . In this case the worst error is around 9%. One example of this behavior is presented in Figure 4

3.3 Efficiency

The efficiency of the Markov model is also measured in terms of the computational time. The measurement of efficiency is made using the ratio between the simulation running time and the Markov model running time $r = \frac{Sim.RunningTime}{Markov.RunningTime}$. The results are presented in the Table 8

In Table 8 can be see that only in the cases with very low variability, the simulation running time is lower than the Markov model running time. In all the other cases the running time of the Markov model is much lower than the simulation model.

4 CONCLUSIONS

The proposed Markov chain model along with the Erlang fitting method provides a practical tool to estimate cycle time for toolsets with operational rules. The accuracy of estimation through this model is verified by simulation. Compared with simulation this model is easier and faster to use and maintain with less input data requirements. Compared

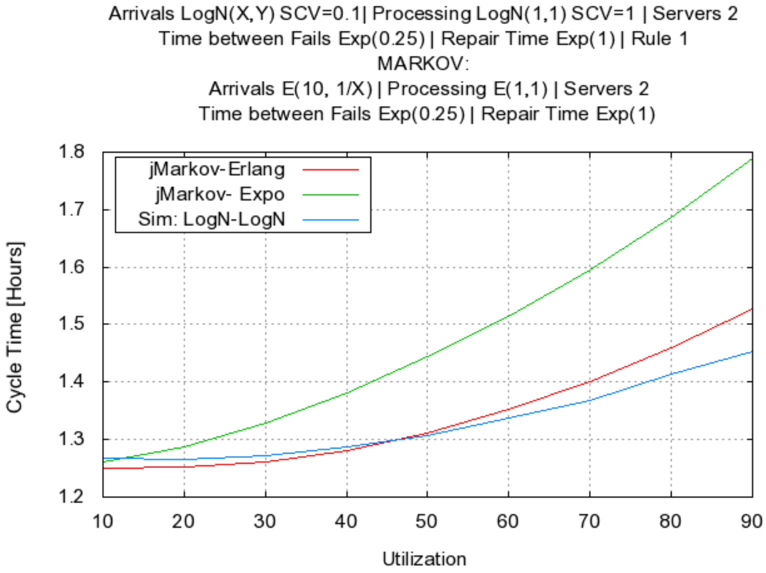

Figure 3: Results for LogN-LogN with Rule 1 and Service time SCV=1 inter -arrival times SCV=0.1

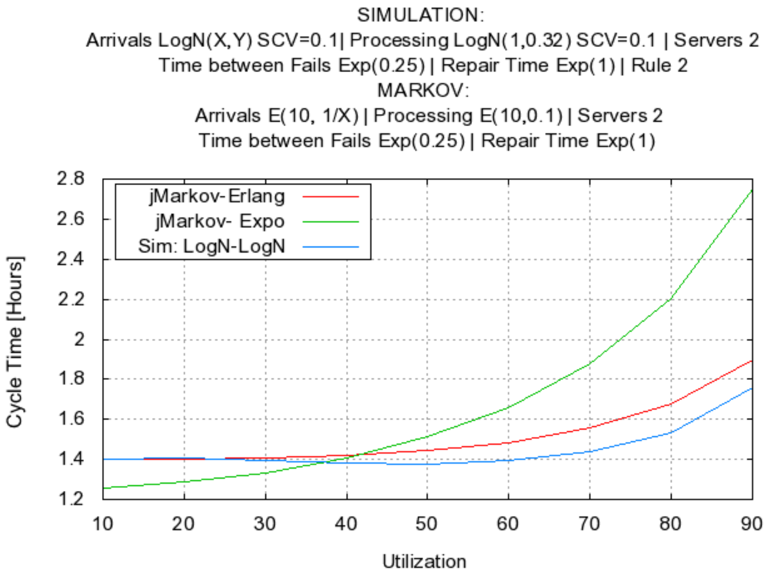

Figure 4: Results for LogN-LogN with Rule 2 and Service Time SCV=0.1 Inter-Arrival Time SCV=0.1

to the classical G/G/m formulas that are not capable of modeling operational rules the proposed model provides a flexible modeling tool with high accuracy. However, some inaccuracy is introduced due to approximation of a general distribution with Erlang which also poses a limit to the use of distributions with $SCV \leq 1$. The next steps to improve this model include the application of phase-type distributions for a more efficient approximation of any positive distribution as well as adding capabilities for modeling multiple types of failures and multiple types of products with priority.

Table 8: Ratio of the running time efficiency of the Markov model and number of states

SCV Serv. Time		*1*	*1*	*1*	*0.5*	*0.5*	*0.5*	*0.1*	*0.1*	*0.1*
SCV Int. Arriv Time		*1*	*0.5*	*0.1*	*1*	*0.5*	*0.1*	*1*	*0.5*	*0.1*
		Ratio	Ratio	Ratio	Ratio	Ratio	Ratio	Ratio	Ratio	Ratio
Utilization	**70%**	100.63	102.56	9.81	97.69	35.23	2.42	1.15	0.22	0.00
	80%	115.60	56.97	13.28	55.45	56.97	2.81	1.24	0.24	0.00
	90%	119.13	121.06	9.70	117.19	40.55	2.74	1.33	0.25	0.00

REFERENCES

Akhavan-Tabatabaei, R., S. Ding, and G. Shanthikumar. 2009. A method for cycle time estimation of semiconductor manufacturing toolsets with correlations. In M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin and R. G. Ingalls, editors, *Proceedings of the 2009 Winter Simulation Conference* 41 (2): 1719 - 1729.

Asmussen, S., O. Nerman, and M. Olsson. 1996. Fitting phase-type distributions via the em algorithm. *Scandinavian Journal of Statistics* 23 (4): 419–441.

Buzaccott, J., and J. G. Shantikumar. 1993. *Stochastic models of manufacturing systems*. Prentice Hall.

Chen, H., J. M. Harrison, A. Mandelbaum, A. van Ackere, and L. M. Wein. 1988. Empirical evaluation of a queueing network model for semiconductor wafer fabrication. *Operations Research* 36 (2): 202–215.

Cox, D. R. 1955. A use of complex probabilities in the theory of stochastic processes. *Mathematical Proceedings of the Cambridge Philosophical Society* 51 (02): 313–319.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of The Royal Statistical Society, Series B* 39 (1): 1–38.

Hopp, W., and M. Spearman. 2002. *Factory Physic, Foundations of Manufacturing Management*, Volume 2nd ed. McGraw-Hill.

Horváth, A., and M. Telek. 2000. Approximating heavy-tailed behaviour with phase-type distributions. In 3rd International Conference on Matrix-Analytic Methods in Stochastic Models.

Lang, A., and J. L. Arthur. 1996. Parameter approximation for phase-type distribution. S. R. Chakravarty and A. S. Alfa. Matrix-Analytic Methods in Stochastic Models, Lecture Notes in Pure and Applied Mathematics.

Latouche, L., and V. Ramaswami. 1994. *Introduction to matrix analitic methods in stochastic modelling*. ASA-SIAM, NY.

Morrison, J., 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.

Neuts, M. F. 1994. *Matrix-geometric solutions in stochastic models: An algorithmic approach*. Dover Publications, NY.

Riska, A., V. Diev, and E. Smirni. 2004. An em-based technique for approximating long-tailed data sets with ph distributions. *Performance Evaluation* 55 (1-2): 147–164.

Shanthikumar, J., S. Ding, and M. Zhang. 2007, oct.. Queueing theory for semiconductor manufacturing systems: A survey and open problems. *Automation Science and Engineering, IEEE Transactions on* 4 (4): 513 –522.

AUTHOR BIOGRAPHIES

RAHA AKHAVAN-TABATABAEI is currently an assistant professor at Universidad de los Andes, Bogota, Colombia. Prior to this she worked for Intel Corporation, AZ, USA, as a senior industrial Engineer for HVM fabs. She has a PhD in Industrial and Systems Engineering from North Carolina State University. Her email address is <r.akhavan@uniandes.edu.co>.

JUAN JOSE UCROS is currently a master student at Universidad de los Andes, Bogota, Colombia. He is also an intern at Colombiana Kimberly Colpapeles. His email address is <jj.ucros122@uniandes.edu.co>.

J. GEORGE SHANTIKUMAR, Ph.D is a professor of industrial engineering and operations research at the University of Purdue, Indiana. Dr. Shanthikumar served as consultant for KLA-Tencor Corp., and worked on joint development projects with AMD, IBM, Intel, LSI, Motorola, Texas Instruments, Toshiba, Fujitsu, TSMC and UMC. His email address is <shanthikumar@purdue.edu>.