Proceedings of the 2001 Winter Simulation Conference
B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, eds.

AN OVERALL FRAMEWORK FOR GENERATING SIMULATION-BASED CYCLE TIME-THROUGHPUT CURVES

Sungmin Park

Managerial Research Team
Korea Gas Corporation
215, Chongja-dong, Pundang-gu
Songnam, Kyonggi-do, 463-754, KOREA (SOUTH)

Gerald T. Mackulak
John W. Fowler

Department of Industrial Engineering
Arizona State University
PO Box 875906
Tempe, AZ 85287-5906, U.S.A.

ABSTRACT

A simulation-based cycle time-throughput curve requires a large amount of simulation output data, and an experimentation framework is needed to enhance the precision and accuracy of a simulation-based cycle time-throughput curve. In this research, approaches and solutions are presented on three prime issues: 1) the establishment of the simulation sampling strategies; 2) the determination of the simulation sequences; and 3) the determination of the length of a simulation run. First, strategic simulation sampling guidelines are proposed as to how to use a fixed amount of samples when trying to generate a precise and accurate cycle time-throughput curve for complex systems. Second, in order to provide good references for the sequential experiments to generate precise simulation-based cycle time-throughput curves, a set of discrete design points is ranked sequentially. Third, a sequential stopping rule is developed to determine the length of a simulation run based on a time series forecasting procedure.

1 INTRODUCTION

1.1 Motivation

Effective manufacturing capacity management is one of the major issues confronting manufacturing industries in increasing the productivity of a system. Specifically, in factory level operations, a production manager wants to concurrently maximize throughput while minimizing cycle time and *Work In Process* (*WIP*). In today's competitive environment, short cycle times are critical. Short cycle times can be achieved by: 1) lowering the throughout of a system; 2) providing more capacity to the system; or 3) decreasing the variability in the system (Fromm 1992). Especially in a technology-driven environment like a semiconductor wafer fabrication facility (*fab*), a reduction in the system variability may be the only short-term cycle time improvement method possible without decreasing throughput. This is because it may take several months to receive a newly ordered piece of equipment.

Because a single estimate of cycle time shows only a local snapshot of a system, it is often preferable to characterize the system by a *Cycle Time-THroughput* (*CT-TH*) curve over a range of throughput rates, which provides a comprehensive profile of the system. This is particularly important in supply chain management activities when one wants to concisely represent performance trade-offs of a factory.

A CT-TH curve is a powerful tool in evaluating the trade-off relationships between cycle time, throughput and WIP. The curve shows the predicted average cycle time plotted against throughput rate, and it also illustrates the variability of cycle time as well as the asymptotic system capacity. The sensitivity of cycle time to throughput can be appraised by examining the curvature or steepness of the curve. Different operating policies can be quantitatively evaluated by reviewing curve shifts (Dayhoff and Atherton 1986, Fowler *et al.* 1997). In a queueing system, the traffic intensity ρ is defined as $\rho = \lambda/\mu$ where λ is the arrival rate, and μ is the service rate. This ρ can substitute for the system throughput (output; i.e. x-axis in Figure 1) since the throughput is the arrival rate (input rate) times the yield. Figure 1 shows the analytical CT-TH curve of an *M/M/1* queueing model with a service rate μ equal to unity.

If the underlying system is simple enough, it may be possible to obtain a mathematical equation for the CT-TH curve. As the system increases in complexity, discrete event simulation becomes the most viable approach for generating the CT-TH curve because simulation can capture any level of manufacturing detail. However, the problem is that extensive detail increases the amount of model construction time and computer run time necessary for statistically precise and accurate output generation. Increased computer run times lead to increased cost so that simulation might ultimately be an expensive option (Brown *et al.* 1997).

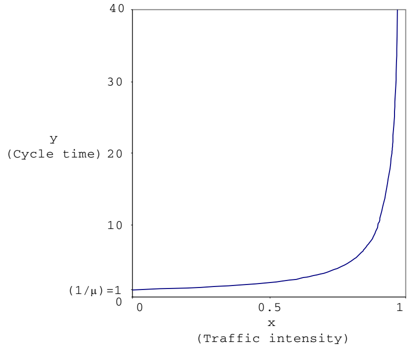

Figure 1: A CT-TH Curve of an *M/M/1* Queueing Model

Within this context, the focus of this research is to develop a methodology that leads to efficient CT-TH curve generation using discrete event simulation. An overall framework for generating simulation-based CT-TH curves presented in this paper should increase the productivity of simulation modeling and experimentation. Consequently, it will assist manufacturing companies in deriving some tangible benefits with respect to: 1) cost reduction by CT-TH curve based operations analysis; 2) pertinent decision making on manufacturing capacity management and equipment purchases through improved precision and accuracy of cycle time and throughput estimates; and 3) improvement of supply chain management decisions by representing the performance trade-offs of a factory.

1.2 Shapes of CT-TH Curves

In manufacturing systems, two different shapes of CT-TH curves are often encountered; i.e. monotonically increasing shape and U-shape. Figure 2 shows the two different shapes of CT-TH curves (Park 2000).

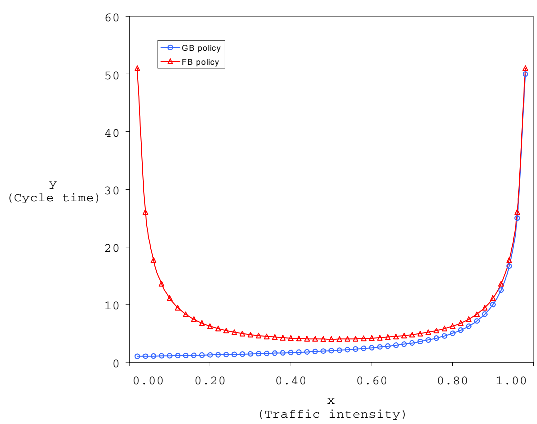

Figure 2: Two Different Shapes of CT-TH Curves

We denote a *Minimum Batch Size* (*MBS*) policy by (*a*, *b*) where *a* and *b* are integers satisfying $1 < a < b$, and *a* and *b* indicate the minimum and maximum batch size, respectively. If *a* is equal to *b*, then the MBS policy becomes a *Full Batch* (*FB*) policy as (*b*, *b*). If *a* is equal to

a *Full Batch* (*FB*) policy as (*b*, *b*). If *a* is equal to unity, the MBS policy becomes a *Greedy Batch* (*GB*) policy as (1, *b*).

A monotonically increasing CT-TH curve is expected when a manufacturing system has no batch processing or adopts a GB policy. If a FB or a MBS policy is adopted, a U-shaped CT-TH curve generally results (Wright Williams & Kelly 1999). For the monotonically increasing CT-TH curve (GB policy in Figure 2), there is one "*knee*", one asymptote and an intercept at the zero traffic intensity limit. When the traffic intensity *x* is close to zero, cycle time is almost equal to the pure processing time. If the traffic intensity *x* approaches the capacity of the system, the cycle time goes to infinity due to an ever-increasing queue. In the vicinity of the knee, the cycle time changes rapidly. The curvature is dependent on characteristics of the system. One of the most important characteristics is the variability in the system. More variability leads to a steeper CT-TH curve. In the extreme case, a system with zero variability would have a CT-TH curve that is a horizontal line; i.e. the cycle time would be the same no matter how much the system is utilized up to its maximum capacity (Fromm 1992). For the U-shaped CT-TH curve (FB policy in Figure 2), there is a region of relatively low traffic intensities where the cycle time actually decreases as the traffic intensity increases. It can take a long time to fill up a batch when the traffic intensity is low. Hence, two distinct knees exist. The two asymptotes illustrate the extreme cases of this behavior.

1.3 Considerations for Generating a Simulation-Based CT-TH Curve

Let's suppose that a simulation experimenter is interested in the range of traffic intensity *x* beyond 0.60 with respect to the CT-TH curve in Figure 1. In Figure 3, the analytical CT-TH curve is reproduced as the curve connecting the black dots within the range of interest. Without the mathematical queueing equation for the CT-TH curve, a simulation-based CT-TH curve can be generated like the curve connecting the big hollow dots in Figure 3.

However, in a naive simulation experimentation, the simulation-based CT-TH curve may have some problems on the precision and accuracy of the curve regardless the extent of simulation efforts.

Note that the design points of traffic intensity $x = 0.70$, 0.80, 0.90 and 0.97 are selected to be investigated for generating cycle time estimates in Figure 3. At each design point, four simulation replicates are generated, and a "*grand*" mean is calculated using the four cycle time estimates. Actually, the simulation-based CT-TH curve is generated using the total of forty replicates, ten replicates at each design point. However, for the illustration purpose in this paper, only four small hollow dots representing four cycle time estimates are shown at each design point in order to highlight the unequal variances through the design points interested. The actual sample variances of every ten replicates are 0.1716, 1.1139, 6.2430 and 22.0620 at the design points $x = 0.70$, 0.80, 0.90 and 0.97 respectively. In the simulation, the length of every single simulation run is 1000 observations, and there is no data truncation.

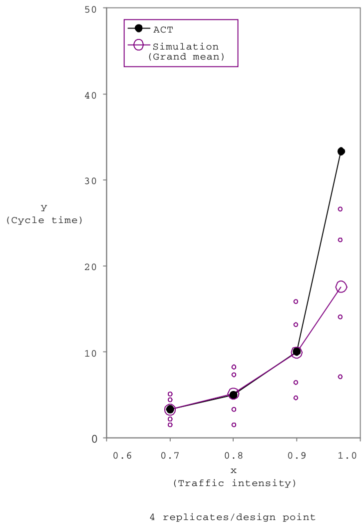

Figure 3: Comparison of the Analytical CT-TH Curve with the Simulation-Based CT-TH Curve

We also note that the simulation-based CT-TH curve is very close to the *Analytical Cycle Time* (*ACT*) up to the design point $x = 0.90$. However, at the design point $x = 0.97$, there is a big bias between the value of the ACT and the grand mean. This is because 1000 observations are not enough to overcome initialization bias.

Simulation experimenters should consider several things when building a simulation experimentation framework. Considerations should be made regarding: 1) determination of the interesting range of traffic intensity to be investigated; 2) determination of specific design points (e.g. specific values of traffic intensity) to be investigated in the range of interest; 3) determination of the sequence of simulation experiments according to the importance of the selected design points in order to improve not only the precision of a simulation-based CT-TH curve but also the efficiency of the simulation experimentation; 4) allocation of simulation efforts to the selected design point (e.g. choose the number of simulation replications at a design point and/or specify the length of a simulation run); and 5) determination of whether or not to truncate initial transient data to reduce the initialization bias. Careful analysis of these general considerations can lead to efficient simulation-based CT-TH curve generation.

2 PHASES TOWARDS THE APPROACHES AND SOLUTIONS

This research is composed of three successive phases listed below to achieve an overall framework for generating simulation-based CT-TH curves.

- Phase 1: *Efficient CT-TH curve generation using a fixed sample size procedure*
- Phase 2: *D-optimal sequential experiments for generating a simulation-based CT-TH curve*
- Phase 3: *A sequential stopping rule for a steady state simulation based on a time series forecasting*

Phase 1 deals with the simulation sampling methods and sampling weights simultaneously in order to establish the simulation sampling strategies for efficient generation of precise and accurate simulation-based CT-TH curves for complex systems. In Phase 1, only the monotonically increasing shaped curve is studied. Phase 1 led to a paper that has been accepted to appear in the *International Journal of Production Research (*Fowler *et al.* 2001).

Phase 2 and 3 reinforce the aspect of the simulation sampling weights more mathematically. Basically, Phase 2 focuses on the improvement of the precision of a simulation-based CT-TH curve. To improve the precision as well as to avoid non-significant simulation experiments, Phase 2 presents a mathematical procedure and referential solutions for optimal experimental sequences. Both shapes of CT-TH curves as shown in Figure 2 are considered in Phase 2. Phase 2 led to a paper that has been accepted to appear in *Operations Research* (Park *et al.* 2001). Finally, Phase 3 is more oriented to a general topic of simulation output analysis for achieving accuracy (i.e. the unbiased stable estimation), which results in developing a sequential stopping rule for a steady state simulation output process using a time series forecasting procedure. Phase 3 led to a paper that has been submitted to *SCS Transactions* (Mackulak *et al.* 2001).

3 PHASE 1: ESTABLISHMENT OF STRATEGIES

3.1 Components of the Strategy

It is reasonable to assume that the budget and resources for any project are limited, so the objective of this phase is to provide guidelines for using a fixed amount of samples when trying to generate a precise and accurate simulation-based CT-TH curve for complex systems. The key of this fixed sample size procedure is the determination of how to organize simulation experimentation for generating a CT-TH curve with sufficiently small variance and bias. Each strategy uses a unique combination of sampling methods and weights to discriminate its performance.

- *Sampling method*: The sampling method determines the required simulation replications. The combination of *Variance Reduction Techniques* (*VRTs*) is related to the sampling method.
- *Sampling weights*: The sampling weights dictate not only the selection of appropriate design points in the experimental design space (i.e. in the interesting range of traffic intensity) but also the importance of each design point (i.e. how many samples are to be allocated to each design point).

Obviously, these two components are not independent, so this phase seeks to find the best combinations of the method and weights depending on the limitation of samples and the complexity of the system.

3.2 Evaluation of the Strategy: Scenarios and Regression Analysis

In order to determine the efficiency of various simulation strategies, several scenarios are developed in this phase. A "*scenario*" describes a simulation sampling strategy for one set of simulation experiments. The scenarios are developed by three stages below.

Using the transformed simulation data, linear regression analysis generates condensed measures of performance of a simulation sampling strategy. *Mean Square Error* (*MSE*) of a fitted regression model is the major performance measure of each simulation sampling strategy.

- Stage 1: *Determine a general simulation structure*
- Stage 2: *Assign the sampling weight at each design point*
- Stage 3: *Develop the sampling methods using VRTs*

3.3 Rationale of the Strategy

3.3.1 Sampling Methods

Common Random Numbers (*CRNs*) and *Antithetic Variates* (*AVs*) are generic VRTs that can be implemented in generating a simulation-based CT-TH curve. When CRNs and AVs are applied to simulation experiments for generating a monotonically increasing CT-TH curve, they show typical behavior with respect to the curvature.

Three different curves are shown in Figure 4-(a). Each curve connects four points. With the symbol •, the ACT is plotted, where the values are calculated by the analytical equation for the *M/M/1* queueing model. Comparing the curve using the ACT with the CRNs (+) and *Independent Random Numbers* (*IRNs*, Δ) curves, it is possible to visualize the bias between the ACT curve and the estimated curve from simulation experiments. The length of all simulation runs is equal to 1000 observations, and no initial transient data are truncated. Using CRNs, if the first design point overestimates cycle time, likely others will as well. CRNs help in generating the right shape (i.e. monotonically increasing shape) of the curve, but there may be a shift (upwards in this example).

Figure 4-(b) shows curves with AVs included. The curve with the ° symbol represents the sample mean of runs with CRNs and AVs at each design point. Notice that this curve is closer to the ACT curve. In the pairs of AVs runs, one is overestimated, the other is underestimated, so that the two CT-TH curves form the outward funnel shape.

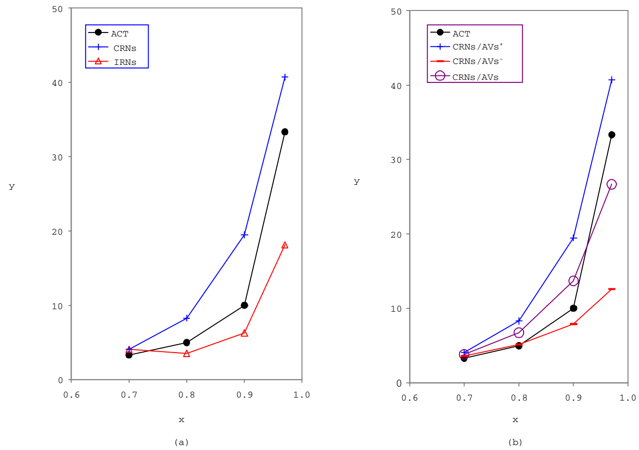

Figure 4: Behavior of VRTs on the Curvature

3.3.2 Sampling Weights

The point with the ° symbol in Figure 5-(a) represents the grand mean cycle time using ten replicates. In creating Figure 5, each run is for 1000 observations, and no initial transient data are truncated.

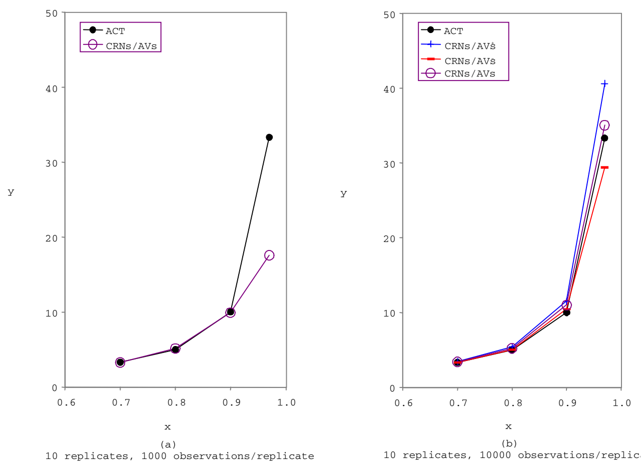

Figure 5: Impact of Sampling Weights on the Curvature

Apart from variance, Figure 5-(a) reveals the potential problem of the initial transient when an equal length of run is applied to simulation experiments with empty and idle initial conditions through all design points. Up to the design point x = 0.90, the length of 1000 observations produces accurate estimates when the combination of CRNs and AVs is incorporated into the sampling method. At the design point x = 0.97, the length of 1000 observations does not overwhelm the initialization bias. Because the nonlinear curvature in the design space near unity is crucial for the overall curvature of a CT-TH curve, it is important to obtain an accurate estimate of the cycle time in that region.

The conditions of Figure 5-(a) are applied for Figure 5-(b), but the simulation run length is elongated up to 10000 observations. The +, - and ° symbols represent the grand sample means derived from the ten independent CRNs runs, the antithetic runs, and the average of the AVs pairs, respectively. The CT-TH curve using ten independent sample means of AVs average replications approximately coincides with the ACT curve. A change of curvature is noticed when compared to Figure 5-(a). Particularly, the advantage of a longer simulation run is realized when the traffic intensity is very high (e.g. x = 0.97). In generating a monotonically increasing CT-TH curve, Figure 5-(b) indicates that simulation efforts should be concentrated on the design points around the highest traffic intensity with respect to the accuracy of the curve.

3.4 Simulation Sampling Strategies

Important simulation sampling strategies established in Phase 1 are summarized as follows. See Fowler *et al.* (2001) for more details.

- *For small samples and/or complex systems*: Strategy with different sampling weights and AVs is most appropriate. AVs are robust to the sample size and effective to reduce the variance as well as bias of a simulation-based CT-TH curve.
- *For large samples*: Strategy with equal sampling weights and CRNs/AVs is most appropriate. CRNs become more effective as the sample gets larger.
- *When the sufficiency of samples and/or the complexity of systems are uncertain*: It is recommended to use different sampling weights and AVs (i.e. a “*conservative*” strategy).
- *For many candidate design points and/or small samples and/or complex systems*: It is recommended to eliminate one or two inner design points (i.e. an “*aggressive*” strategy).

4 PHASE 2: DETERMINATION OF SEQUENCES

Some design points seem to be more important to building the right shape of a CT-TH curve in Figure 2. For example, design points located around the knee and the center of the curve or the two extremities of the design space seem to dominant points at other locations.

In Phase 2, a set of discrete design points (i.e. ten design points of x in Figure 2, varying from 0.10 to 0.90 in increments of 0.10 and the high traffic case of 0.97) is prioritized sequentially using a determinant criterion for a general class of ideally developed CT-TH curves shown in Figure 2.

First, a nonlinear expectation function is established to describe the curvature of each shape of CT-TH curve in Figure 2. Second, using the nonlinear expectation function, a D-optimal design pinpoints p starting design points all together simultaneously, and then the remaining n - p candidate design points are ranked sequentially for each shape of curves in Figure 2. The sequences proposed in this phase provide a good source of information with which a simulation experimenter can design simulation experiments without non-significant simulation sampling. Manufacturing production lines with respect to different batch policies can be systematically simulated based on the sequences.

4.1 Generalized Nonlinear Expectation Functions

Initially, it is assumed that a simulation experimenter is capable of defining the basic shape of a CT-TH curve for the system concerned. From a nonlinear regression point of view, this means that a p-parameters nonlinear expectation function for fitting the CT-TH curve can be established prior to developing a type of optimal design. In Phase 2, the D-optimality criterion is used to determine the simulation sequences.

To establish two different kinds of generalized nonlinear expectation functions, two approximations of the average cycle time are used associated with: 1) a *G/G/1* queueing model (Hopp and Spearman 1996); and 2) a $G/G^{(b_p)}/c$ queueing model with a FB policy in a batch queue with general arrivals, a general service, and multiple servers (Phojanamongkolkij 2000).

A generalized nonlinear expectation function of a CT-TH curve is required to satisfy some necessary conditions (Park 2000). Additionally, to decide the appropriate value of p, the minimum number of data points required for generating a CT-TH curve should be considered. In Figure 2, three principal data points can configure the monotonically increasing CT-TH curve. Two data points could be located around the two asymptotes, and one data point could be placed at the knee of the curve as an “*anchor*” point. Similarly, for the U-shape CT-TH curve in Figure 2, four data points are needed since the U-shape CT-TH curve requires two anchor points at the two different knees.

Considering the necessary conditions and the minimum number of data, two cycle time approximations are modified into generalized 3- and 4-parameters nonlinear expectation functions. Assuming that all parameters are equal to unity, the two CT-TH curves in Figure 2 are generated by Equations (1) and (2).

$$f(x,\boldsymbol{\theta}) = \frac{\theta_1 x}{\theta_2 - x} + \theta_3 \quad (1)$$

$$f(x,\boldsymbol{\theta}) = \frac{\theta_3}{x} + \frac{\theta_1 x}{\theta_2 - x} + \theta_4 \quad (2)$$

Obviously, the values of the parameters define the shape of each curve. For example, in Equation (1), when x (i.e. the traffic intensity) approaches zero, the value of $f(x,\boldsymbol{\theta})$ gets close to θ_3. However, in Equation (2), the value of $f(x,\boldsymbol{\theta})$ goes to infinity as x becomes close to zero. When x approaches θ_2 (i.e. the capacity), the values of $f(x,\boldsymbol{\theta})$ in Equations (1) and (2) go to infinity.

4.2 Procedure for the Determination

Park (2000) illustrated a detailed mathematical procedure for determination of the sequence of simulation experiments for generating two different shapes of CT-TH curves in Figure 2. The procedure is divided into three stages as follows.

- Stage 1: *Building the derivative matrix*
- Stage 2: *Solving a nonlinear optimization problem for the p starting design point*
- Stage 3: *Rank the remaining n - p candidate design points*

4.3 D-optimal Sequences

The U-shape CT-TH curve in Figure 2 can be generated by the D-optimal sequence below, and it shows some features.

(0.10, 0.50, 0.90, 0.97) → 0.40 → 0.60 → 0.80 →0.20 → 0.30 → 0.70

First, among the four starting design points, two design points x = 0.10 and 0.97 seem to be attributed to the asymptotes, and the design point x = 0.90 can be regarded as the right anchor point. As for the design point x = 0.50 (an approximate center point in the design space between x = 0.10 and 0.97), it seems to be selected in order to explain the flat region in the middle of the U-shape CT-TH curve. Second, the first two sequential design points x = 0.40 and 0.60 reinforce the precise estimation on the middle of the curve. Then the remaining design points are arranged to approach from the two extremities to the center of the design space.

For generating the monotonically increasing CT-TH curve in Figure 2, the D-optimal sequence of design points is determined as below. With the monotonically increasing CT-TH curve, all sequential design points are arranged to approach from the two extremities to the center of the design space.

(0.10, 0.90, 0.97) → 0.20 → 0.30 → 0.80 →0.40 → 0.50 → 0.70 → 0.60

5 PHASE 3: DETERMINATION OF THE LENGTH OF A SIMULATION RUN

If the important design points are already identified, then a subsequent issue might be how long a simulation run should be at a specific design point to generate accurate data.

When a simulation output process is a covariance stationary stochastic process generated from a steady state simulation, it is known that a time series of the cumulative sample means of the process has some "*asymptotic*" properties in favor of detecting the point where a precise and accurate estimate can be obtained, not where the steady state begins (Law 1977, Gafarian, Ancker and Morisaku 1978).

Because the cumulative sample mean averages out the raw data and produces a "*smooth*" statistic, it is expected that the cumulative sample mean will converge to a certain constant level. Therefore, if a sequential stopping rule seeks to be a simple algorithm as well as provide a stable estimate, a time series of the cumulative sample means can be a good data flow for those purposes.

In Phase 3, a sequential stopping rule is developed, which is a compromise between the efficiency and the simplicity of the algorithm for determining the length of a simulation run. But the problem remains as to how a simulation model detects a stopping point by itself. A time series forecasting procedure is incorporated into the sequential stopping rule to monitor the time series of the cumulative sample means continually and finally to detect a stopping point. See Mackulak *et al.* (2001) for more details.

5.1 Forecasting Procedure

Suppose that an observation Y_T is collected at time T from a steady state process. The cumulative sample mean at time T, C_T is updated as Equation (3).

$$C_T = \frac{1}{T}\sum_{t=1}^{T} Y_t \quad (3)$$

After the "*cumulative mean curve*" sufficiently flattens out to a near constant level, it is presumed that the cumulative sample mean C_T is almost unchanged over time, or, if it is changing, it is doing so very slowly. At this point, it is supposed that the level is near the steady state mean μ, and that C_T is regarded as a stable estimate of μ. Hence, the simulation output process can be terminated.

In this case, an appropriate model for a time series of the cumulative sample means can be defined as Equation (4),

$$C_T = \mu + \varepsilon_T \tag{4}$$

where ε_T is a random component having mean 0 and variance σ_ε^2.

If a time series follows a constant mean process as shown in Equation (4), the simple exponential smoothing method can be used as a forecasting procedure. Simple exponential smoothing is defined as equation (5),

$$S_T^{[1]} \equiv S_T = \alpha C_T + (1-\alpha) S_{T-1} \tag{5}$$

where S_T is called "*smoothed value*", and the fraction α is called the "*smoothing constant*" (Montgomery, Johnson and Gardiner 1990).

In general, the τ-period-ahead forecast error computed for period *T* is the actual value in period *T* minus the forecast for period *T* made at the end of period $T-\tau$ as defined in Equation (6),

$$e_\tau(T) = C_T - \hat{C}_T(T-\tau) = C_T - S_{T-\tau} \tag{6}$$

where $\hat{C}_T(T-\tau)$ is the forecast for period *T* made at the end of period $T-\tau$.

In considering the pattern of the cumulative mean curve, it may be logical to: 1) give more weight to recent forecast errors than to older ones; and 2) use the absolute values of the forecast errors to circumvent the oscillations of the curve. Hence, in the sequential stopping rule, the cumulative sum of the last *N* absolute 1-period-ahead forecast errors, E_T, as shown in Equation (7) can be a logical measure.

$$E_T = \sum_{t=T-N+1}^{T} |e_1(t)| \tag{7}$$

Consequently, if the decision criterion defined as equation (8) is satisfied, the simulation output process is terminated,

$$E_T < (\gamma \times C_T) \tag{8}$$

where γ is the relative percentage error applied to C_T. This means that if E_T is less than 100γ % of the current cumulative sample mean, it is assumed that the process can provide a stable estimate.

5.2 Algorithm

The algorithm of the sequential stopping rule has an initialization step and three main steps like below. The three main steps are activated whenever new data is collected from a process. In Figure 6, the detailed logic flow of the algorithm is presented.

- Step 0: *Initialize the parameters*
- Step 1: *Collect the new data Y_T and update C_T*
- Step 2: *Perform the forecasting procedure*
- Step 3: *Compare E_T with ($\gamma \times C_T$)*

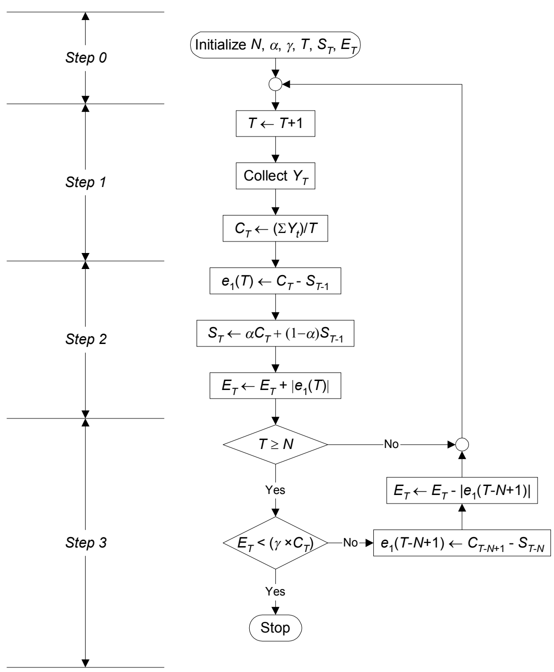

Figure 6: Logic Flow Diagram of the Algorithm

6 OVERALL FRAMEWORK

To assist simulation experimenters trying to overcome the typical problems associated with generating a simulation-based CT-TH curve, this research presented the approaches and solutions for efficient CT-TH curve generation techniques using discrete event simulation.

In this research, some queueing models (e.g. *M/M/1*, *G/G/1 and* $G/G^{(b_p)}/c$) were analyzed, and then more complex systems (e.g. a flow shop system and a semiconductor wafer fab) were employed and tested for validation purposes.

In addition to discrete event simulation methodologies (e.g. VRTs and simulation output analysis), this research also utilized several statistical methodologies including: 1) linear and nonlinear regression analysis; 2) D-optimal design of experiments; and 3) a time series forecasting procedure. See Fowler *et al.* (2001), Mackulak *et al.* (2001) and Park *et al.* (2001) for more details.

An integrated overall framework for generating simulation-based CT-TH curves is illustrated in Figure 7. The overall framework takes into account the following system and experimentation characteristics: 1) the complexity of the system; 2) the size of the simulation sample; 3) the process type of the system with respect to batch policies; and 4) the simulation experimenter's preference on the experimental framework. Based on these characteristics, the simulation modeling and experimental decisions are made concerning: 1) the simulation sampling weights; 2) the simulation sampling method; 3) the starting design of the simulation experiments; 4) the sequence of additional simulation design points; and 5) the length of a simulation run.

In Figure 7, associated with the blocks of the starting designs, *EXT* indicates an extreme design point in the design space. *ANC* and *CEN* mean an anchor point and a center point in the design space, respectively. Detailed explanation on the framework shown in Figure 7 is given below.

- *First, simulation sampling strategy:* A simulation experimenter should see whether or not the system concerned is to be regarded as a simple system. If the system might be viewed as a simple system, and the available sample is large enough, then it is recommended to use the simulation sampling strategy with equal sampling weights and CRNs/AVs. When the system is not simple and/or the size of sample is small, it is expected that the simulation sampling strategy with different sampling weights and AVs performs well.
- *Second, starting design*: If the production process contains any batching processes like FB or MBS policy, and thus the CT-TH curve is expected to

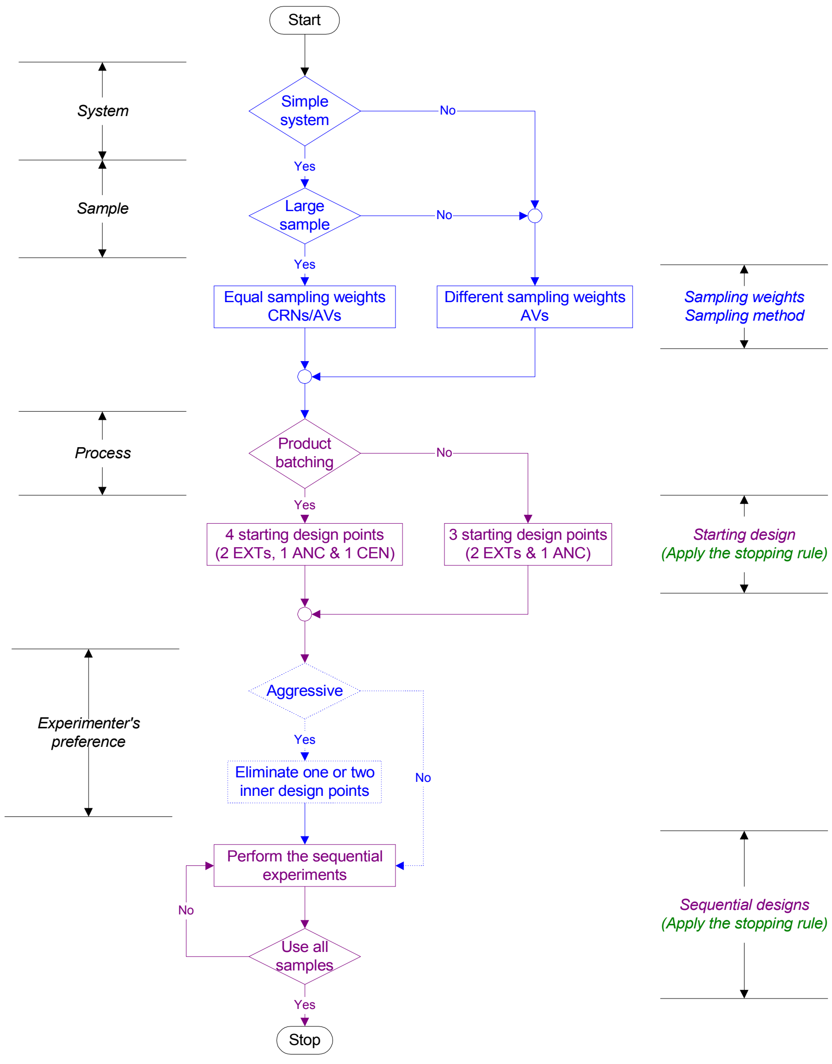

Figure 7: Overall Framework for Generating Simulation-Based CT-TH Curves

form a U-shape, the simulation experimenter might select a starting design with 4 design points. The referential starting design points (0.10, 0.50, 0.90, 0.97) can guide the simulation experimenter on selecting the location of design points to be investigated. It is preferred to choose design points, two for extreme points, one for an anchor point and one for a center point. Meanwhile, if the system experimenter considers a system associated with a monotonically increasing CT-TH curve, he or she can set up a starting design with 3 design points located around (0.10, 0.90, 0.97).

- *Third, sequential designs*: Once sufficient simulation experiments are dedicated to the starting design points, the simulation experimenter can continue to investigate the sequential design points based on the referential sequences. At this time, the experimenter's preference on the simulation experimentation framework could be taken into consideration. If the experimenter wants to obtain more precise and accurate cycle time estimates on the dominant design points (i.e. the anchor, the center or extreme design points), then he or she might eliminate one or two candidate sequential design points and allocate samples more on the survived design points until all samples are exhausted.
- *Fourth, applying the stopping rule*: The sequential stopping rule developed in Phase 3 can be used for every simulation replication not only at the starting design points but also at the sequential design points. The experimenter can adjust the parameters in the algorithm simultaneously to make the sequential stopping rule more (or less) sensitive to the pattern of the time series.

ACKNOWLEDGMENTS

The authors of this paper would like to especially thank Dr. J. Bert Keats, Dr. Dan L. Shunk and Dr. W. Matthew Carlyle for their valuable recommendations and contributions to this research. The second and third authors are partially supported by grant 2001-NJ-878 from the Semiconductor Research Corporation and International SEMATECH. The third author is also partially supported by SRC grant 97-FJ-492 and National Science Foundation grants DMI-9713750 and DMI-0075655.

REFERENCES

Brown, S., Chance, F., Fowler, J. W., and Robinson, J. 1997. A centralized approach to factory simulation. *Future Fab International* 1 (3): 83-86.

Dayhoff, J. E., and Atherton, R. W. 1986. Signature analysis of dispatch schemes in wafer fabrication. *IEEE Transactions on Components, Hybrids, and Manufacturing Technology* CHMT-9 (4): 518-525.

Fowler, J. W., Brown, S., Gold, H., and Schoemig, A. 1997. Measurable improvements in cycle-time-constrained capacity. In *IEEE International Symposium on Semiconductor Manufacturing Conference Proceedings*, A21-A24.

Fowler, J. W., Park, S., Mackulak, G. T., and Shunk, D. L. 2001. Efficient cycle time-throughput curve generation using a fixed sample size procedure. Working paper. Department of Industrial Engineering, Arizona State University, U.S.A. to appear in the *International Journal of Production Research.*

Fromm, H. 1992. Some remarks on cycle time, variability, zero inventories, and costs in microelectronics manufacturing lines. IBM Technical Report TR 28.167, German Manufacturing Technology Center, Sindelfingen.

Gafarian, A. V., Ancker, C. J. Jr., and Morisaku, T. 1978. Evaluation of commonly used rules for detecting "steady state" in computer simulation. *Naval Research Logistics Quarterly* 25: 511-529.

Hopp, W. J., and Spearman, M. L. 1996. *Factory Physics: Foundations of Manufacturing Management*. Chicago: Irwin.

Law, A. M. 1977. Confidence intervals in discrete event simulation: a comparison of replication and batch means. *Naval Research Logistics Quarterly* 24: 667-678.

Mackulak, G. T., Park, S., Keats, J. B., and Fowler, J. W. 2001. A sequential stopping rule for a steady state simulation based on a time series forecasting. Working paper. Department of Industrial Engineering, Arizona State University, U.S.A. submitted to *Transactions of the Society for Computer Simulation International.*

Montgomery, D. C., Johnson, L. A., and Gardiner, J. S. 1990. *Forecasting and Time Series Analysis*. 2nd ed. New York: McGraw-Hill, Inc.

Park, S., 2000. Efficient cycle time-throughput curve generation using discrete event simulation. Doctoral dissertation, Department of Industrial Engineering, Arizona State University, U.S.A.

Park, S., Fowler, J. W., Mackulak, G. T., Keats, J. B., and Carlyle, W. M. 2001. D-optimal sequential experiments for generating a simulation-based cycle time-throughput curve. Working paper. Department of Industrial Engineering, Arizona State University, U.S.A. to appear to *Operations Research.*

Phojanamongkolkij, N. 2000. Analytical models of batch processing for optimal control of semiconductor manufacturing. Doctoral dissertation, Department of Industrial Engineering, Arizona State University, U.S.A.

Wright Williams & Kelly. 1999. *Factory Explorer*™ *User's Guide and Reference Topics* ©.

AUTHOR BIOGRAPHIES

SUNGMIN PARK is a Research Associate of Managerial Research Team at Korea Gas Corporation. He received his Ph.D. from Arizona State University. His research interests are in simulation-based system modeling and analysis, statistical design of simulation experiments and applied statistics. His email address is <sumpark@kogas.or.kr>.

GERALD T. MACKULAK is an Associate Professor of the Department of Industrial Engineering at Arizona State University. He received his Ph.D. from Purdue University. His research interests are in simulation, applied operations research to manufacturing systems and artificial intelligence for virtual manufacturing and production control. His email address is <mackulak@asu.edu>.

JOHN W. FOWLER is an Associate Professor of the Department of Industrial Engineering at Arizona State University. He received his Ph.D. from Texas A&M University. His research interests are in manufacturing systems analysis, discrete event simulation and applied operations research. His email address is <john.fowler@asu.edu>.