Estimating Manufacturing Cycle Time and Throughput in Flow Shops with Process Drift and Inspection

Mandar M. Chincholkar
Intel Corporation
Hillsboro, Oregon

Timothy Burroughs
North Carolina A & T State University
Greensboro, North Carolina

Jeffrey W. Herrmann*
Institute for Systems Research and
Department of Mechanical Engineering
University of Maryland
College Park, Maryland 20742.
jwh2@umd.edu
(301) 405-5433

May 12, 2004

Timothy Burroughs' participation was supported by NSF Grant #0243803, and his conclusions do not necessarily reflect the opinions of the funder.

*Corresponding Author

Abstract

Process drift is a common occurrence in many manufacturing processes where machines become dirty (leading to more contamination) or processing parameters degrade, negatively affecting system performance. Statistical process control tracks process quality to determine when the process has gone out of control (has drifted beyond its specifications). This paper considers the case where parts examined at a downstream inspection station are used to determine when the upstream process is out of control. The manufacturing cycle time from the out of control process to the downstream inspection process influences the detection time that elapses until the out of control process is noticed and repaired. Because an out of control process produces more bad parts, the detection time affects the number of good parts produced and the throughput of the manufacturing system.

This situation is common in many industries but no models of the phenomena exist. This paper presents a novel manufacturing system model based on queueing network approximations for estimating the manufacturing cycle time and throughput of such systems. These are important performance measures since they influence economic measures such as inventory costs and revenue. The model can be used for a variety of system design and analysis tasks. In particular, the model can be used to evaluate the placement of inspection stations in a process flow.

1 Introduction

Process drift is a common occurrence in many manufacturing processes where machines become dirty (leading to more contamination) or processing parameters degrade, negatively affecting system performance. Statistical process control (SPC) tracks process quality to determine when the process has gone out of control (has drifted beyond its specifications). SPC depends upon inspecting the parts produced, measuring critical attributes of the parts, and using these to determine process quality.

Due to variability, all processes produce both good and bad parts. In many cases, a manufacturing process has limited abilities to measure part quality or to discover and discard the bad parts that the process creates. (A “bad” part has unacceptable performance or appearance and will be discarded without being sold.) Thus, the manufacturing system must have inspection stations where human (or automatic) inspectors assess part quality, perform SPC, and discard the bad parts found. The placement of inspection stations in the process flow can have a significant impact on the performance of the manufacturing system, as discussed below.

These bad parts must be identified and discarded. In general, yield is the ratio of the number of good parts produced to the number of parts processed. Some types of flaws are obvious and can be caught immediately, while others require careful examination of trained inspectors using special equipment or procedures. An out of control process produces bad parts at an increased rate. Ideally, an out of control manufacturing process would be detected, halted, and fixed as soon as it went out of control. However, in practice, the delay until the detection increases the number of bad parts created.

The delay occurs because the defective parts discovered downstream are used to determine when the upstream process is out of control. This situation is common in many industries, especially semiconductor manufacturing and electronics assembly. For instance, when a lithography step undergoes a process drift, it will create a flaw that leads to bad parts that must be scrapped. Detecting the drift quickly after it occurs is essential to reduce the number of bad parts.

The manufacturing cycle time from the out of control process to the downstream inspection process influences the detection time that elapses until the out of control process is noticed and repaired. Because an out of control process produces more bad parts, the detection time affects the number of good parts produced and the throughput of the manufacturing system.

Understanding the impact of process drift is important when designing a manufacturing system in which detection times will be significant. Placing inspection stations and choosing equipment can amplify (or reduce) the impact of process drift on system performance. Thus, it is critical to have models that can evaluate system performance in the presence of process drift.

The time that a work order (also known as a job, a batch, or a lot) spends at a workstation from its arrival at the workstation to its completion is known as the *manufacturing cycle time*. The time that a job spends in the manufacturing system between order release and completion is known as the *total manufacturing cycle time*, which is, for a flow shop, the sum of the workstation manufacturing cycle times. (Note that some authors refer to manufacturing cycle time as *throughput time* or *flow time*.)

Reducing the total manufacturing cycle time has many benefits, including lower inventory, reduced costs, faster response to customer orders, and increased flexibility.

Another important performance measure is the throughput of the system. The throughput is the rate at which the system produces good parts. Increasing the throughput yields more sales and increases revenue.

Previous research has examined some of the links between total manufacturing cycle time, throughput, and yield. Srinivasan *et al.* [Srinivasan 95] enumerate benefits of reducing total manufacturing cycle time towards improving system yield for semiconductor manufacture. Their work relates the process yields to deviation of total manufacturing cycle time from its nominal value along with a simulation model to quantify the relationship. Cunningham and Shanthikumar [Cunningham 96] analyze the effects of reducing total manufacturing cycle time on improving die yield of semiconductor wafers. They present two conjectures on how reducing total manufacturing cycle time improves yield. The informational conjecture states that the completed jobs can be studied for defects and improved. The physical conjecture states that a reduced total manufacturing cycle time means lower contamination of completed jobs.

Models of manufacturing systems are useful for obtaining information about a system being designed or modified when it is not possible or desirable to experiment with the real system. This is especially true in manufacturing since the system is a large, complex, and unique operation. Models are needed throughout the manufacturing system life cycle to provide information that is needed to make good decisions about the design and operation of the system.

As mentioned before, the placement of inspection stations in the process flow can have a significant impact on the performance of the manufacturing system. Inspecting parts after each step can prevent bad parts from being processed at subsequent operations. For example, if lithography is producing bad parts, it is wasteful to etch those parts. However, having too many inspection stations is costly and causes delays. Shi and Sandborn [Shi 03] study the problem of placing inspection stations and use optimization to find a solution to minimize the yielded cost (the cumulative cost divided by the final yield). Narhari and Khan [Narhari 96] analyze reentrant manufacturing systems with inspection stations, which may accept parts, reject parts, or send parts to an earlier process to be reworked. Process drift is not considered. They present queueing models for estimating total manufacturing cycle time and throughput, discuss the inspection station placement problem, and provide references to other work on that problem. They present an example that shows how the placement of inspection stations affects throughput.

The existing literature on inspection station placement has not, to the best of our knowledge, addressed the total manufacturing cycle time and throughput of manufacturing systems with process drift. When deciding where to place inspection stations, a firm can use the model presented in this paper to evaluate alternatives and can incorporate the model in an optimization approach. Though optimization is beyond the scope of the current paper, Section 4 uses the model for comparing different locations for an inspection station.

Queueing networks are popular and useful models for manufacturing systems. For more information on queueing network models, see Papadopoulos *et al.* [Papadopoulos 93] and Buzacott and Shanthikumar [Buzacott 93], who present queueing network models for manufacturing systems. Connors *et al.* [Connors 96] modeled semiconductor wafer fabrication facilities using a sophisticated queueing network model to analyze these facilities quickly by avoiding the effort and time needed to create and run simulation models. They present numerical results that show how the queueing network model yields results that are similar to those that a simulation model yields. Queueing network models are also the mathematical foundation of manufacturing system analysis software like rapid modeling [Suri 89]. Koo *et al.* [Koo 95] describe software that integrates a capacity planning model and queueing network approximations. They report that the approximations are reasonable when variability is moderate. Govil and Fu [Govil 99] provide a comprehensive survey of research and software using queueing theory to study manufacturing systems. Herrmann and Chincholkar [Herrmann 02] present a queueing network model for a manufacturing system with no reentrant flow and no process drift.

Despite the extensive work on queueing networks, there exist no models using process drift to relate total manufacturing cycle time, yield, and throughput. This paper describes a novel manufacturing system model for estimating total manufacturing cycle time and throughput of manufacturing systems with process drift and inspection. The model is based on queueing network approximations. To make the presentation more clear, this paper focuses on the single-product case. In addition, all resources have perfect availability, and all resources at a workstation are identical. Chincholkar [Chincholkar 02] presents a more general model for manufacturing systems with multiple products.

This analytical model is able to provide insights into how the manufacturing system parameters (including processing times and arrival rate) impact manufacturing system performance (including total manufacturing cycle time and throughput). In particular, the models shows that counter-intuitive behavior can occur. For instance, increasing the processing time at a workstation increases that workstation's manufacturing cycle time but could reduce the yield (and the throughput) and the total manufacturing cycle time. Also, increasing the arrival rate increases the total manufacturing cycle time but could reduce the yield and throughput.

Experimental results show that the analytical model provides results similar to those of discrete-event simulation. The analytical model requires less data and less computational effort than the simulation model and is therefore more appropriate for situations where a decision-maker needs to compare many scenarios quickly. Thus, the model is a useful tool for comparing the placement of inspection stations.

The remainder of this paper is organized as follows. Section 2 describes process drift and defines the concepts that will be used to develop the mathematical model. Section 3 presents the mathematical model that estimates total manufacturing cycle time and throughput and discusses insights that the model provides. Section 4 describes the results of experiments done to compare the analytical model to a discrete-event simulation model. These demonstrate the impact of processing time changes and operation sequence changes. Section 5 summarizes the paper and highlights important conclusions.

2 System Description

This section defines the concepts that will be used to develop the mathematical model.

Some of the parts that continue from a processing station to the next one have undetected flaws (which will be found at the inspection station). *Normal yield* is the size of the fraction with undetected flaws when the process is operating within its specifications. *Reduced yield* (which is lower than the normal yield) is the size of the fraction with undetected flaws when the process behavior has drifted beyond its specifications.

The rate at which a process goes out of control is the *drift rate* (the average time until the next drift is the reciprocal of the drift rate). The time that the process remains out of control depends upon how long it takes a job to move from that workstation to the downstream inspection station. This time is called the *detection time.* (In some settings, this is called the *metrology delay.*)

When the job is inspected, the drift is noticed (through a statistical process control method), the process is fixed, and the process resumes operating within its specifications.

Process drift affects not only the job that is running when the drift occurs but also every subsequent job until the drift is detected. Clearly, a larger detection time implies that the process will operate out of control (at the reduced yield) for a longer period of time, which increases the number of bad parts with undetected flaws. The process drift cannot be detected until the downstream inspection station completes a job that was completed while the process is out of control. The detection time depends on the manufacturing cycle time at the workstations that follow the process that is out of control.

In some situations, a process drifts continuously. However, because SPC looks for a specific condition and then triggers a repair based upon that condition, the model presented here regards process drift as a distinct event (the event when a certain condition is satisfied). This conceptual view of process drift captures the key characteristics of the problem situation, especially the binary nature of "under control" and "out of control" that underlies SPC.

The mathematical model is based upon these concepts and the following, additional assumptions: Each job visits each workstation exactly once, and all jobs follow the same sequence. (That is, the system is a flow shop with no re-entrant flow.) When a process drift is detected at an inspection station, the workstation is fixed immediately. An inspection station finds (and discards) all of the bad parts in a job. Inspection stations do not undergo process drift. The movement of jobs from one station to the next is instantaneous. The processing times, setup times, and interarrival times are all independent random variables. Note that this model can be extended to manufacturing systems that process multiple products and where resources are not always available (due to failures and repairs). Chincholkar [Chincholkar 02] presents this extended model in detail.

3 Model Development

This section presents the analytical model for estimating the total manufacturing cycle time and throughput of the manufacturing system. The development of this model follows the standard decomposition approach for queueing network approximations [Buzacott 93].

3.1 Data Requirements

The manufacturing system model requires the following data about the product: the part arrival rate (number of parts per time unit of factory operation), the batch size (number of parts) at order release, and the sequence of workstations that each job must visit. The model also requires the following data for each workstation: the number of identical resources available, the mean job setup time and its SCV, the mean part processing time and its SCV, the drift rate, the normal yield, and the reduced yield. The squared coefficient of variation (SCV) of a random variable equals its variance divided by the square of its mean.

The model uses the following notation for the input data:

$$T^a = \text{part arrival rate (parts per time unit)}$$

$$B_0 = \text{batch size (number of parts) at order release}$$

$$c^r = \text{SCV of job interarrival times}$$

$$\begin{aligned} R &= \text{sequence of stations that each job must visit} \\ &= (1, \ldots, q) \\ \mathcal{J} &= \text{the set of processing stations in } R \\ \mathcal{F} &= \text{the set of inspection stations in } R \\ Q_j &= \text{the subsequence of } R \text{ that starts with the station that follows} \\ & \quad\ j \text{ and ends with the first inspection station after } j \end{aligned}$$

$$n_j = \text{number of identical resources at station } j$$

$$t_j = \text{mean part process time at station } j$$

$$c_j^t = \text{SCV of the part process time at station } j$$

$$s_j = \text{mean job setup time at station } j$$

$$c_j^s = \text{SCV of the job setup time at station } j$$

$$\rho_j = \text{drift rate for station } j$$

$$y_j^n = \text{normal yield at station } j$$

$$y_j^r = \text{reduced yield at station } j$$

Note that Q_j is empty if no inspection station follows j in R. Otherwise, Q_j includes exactly one inspection station, which occurs at the end of the subsequence.

3.2 Aggregation and Approximation

Given the input data, the analytical model aggregates the part processing times and job setup times to estimate the mean and SCV of the job processing times at each workstation. Then, it approximates the manufacturing cycle time at each workstation, the total manufacturing cycle time, and the throughput.

Yield. The drift rate ρ_j is the rate at which the process drifts (goes out of control) when the process is operating correctly. $D_j = 1/\rho_j$ is the mean time from the detection (and repair) of one process drift to the occurrence of the next one. (During this interval the process yield equals y_j^n.) The detection time DT_j is the expected delay from the occurence of a process drift that occurs at station j to the detection of the process drift. (During this interval the process yield equals y_j^r.) The detection time equals the sum of the manufacturing cycle time at the workstations that follow the processing station up to and including the next inspection station:

$$DT_j = \sum_{g \in Q_j} CT_g^{\star}; \quad \forall j \in \mathcal{J} \tag{1}$$

For station j, the *average yield* z_j is the time-weighted average of the normal and reduced yields. If there is no process drift at station j, $\rho_j = 0$ and $z_j = y_j^n$.

$$z_j = \frac{D_j y_j^n + DT_j y_j^r}{D_j + DT_j} \tag{2}$$

The average batch size at a workstation is influenced by the yields of the preceding operations. Let B_j be the average batch size of jobs that leave station j. The number of parts in a job remains the same at processing stations. A job starts with all good parts, and the number of good parts in a job changes as it is processed at processing workstations. The cumulative effect is measured by the *cumulative yield* Z_j, which is the average fraction of good parts in jobs that leave station j.

$$B_j = B_{j-1} \forall j \in \mathcal{J} \tag{3}$$

$$Z_0 = 1 \tag{4}$$

$$Z_j = Z_{j-1} z_j \forall j \in \mathcal{J} \tag{5}$$

Since inspection stations discard bad parts, the batch size of a job changes at the inspection stations, and the average batch size equals the average number of good parts that arrive. The jobs leaving an inspection station have all good parts.

$$B_j = B_{j-1} Z_{j-1} \, \forall j \in \mathcal{F} \quad (6)$$

$$Z_j = 1 \, \forall j \in \mathcal{F} \quad (7)$$

Processing Times. The mean job processing time at station j, which is the sum of the part processing times and the job setup time, is t_j^+:

$$t_j^+ = B_{j-1} t_j + s_j \quad (8)$$

Note that, at station j, the variance of the part processing times equals $t_j^2 c_j^t$ and the variance of the job setup time is $s_j^2 c_j^s$. The variance of the job process time (the sum of the variance of the part processing times and the variance of the job setup time) divided by the square of the mean job processing time is c_j^+, the SCV of the job processing time:

$$c_j^+ = \frac{B_{j-1} t_j^2 c_j^t + s_j^2 c_j^s}{(t_j^+)^2} \quad (9)$$

Arrival and Departure Processes. The job arrival rate x (jobs per time unit) is the part arrival rate divided by the initial batch size: $x = T^a / B_0$. At the first station, the arrival variability (the SCV of the job interarrival times) depends upon the external job arrivals. That is, $c_1^a = c^r$.

For station j, the departure variability c_j^d depends upon c_j^a, the arrival variability, and c_j^+, the variability (SCV) of the job process times. The following is a commonly used estimate for departure variability [Hopp 01]:

$$c_j^d = 1 + \frac{u_j^2}{\sqrt{n_j}}(c_j^+ - 1) + (1 - u_j^2)(c_j^a - 1) \quad (10)$$

The arrival variability c_j^a at the subsequent stations depends upon the departure variability at the previous workstation. Thus, for $2 \leq j \leq q$, $c_j^a = c_{j-1}^d$.

Performance Measures. The performance measures of interest are $CT_j^\star$, the manufacturing cycle time at each workstation; CT, the total manufacturing cycle time; and T^o, the throughput (parts per time unit). Another important quantity is u_j, the utilization of the resources at station j.

The model considers each workstation as a GI/G/m queueing system. The manufacturing cycle time at each workstation is the sum of the average waiting time in the queue plus the average job processing time. (See [Whitt 93] for more information about the approximation.) The total manufacturing cycle time is the sum of the workstation manufacturing cycle times.

$$
\begin{aligned}
u_j &= \frac{t_j^+ x}{n_j} && (11) \\
CT_j^\star &= \frac{1}{2}(c_j^a + c_j^+)\frac{u_j^{(\sqrt{2n_j+2}-1)}}{n_j(1-u_j)}t_j^+ + t_j^+ && (12) \\
CT &= \sum_{j \in R} CT_j^\star && (13)
\end{aligned}
$$

The throughput T^o is the job arrival rate multiplied by the average batch size after the last workstation (workstation q). Note that the throughput is less than the part arrival rate due to yield losses, which are affected by the process drift and detection time.

$$
T^o = xB_q = \frac{B_q}{B_0}T^a \qquad (14)
$$

3.3 Sensitivity Analysis

This analytical model allows one to evaluate how system performance (both manufacturing cycle time and throughput) changes when parameters such as the processing time or arrival rate change.

3.3.1 Processing Time

Increasing the part processing time at a workstation increases the manufacturing cycle time at that workstation. In a system with process drift, this increase may, in turn, reduce the throughput and the total manufacturing cycle time.

This can occur because the manufacturing cycle time increase consequently delays the detection of process drift at a preceding workstation (its detection time increases). As a result, the average yield decreases, the cumulative yield decreases, and the batch size leaving the next inspection station is smaller. The reduced batch size decreases the job processing times, utilization, and manufacturing cycle time at the workstations following the inspection station.

Consider, for example, the production line with three stations shown in Figure 1. Suppose we have the following conditions:

1. $R = (1, 2, 3)$. $\mathcal{J} = \{1, 3\}$. $\mathcal{F} = \{2\}$.

2. $s_j = 0$ and $n_j = 1$ for $j = 1, 2, 3$.

3. $c^r = 1$, $\rho_1 = \rho$, $D_1 = 1/\rho$.

4. $\rho_3 = 0$, and $z_3 = y_3^n = 1$.

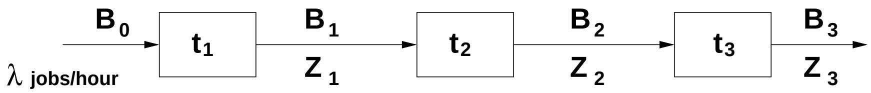

Figure 1: A three station line

Let $x = \lambda = T^a/B_0$. Since $B_1 = B_0$, utilizations may be calculated as follows:

$$u_1 = \lambda B_1 t_1 \quad (15)$$

$$u_2 = \lambda B_1 t_2 \quad (16)$$

$$u_3 = \lambda B_2 t_3 \quad (17)$$

To simplify the sensitivity analysis, we will use the M/M/1 queueing model to approximate the workstation manufacturing cycle times. This leads to the following results:

$$CT_1^\star = \frac{B_1 t_1}{1-u_1} \quad (18)$$

$$CT_2^\star = \frac{B_1 t_2}{1-u_2} = \frac{B_1 t_2}{1-\lambda B_1 t_2} \quad (19)$$

$$CT_3^\star = \frac{B_2 t_3}{1-u_3} = \frac{B_2 t_3}{1-\lambda B_2 t_3} \quad (20)$$

Because $DT_1 = CT_2^\star$,

$$B_2 = B_1 Z_1 = B_1 z_1 = B_1 \frac{D_1 y_1^n + CT_2^\star y_1^r}{D_1 + CT_2^\star} \quad (21)$$

Thus, as t_2 increases, u_2 and $CT_2^\star$ increase. Because $y_1^r < y_1^n$, z_1 decreases. Consequently, Z_1, B_2, u_3, and $CT_3^\star$ decrease. Because $B_3 = B_2$ and $T^o = \lambda B_3 = \lambda B_2$, the throughput T^o also decreases.

The total manufacturing cycle time $CT = CT_1^\star + CT_2^\star + CT_3^\star$. Differentiating with respect to t_2,

$$\frac{dCT}{dt_2} = 0 + \frac{dCT_2^\star}{dt_2} + \frac{dCT_3^\star}{dt_2} \tag{22}$$

$$\frac{dCT_2^\star}{dt_2} = \frac{B_1}{(1-u_2)^2} \tag{23}$$

$$\frac{dCT_3^\star}{dt_2} = \frac{dB_2}{dt_2}\frac{t_3}{(1-\lambda B_2 t_3)} + \frac{\lambda B_2 t_3^2}{(1-\lambda B_2 t_3)^2}\frac{dB_2}{dt_2} \tag{24}$$

$$\frac{dB_2}{dt_2} = B_1 y_1^r \frac{dCT_2^\star}{dt_2}\frac{1}{D_1 + CT_2^\star} - B_1 \frac{D_1 y_1^n + CT_2^\star y_1^r}{(D_1 + CT_2^\star)^2}\frac{dCT_2^\star}{dt_2} \tag{25}$$

$$= \frac{B_1^2 D_1 (y_1^r - y_1^n)}{(1-u_2)^2 (D_1 + CT_2^\star)^2} \tag{26}$$

Note that, because $y_1^r < y_1^n$, $\frac{dB_2}{dt_2} < 0$. Moreover, $\frac{dT^o}{dt_2} = \lambda \frac{dB_2}{dt_2} < 0$.

$$\frac{dCT}{dt_2} = \frac{B_1}{(1-u_2)^2} + \frac{B_1^2 D_1 (y_1^r - y_1^n)}{(1-u_2)^2 (D_1 + CT_2^\star)^2}\frac{t_3}{(1-u_3)^2} \tag{27}$$

Substituting Equation 19 for $CT_2^\star$ and rearranging terms, we can show that $dCT/dt_2 <$ 0 if and only if

$$\frac{B_1 D_1 t_3 (y_1^n - y_1^r)}{(1-u_3)^2} > \left(D_1 + \frac{B_1 t_2}{1-u_2}\right)^2 \tag{28}$$

Equation 28 represents the condition under which increasing the processing time at the second workstation reduces the total manufacturing cycle time. Figure 2 illustrates this phenomenon for a particular example, and Table 1 lists the system parameters not given above. This phenomenon has been observed in other systems such as polling systems. (For more about polling systems, see Takagi [Takagi 86].)

3.3.2 Arrival Rate

Now consider the impact of increasing the part arrival rate on both the total manufacturing cycle time and the throughput. In general, a larger arrival rate increases the

Variable	Value
Batch size B_0 (parts/job)	100
Arrival rate T^a (parts/time unit)	4
Interarrival SCV c^r	1
Drift rate ρ_1 (time units^{-1})	0.001
Part process time t_1 (time units)	0.2
Part process time t_3 (time units)	0.27
Normal yield y_1^n	0.9
Reduced yield y_1^r	0.5

Table 1: Specifications for the three station example

utilization and manufacturing cycle time at a workstation. However, as discussed in the previous section, increasing the manufacturing cycle time at one workstation can increase a previous workstation's detection time, which decreases the average yield. This in turn can reduce the throughput and the total manufacturing cycle time.

For example, consider the production line with three stations shown in Figure 1. A larger arrival rate increases the utilization and manufacturing cycle time at the first two workstations. This increases the first station's detection time, so the average yield of the first station and the final batch size B_3 decrease. Throughput does not increase proportionally to arrival rate. As the part arrival rate continues to increase and the detection time gets even larger, the throughput can even decrease.

This can be demonstrated numerically with the example presented above (with $t_2 = 0.24$ time units). As shown in Table 2, the models presented in Section 3.2 predict that the overall yield decreases and the throughput declines when the arrival rate increases from 4.0 to 4.1 parts per time unit. It is possible to prove analytically that, under certain conditions, increasing the part arrival rate T^a leads to a decrease in the throughput T^o.

T^a (parts/time unit)	3.9	4.0	4.1
CT_1^* (time units)	56	60	66
CT_2^* (time units)	95	132	278
CT_3^* (time units)	32	30	25
CT (time units)	183	222	369
T^o (parts/time unit)	3.37	3.41	3.33
Yield $T^o/T^a = B_3/B_0$	0.87	0.85	0.81

Table 2: Impact of arrival rate on total manufacturing cycle time and throughtput

4 Experimental Results

The analytical model presented in Section 3.2 was evaluated by comparing its results to those from a discrete event simulation model. This section explains the experiments performed.

The purpose of the experiments was to compare the results that the analytical model produces to the results that a discrete-event simulation model produces. In particular, the experiments studied how increasing the detection time reduces throughput and how changing the placement of an inspection station affects throughput. A comparison between the analytical and experimental results is useful to demonstrate the validity of the analytical model approximation. While the assumptions of the analytical model are necessary to derive that model, manufacturing systems will not satisfy all of them in practice. For example, the analytical model assumes that that the number of parts in a job decreases deterministically (based on the yield). However, in practice, some jobs will have more good parts and some less. A simulation model can represent this batch size variability. In addition, a simulation can capture the variation in drift times and detection times. Thus, another purpose of the simulation study is to determine the robustness of the analytical model when these assumptions do not hold.

To avoid unnecessary complexity, the experiments used simulation models and analytical approximation models of a three station manufacturing system that has a processing workstation, then an inspection workstation, and finally a processing workstation. It is important to note that the experiments used the approximation model presented in Section 3.2, not the equations used in Section 3.3.1 for discussing sensitivity analysis.

There were three sets of scenarios, and each set contained sixteen scenarios. Most of the parameters were set to fixed values. To create the baseline scenarios in the first set, the experiments varied four parameters to capture the impact of those parameters and to compare the models' performance across a range of scenarios. The first set of scenarios is the baseline set (Scenarios 1A to 16A). Each scenario in the second set of scenarios (Scenarios 1B to 16B) varies the corresponding baseline scenario by increasing the per-part inspection time. Each scenario in the third set of scenarios (Scenarios 1C to 16C) varies the corresponding baseline scenario by switching the sequence of the last two workstations. That is, the inspection workstation becomes the last workstation.

4.1 Simulating Process Drift

The simulation model represents the process drift of the first workstation as a distinct and independent random process. When the workstation begins operating normally, the model samples from an exponential distribution with mean $D_1 = 1/\rho_1$. This value is the time until the next drift event. When this event occurs, the workstation is out of control. Any jobs that begin processing after this event will have more bad (but undetected) parts. When the first of these "contaminated" jobs reaches the downstream inspection station and completes the inspection process, the drift is detected, and the workstation resumes operating normally. Any jobs that begin processing after this event will have fewer bad (but undetected) parts.

At the first workstation, the number of good parts in a job has a binomial distribution with the probability of success equal to the yield (either the normal yield or the reduced yield) and the number of trials equal to the initial batch size.

4.2 Experiments

As mentioned before, some input parameters of the model remained fixed in all scenarios, while others were changed. The following parameters were fixed: $T^a = 4$ parts per time unit. $B_0 = 20$ parts. $n_j = 1$, $c_j^t = 1$, and $s_j = 0$ for $j = 1, 2, 3$. $y_1^n = 0.9$. $y_3^n = 1$, and $\rho_3 = 0$.

Table 3 shows the values for each of the four parameters that were changed to create the baseline scenarios.

The job interarrival times were exponentially distributed with a mean of five time units. The drift interarrival times were exponentially distributed. The part processing times (at each workstation) were exponentially distributed. Thus, the job processing times (which are the sum of the part processing times) had a gamma distribution that depended upon the number of parts in the job when the job was processed.

For Scenarios 1A to 16A, $t_2 = 0.18$ time units. For Scenarios 1B to 16B, $t_2 = 0.24$ time units. For Scenarios 1C to 16C, $t_2 = 0.18$ time units, but the inspection station was the third station, not the second. For each of the 48 scenarios, ten replications of the simulation model were run. Each replication ran for 25,000 time units each, with no warm-up period. The simulation model was constructed using Arena©. (Arena is a registered trademark of Rockwell Automation.) Running ten replications of the model took approximately one minute on a personal computer. (Note that the simulation models are available from the following URL: http://www.isr.umd.edu/Labs/CIM/projects/dfp/index.html.)

Variable	Values
Drift rate ρ_1	0.01, 0.001
Reduced yield y_1^r	0.5, 0.8
Part processing time t_1	0.18, 0.24
Part processing time t_3	0.18, 0.24

Table 3: Scenarios for model comparison

4.3 Results

The most significant feature of the analytical model is the ability to capture the impact of manufacturing cycle time on the throughput of the manufacturing system. The following results compare how the analytical and simulation models estimate this impact. The simulation results represent a 95% confidence interval.

First, consider Scenarios 1A to 16A and Scenarios 1B to 16B. Then, consider Scenarios 1A to 16A and Scenarios 1C to 16C. In each case, the 32 scenarios can be viewed as 16 pairs of scenarios.

In the first case, $t_2 = 0.18$ in the first scenario of each pair, and $t_2 = 0.24$ in the second scenario. Figure 3 shows the decrease in throughput for each pair. Table 4 shows the numerical results for the pairs with the most change in throughput. In the second case, the inspection station is immediately after the first process in the first scenario of each pair, while the inspection station is last in the second scenario. (This increases the detection time for the first process.) Figure 4 shows the decrease in throughput for each pair. Table 5 shows the numerical results for the pairs with the most change in throughput. In general, the analytical model is able to estimate the change in throughput very well.

Scenario	Predicted Throughput	Lower Limit	Upper Limit
3A	3.507	3.389	3.455
3B	3.167	3.101	3.153
4A	3.577	3.525	3.594
4B	3.492	3.424	3.486
7A	3.507	3.381	3.455
7B	3.169	3.046	3.136
8A	3.577	3.566	3.626
8B	3.492	3.465	3.497
11A	3.532	3.364	3.472
11B	3.393	3.201	3.281
15A	3.532	3.405	3.504
15B	3.393	3.252	3.304
16A	3.583	3.517	3.576
16B	3.548	3.489	3.548

Table 4: Change in throughput when t_2 increases

5 Summary and Conclusions

This paper presented an analytical model for estimating the performance of a manufacturing system with process drift and inspection. In particular, the manufacturing system is a flow shop that produces a single product. This analytical model is able to

Scenario	Predicted Throughput	Lower Limit	Upper Limit
7A	3.507	3.381	3.455
7C	3.134	3.015	3.107
8A	3.577	3.566	3.626
8C	3.483	3.441	3.482
15A	3.532	3.405	3.504
15C	3.345	3.208	3.286

Table 5: Change in throughput when inspection station is last

provide insights into how the manufacturing system parameters (including processing times, arrival rate, and placement of an inspection station) affect manufacturing system performance (including total manufacturing cycle time and throughput). An especially important result is that increasing manufacturing cycle time at one workstation can reduce both total manufacturing cycle time and throughput. This shows that these performance measures have a complex relationship in systems with process drift and inspection.

The queueing network approximations offer some advantages and also have limitations. Compared to simulation models or more sophisticated queueing network analysis techniques, these approximations are less accurate, especially for very complex systems, and cannot provide the same range of performance measures. However, they require less data and less computational effort than the simulation models and other analysis techniques. Therefore, they are more appropriate for situations where a decision-maker needs to compare many scenarios quickly. Ultimately, they are sufficiently detailed to offer insight into the behavior of the phenomena being studied.

The models presented here have many potential uses, including evaluating the system-level performance of alternative system designs that place inspection stations at different points in the processing sequence (as discussed in Section 4) and alternative system designs that use different equipment with different drift rates and yields.

Acknowledgements

This research was supported by the National Science Foundation under grants DMI 97-13718 and EEC 02-43803 and was conducted in the facilities of the Computer Integrated Manufacturing Laboratory, a constituent lab of the Institute for Systems Research. Jeffrey W. Herrmann has a joint appointment with the Department of Mechanical Engineering and the Institute for Systems Research. The authors appreciate the help provided by Ms. Sara Hewitt, who conducted preliminary simulation experiments and created the Learning Historian, a simulation experimentation tool.

References

[Buzacott 93] Buzacott, J.A., and J.G. Shanthikumar (1993) *Stochastic Models of Manufacturing Systems*, Prentice-Hall, Englewood Cliffs, New Jersey.

[Chincholkar 02] Chincholkar, Mandar (2002) *Design for Production: Using Manufacturing Cycle Time Information to Improve Product Development*, PhD Dissertation, University of Maryland, College Park, USA.

[Connors 96] Connors, D.P., G.E. Feigin, and D.D. Yao (1996) A Queueing Network Model for Semiconductor Manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, **9**, no.3.

[Cunningham 96] Cunningham, S.P., and J.G. Shanthikumar (1996) Empirical Results on the Relationship Between Die Yield and Cycle Time in Semiconductor Wafer Fabrication. *IEEE Transactions on Semiconductor Manufacturing*, **9**, no.2.

[Govil 99] Govil, Manish K., and Michael C. Fu (1999) Queueing theory in manufacturing: a survey. *Journal of Manufacturing Systems*, **18**, no.3, 214-240.

[Hopp 01] Hopp, Wallace J., and Mark L. Spearman (2001) *Factory Physics: Foundations of Manufacturing Management*, second edition, Irwin McGraw-Hill, Boston.

[Herrmann 02] Herrmann, Jeffrey W., and Mandar Chincholkar (2001/2002) Reducing Manufacturing Cycle Time during Product Design. *Journal of Manufacturing Systems*, **20**, no.6, 416-428.

[Koo 95] Koo, P.-H., C.L. Moodie, and J.J. Talavage (1995) A spreadsheet model approach for integrating static capacity planning and stochastic queueing models. *International Journal of Production Research*, **33**, no.5, 1369-1385.

[Narhari 96] Narhari, Y., and L.M. Khan (1996) Modeling Reentrant Manufacturing Systems with Inspection Stations. *Journal of Manufacturing Systems*, **15**, no.6, 367-378.

[Papadopoulos 93] Papadopoulos, H.T., C. Heavey, and J. Browne (1993) *Queueing Theory in Manufacturing Systems Analysis and Design*, Chapman and Hall, London.

[Shi 03] Shi, Zhen, and Peter Sandborn (2003) Optimization of Test, Diagnosis, Rework Locations and Characteristics in Electronic Systems Assembly Using Real-Coded Genetic Algorithms, *Proceedings of the International Test Conference*, October 2003.

[Srinivasan 95] Srinivasan, K., R. Sandell, and S. Brown (1995) Correlation Between Yield and Waiting Time: A Quantitative Study, in *Transactions of 1995 IEEE/CPMT International Electronics Manufacturing Technology Symposium.*

[Suri 89] Suri, Rajan (1989) Lead time reduction through rapid modeling. *Manufacturing Systems*, 66-68.

[Takagi 86] Takagi, Hideaki (1986) *Analysis of Polling Systems*, MIT Press, Cambridge, Massachusetts.

[Whitt 93] Whitt, W. (1993) Approximating the GI/G/m queue. *Production and Operations Management*, Volume 2, Number 2, pages 114-161.

Author's Biographies

Mandar M. Chincholkar earned his Ph.D. in mechanical engineering from the University of Maryland in 2002. He graduated with a master's degree in mechanical engineering (design engineering) from the Indian Institute of Technology, Bombay. He currently works for the Logic Technology Development Group at Intel Corp, in Hillsboro, Oregon. His research interests include computer aided design and manufacturing, integrated product and manufacturing systems design and geometric modeling.

Timothy Burroughs is a student at North Carolina A & T State University. He participated in a Research Experience for Undergraduates program sponsored by the Institute for Systems Research at the University of Maryland.

Jeffrey W. Herrmann is an associate professor with a joint appointment in the Department of Mechanical Engineering and the Institute for Systems Research at the University of Maryland. He is the director of the Computer Integrated Manufacturing Laboratory. He earned his B.S. in applied mathematics from Georgia Institute of Technology. As a National Science Foundation Graduate Research Fellow from 1990 to 1993, he received his Ph.D. in industrial and systems engineering from the University of Florida. His publications cover topics in process planning, production scheduling, design for manufacturing, and manufacturing facility design. His current research interests include design for production, manufacturing system design, and decision-making in product development.

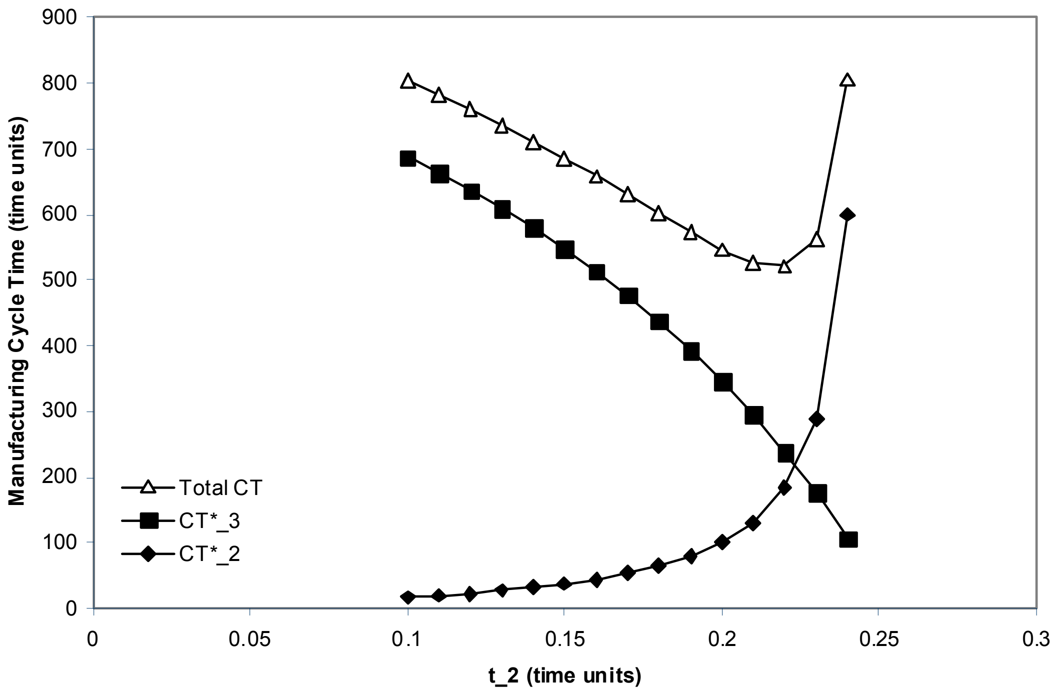

Figure 2. Manufacturing Cycle Time versus Part Processing Time.

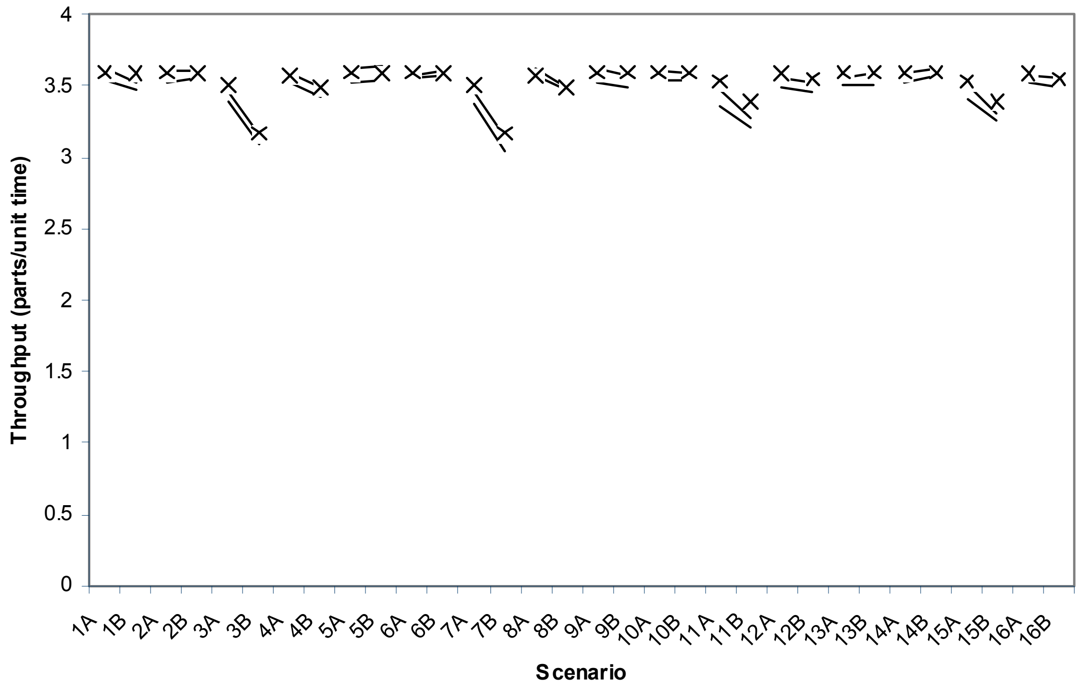

Figure 3. Throughput for Scenarios 1A to 16A and 1B to 16B.
(For each scenario, the "X" is the predicted value,
and the lines are the bounds on the 95% confidence interval.)

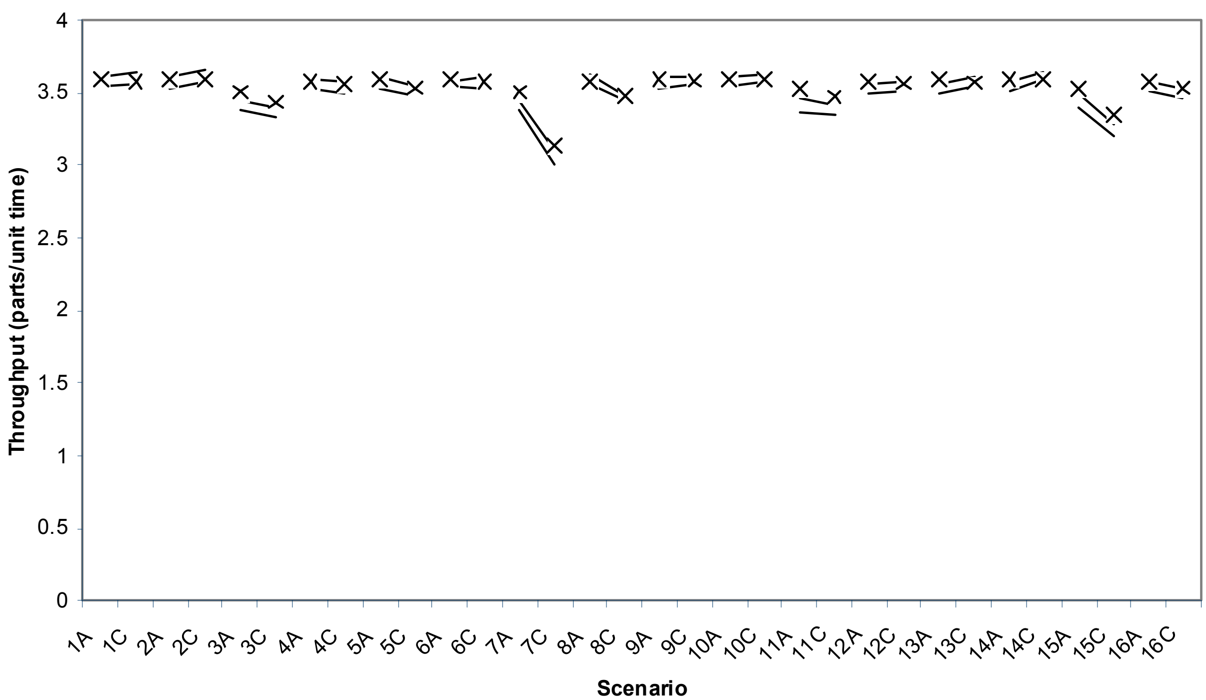

Figure 4. Throughput for Scenarios 1A to 16A and 1C to 16C.
(For each scenario, the "X" is the predicted value,
and the lines are the bounds on the 95% confidence interval.)