Proceedings of the 1996 Winter Simulation Conference
ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain

EVALUATION OF LOT RELEASE POLICIES FOR SEMICONDUCTOR MANUFACTURING SYSTEMS

Raka Sandell

MOS12 Die Manufacturing
Semiconductor Products Sector
Motorola Inc.
1300, North Alma School Road,
Chandler, Arizona- 85224, U.S.A

Krishna Srinivasan

Worldwide Enterprise Modeling Group
Motorola Manufacturing Systems
Motorola Inc.
1299, E. Algonquin Road
Schaumburg, Illinois 60196, U.S.A.

ABSTRACT

This paper describes a simulation based approach to evaluate and benchmark various lot release strategies in semiconductor manufacturing systems. A full-factorial experiment, with 5 factors, has been designed and executed for this purpose. A variation of the Bonferroni inequality is used to compute confidence intervals for differences in system performance for different operating strategies. The results show that no lot release policy dominates across all scenarios.

1 INTRODUCTION

Semiconductor manufacturing is generally characterized by investments in process technology which are over a billion dollars. The industry is characterized by extremely short product life-cycles, frequently decreasing profit margins and intense competition. In such a volatile scenario, maintaining a competitive advantage and remaining profitable, in operational terms requires minimization of cycle time and work-in-process inventory and the maximization of throughput. Cycle time is the time it takes for a product to go from start to finish. Cycle time reduction is important as it leads to a shorter order lead time and hence better customer satisfaction. Also, for the same level of throughput, a shorter cycle time results in a smaller work-in-process inventory that not only reduces the capital tied up, but also minimizes the inventory buffer that needs to be maintained at the downstream end of the plant. When the market demand changes and product designs become obsolete, such inventory may lose value. There is also a technological reason for reducing the cycle time. The shorter the period that the wafers are exposed to aerial contaminants while waiting for processing, the smaller is the yield loss. It is also important to reduce the variance of cycle time. Low variability in cycle time allows a more accurate prediction of production completion time and ability to meet due dates. This also facilitates improved downstream coordination of operations on completed wafers such as wafer probe, assembly and test.

The performance measures of interest for a product generally depend on the nature of business one is in. For make-to-order parts as in ASIC (Application Specific Integrated Circuit) factories, adhering to due dates is of paramount importance. It is also important to minimize the mean cycle time. Hence, a performance measure like mean + 3 × standard deviation, also known as 98% cycle time, is of great significance. For a make-to-market environment, like Dynamic Random Access Memory (DRAM) chip manufacturing, typical objectives are maximization of throughput while maintaining cycle time goals. For extremely expensive chips like microprocessors (pentium, power PC etc.), yield is an important criterion. Thus in all types of industries, the reduction of cycle time assumes great significance.

Different types of control policies are implemented to ensure that the factory's cycle time is minimized under different operating characteristics. Some examples of such policies include determining the instants of time when wafers or lots are to be brought into the factory, deciding how operators choose parts for processing given that there are many wafers to choose from, and determining the optimal schedules for machine maintenance. Certain rules are formally defined in order to address such operational problems in wafer fabrication facilities. These rules are referred to as scheduling policies.

In general, the scheduling policies used to exercise control over fab operations are of two types. In the first type, one can specify rules for determining when new lots are to be released into the plant. This is referred to as the release policy. Clearly, the release policy must meet some constraints such as maintaining an average release rate of lots. Examples of lot release policies are deterministic release and closed-loop release which are explained in subsequent sections. Second, for lots already in the plant, one has to decide which lot is to be

processed next at each machine as it becomes available. These are referred to as dispatching policies. Such policies determine the sequence in which jobs in front of a machine are processed by the machine. Examples of such policies are Shortest Processing Time (SPT) and Earliest Due Date (EDD). The dispatching policy need not be a global rule i.e., the dispatching rule at different workstations may be different.

Many papers in literature have compared the cycle time performance of lot release and dispatching policies. Most of them have concluded that lot release policies have a much larger impact on fab performance than dispatching policies (Wein(1988), Glassey and Resende(1988a, 1988b)). In addition, these papers have concluded that one policy generally dominates over other release policies investigated. For example, Spearman and Zazanis(1992) conclude that the CONWIP release policy is more superior than open-loop release policies. Wein(1988) shows that the workload regulating rule is much superior to deterministic and Poisson release policies. In this study, we adopt the premise that lot release policies have a much larger impact than dispatching policies on fab performance. While earlier authors have tested the quality of policies on one manufacturing facility, this study attempts to provide a comparison of release policies across vastly differing manufacturing facilities. The study explores whether conclusions derived by other authors can be extended and held to be true under all types of manufacturing scenarios. Preliminary results showed that under different circumstances, the performance of policies tend to vary greatly. Hence, a full factorial experiment is constructed and executed to test the impact of common lot release policies on the cycle time performance of vastly differing manufacturing facilities.

As this study will strive to provide a quantitative comparison of the effectiveness of different lot release strategies, a basis for comparison has to be defined *a priori*. As described earlier, the objective of most scheduling policies for semiconductor manufacturing is the minimization of mean cycle time and average WIP and the maximization of throughput. Glassey and Resende(1988a) suggest guidelines for the comparison of scheduling policies. Scheduling policies can be compared based on a tradeoff curve plotted between the mean cycle time and mean throughput for a system. Figure 1 describes the cycle time D(t) as a function of the fab throughput t. It is a well known fact that as throughput approaches the capacity of the fab (i.e. the utilization of machines becomes very high) the average queuing time (cycle time) approaches infinity. As shown in Figure 1, scheduling policy SA is superior to policy SB for a given throughput t if DA(t) < DB (t), where DA (t) and DB (t) are the tradeoff curves for policies SA and SB, respectively.

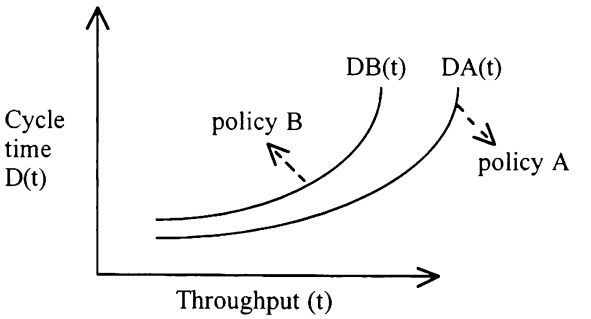

Figure 1: Scheduling policies: Basis for comparison

The experiment described in subsequent sections will use the above described scheme for testing release policies. The scheduling policies will be tested under different throughput rates for each of the system configurations in order to make the above comparison. A variation of the Bonferroni inequality will be used to create confidence intervals for the difference between the policies.

The paper is organized in the following manner. Section 2 provides an overview of past work in this area. Section 3 elaborates on the experimental design and the factors used in the experiment. Section 4 describes the use of a modified version of the Bonferroni inequality to compare release policies. The results are discussed in Section 5 and analysis of release policies for an actual system and conclusions are provided in Sections 6 ad 7 respectively.

2 LITERATURE SURVEY AND DEFINITIONS

This section describes recent forays in the area of dispatching and lot release policies for semiconductor manufacturing systems. Wein(1988) is concerned with assessing the impact that scheduling can have on the performance of semiconductor wafer fabrication facilities. The performance measure considered here is the mean cycle time for a lot of wafers. Several release policies and dispatching rules have been evaluated using a simulation model of a fictitious, single product semiconductor facility. Three versions of the model, which differ by the number of servers present at particular stations are used. Four different types of release policies were considered.

The first release policy evaluated is the deterministic input policy. Given the total number of wafers that need to be processed in a week, the average time between the entry of wafers of equivalent quantities is obtained from the ratio between the total time available and the number of wafers to be processed. In other words, if the total available time in a week is Z units and the total number of wafers that need to be processed is N, then the average

inter-arrival time between wafers is $(Z \div N)$ units. For the deterministic release policy, the inter-arrival times are sampled from a constant distribution. The second type of release is the Poisson release policy. For the Poisson release, the inter-arrival times are sampled from an exponential distribution with mean $(Z \div N)$. The third release policy studied is the closed loop input policy where the number of lots in the system is held constant. The departure of a job/wafer from the system immediately triggers the arrival of a new job into the system. Here the throughput/ wafers started is controlled by the maximum WIP permitted in the system. The fourth release policy is the workload regulating input policy, which releases wafers in lots into the system whenever the total amount of work in the system for any bottleneck falls below a pre-specified level.

Wein(1988) showed that the workload regulating policy dominates the other three policies. As mentioned before, these results were applicable to a single product fictitious facility where the number of servers were perturbed and hence, by no means the experiment tested the robustness of the applicability of the result to other manufacturing configurations.

Glassey and Resende(1988a, 1988b) also carried out comparisons of the four release policies and another policy called the starvation avoidance policy. Based on experiments carried out with the fictitious one-product facility used by Wein(1988) and based on perturbations to the number of servers, they concluded that the starvation avoidance policy was superior to the other four policies. Here again, the authors had not tested the robustness of the result to alternate manufacturing configurations. Similar results were observed in Lu, Ramasamy and Kumar (1992).

Spearman and Zazanis(1992) demonstrated, through analytical means the dominance of the CONWIP release policy over Poisson release policy for a network of tandem exponential queues. The network comprises a set of machines where the process occurs in tandem (as in an assembly line) and each of the machines that comprise the network are single server systems where the service time is assumed to be sampled from an exponential distribution.

These results presented in the literature survey have not been clearly understood by the practicing community. Even though the workload regulating policy and the starvation avoidance policies have been shown to be quite effective in improving performance metrics, practitioners have found them impractical to implement them on the factory floor. This is primarily due to the difficulties involved in the detection of the bottleneck on a dynamic basis in a large sized factory. The bottleneck tends to shift constantly as a function of various random events and fluctuations in the product- mix. Therefore, policies such as workload regulation and starvation avoidance, which are based on bottleneck activity are not examined in great detail in this study. Instead, this study concentrates on simplistic policies like deterministic input, CONWIP and Poisson release which are easily implementable on the factory floor.

Among the three release policies described above - Poisson, deterministic and CONWIP, there lies a great deal of confusion with the understanding of the operation and overall relative effectiveness of CONWIP type closed-loop policies for different factory configurations. Many fab managers tend to assume on the basis of results reported by Spearman and Zazanis(1992) for tandem manufacturing lines that CONWIP is the panacea for all operational problems faced in their factories. In many cases, the authors have observed that the results presented by Spearman and Zazanis(1992) have been misinterpreted as a generalized proof of the superiority of CONWIP over all release policies for all types of performance measures. This contradicts results obtained by Wein(1988) and Glassey and Resende(1988a, 1988b) where, the CONWIP policy has been shown to be inferior to an open-loop policy like the deterministic release policy where the inferiority of a policy was measured based on the amount of increase in mean cycle time with an increase in average throughput. This study will attempt to resolve this confusion and make recommendations with respect to the relative performance of these release policies for different manufacturing configurations.

This study will show that one release policy does not dominate others for all types of manufacturing systems. In fact, the results that have been reported in all these studies have been based on different types of manufacturing systems. While those reported by Wein(1988) and Glassey and Resende(1988a, 1988b) are for a fictitious one-product semiconductor manufacturing facility, the results reported by Spearman and Zazanis(1992) are based on tandem (assembly line) manufacturing systems and hence might not apply to complex systems like semiconductor manufacturing systems. Such disparities in the base assumptions, the concomitant variations in the conclusions and the resulting confusion in the practicing community calls for a detailed experimental design that tests the results obtained using scheduling policies for multiple manufacturing systems.

3 EXPERIMENTAL DESIGN

The earlier section brought to light the confusion that exists over the relative effectiveness of the deterministic, Poisson and CONWIP policies for different types of manufacturing systems. In this section, a full-factorial

experiment is designed which will permit the evaluation of performance of these policies for a wide spectrum of factory configurations.

3.1 Experimental Factors Perturbed

While designing a full-factorial experiment, it is important to intuitively determine a list of factors that are expected to affect the output response in a significant manner. In this case, the configurations used and types of analysis carried out by earlier researchers provide pointers to the factors that need to be perturbed in this experiment. In addition, it has been conjectured in past literature that the superiority of CONWIP type closed-loop policies may be attributed to the fact that there exists a negative correlation between the arrival process and the WIP in the system. A large WIP in the system limits the rate of arrival into the system while a low WIP level accelerates the arrival rate. In other words, variability in WIP translates to arrival variability. This phenomenon may prove favorable to performance measures of some systems. The factors that are considered for this study include those influence the system configuration and those that greatly impact the WIP in the system. These factors, their levels of perturbation and the respective reasons for inclusion are described below.

i) System configuration

This factor refers to the size of the system and is a function of the process flow and the number of work stations in the system. The inclusion of this factor is motivated by the fact that earlier researchers have published results based on experiments carried on different sets of systems. Therefore, the use of this factor will help determine if the size/type of system has any bearing on the relative ranking of the three lot release policies.

The system configuration factor is varied over three levels. Each level corresponds to a process flow and a corresponding tool list. The first system involves three machines and five process steps. The second system involves five machines and thirteen process steps, and the third system configuration comprises nine machines and twenty five process steps. A fourth system is simulated separately. It uses the process flow and tool list corresponding to an actual semiconductor manufacturing system. The process flows are shown in Figure 2, 3 and, 4 and the tool list characteristics are displayed in Tables 1, 2, and 3 respectively. The number of identical tools at each work station is computed by ensuring that the utilization of the station given a product start rate, is less than 1. The product start rate is derived from the average inter-arrival time of wafers. For the purpose of tool-set determination, the average inter-arrival time between lots of 25 wafers each is assumed to be 0.85 hours. This number has been found to be consistent with the scale of operation of typical medium to large-sized semiconductor manufacturing facilities.

Table 1. Tool List for 3 machine 5 step system

Tool no.	No. of passes	Tool TP (wafers/ hour)	MTBF (hours)	MTTR (hours)	No. of tools	Utilization
1	2	75	9	1	1	0.98
2	2	60	10	1	2	0.61
3	1	42	9	1	1	0.87

Sequence of stations: 1-2-1-2-3

Figure 2: Process Flow for 3 machine 5 step system

Table 2. Tool List for 5 machine 13 step system

Tool No.	No. of passes	Tool TP (wafers / hour)	No. of tools	Utilization
1	3	72	2	0.76
2	3	60	2	0.92
3	3	42	3	0.87
4	2	55	2	0.66
5	2	45	2	0.82

Sequence of stations: 1-2-1-2-3-1-2-3-4-5-3-4-5

Figure 3: Process Flow for 5 machine 13 step system:

Table 3. Tool List for 9 machine 25 step system

Tool No.	No. of passes	Tool TP (wafers / hour)	No. of tools	Utilization
1	3	72	2	0.76
2	3	60	2	0.91
3	3	42	3	0.87
4	3	55	3	0.66
5	3	45	3	0.82
6	3	34	4	0.81
7	3	36	4	0.76
8	2	40	2	0.92
9	2	82	1	0.90

1-2-1-2-3-1-2-3-4-5-3-4-5-6-7-4-5-6-7-8-9-6-7-8-9

Figure 4: Process Flow for 9 machine 25 step system:

ii) Release policy

This is the main focus of the experimental design. The

output response derived from the performance measures of interest are observed as a function of release policies under different system scenarios. As described earlier, this factor is varied over three levels.

The first level represents the deterministic release policy where jobs are released into the factory such that the inter-arrival time is a constant. The second level involves releasing jobs into the factory such that the inter-arrival times are sampled from an exponential distribution. The third level studies an implementation of CONWIP for releasing jobs into the factory. Here the system is modeled as a closed queuing network where the departure of a job from the system triggers a new arrival into it. Thus the coefficient of variation (the ratio between the standard deviation and mean of a random process) of inter-arrival time of jobs into the system is identical to the coefficient of variation of the departure process of the last machine.

iii) Start rate

The start rate of wafers is controlled by specifying the mean inter-arrival time of wafers into the system for deterministic and Poisson release policies. Variations in start rates alter the utilization of stations in the system. This factor enables the testing of robustness of the results as a function of factory congestion and equipment utilization. Typically this factor is varied over six levels.

The range of mean inter-arrival times used for this experiment is from 0.85 hours per lot of 25 wafers to 1.3 hours per lot in steps of 0.05 or 0.1. Specifically, the levels of the factors of inter-arrival time are 0.85, 0.9, 0.95, 1.0, 1.05, 1.1 and 1.2. These levels of inter-arrival times are meaningless with respect to CONWIP release policy. This is because, while the start rate for open-loop policies can be controlled by an external arrival rate, the start rate for closed-loop systems is controlled by altering the permitted WIP in the system. Sample preliminary simulation runs are required while ensuring that the start rate resulting from WIP fluctuations in CONWIP/ closed-loop systems is comparable with the start rates achieved by external release in open-loop systems using Poisson and deterministic policies. The different levels of WIP required to achieve average lot inter-arrival times equivalent to 0.85, 0.9, 0.95, 1.0, 1.05, 1.1 and 1.2 hours is determined in an iterative fashion.

iv) Variability in service time

The variability in service time is controlled by altering the probability distribution for service times. This factor is varied over two levels. The arrival stream for CONWIP systems are controlled by WIP fluctuations. The WIP fluctuations are in turn impacted by the variability in service time. Thus, this factor tests the validity of conclusions in the presence of large variability in WIP induced by variability in service time.

The service time is perturbed over two levels. These two levels are deterministic (or constant) service time and exponentially distributed service time with CV of the distribution 0 and 1 respectively.

v) Tool reliability

Semiconductor manufacturing systems are characterized by large variations in tool reliability due to the tight process specifications that need to be maintained on the factory floor. These large fluctuations in tool reliability greatly impact the resultant variability in processing time (Segal and Whitt (1988)). As described in the earlier section, the variability in service time impacts the variability in WIP in a significant manner and is considered an important factor in this experimental study.

Variations in tool reliability are captured through changes in the Mean Time Between Failures (MTBF) and the Mean Time To Repair (MTTR). The distributions of MTBF and MTTR considered are constant and exponential. This factor is varied over three levels. The table below shows the three levels of this factor considered for the experiment.

Table 4. Three levels of tool reliability

Tool No	LOW		MEDIUM		HIGH	
	MTBF (hours)	MTTR (hours)	MTBF (hours)	MTTR (hours)	MTBF (hours)	MTTR (hours)
	Const.	Const.	Expon.	Const.	Expon.	Expon.
1	9	1	9	1	4.5	1
2	10	1	10	1	5	1
3	9	1	9	1	4.5	1
4	10	1	10	1	5	1
5	8	1	8	1	4	1
6	10	1	10	1	5	1
7	8	1	8	1	4	1
8	11	1	11	1	5.5	1
9	7	1	7	1	3.5	1

3.2 Output Responses Studied

This study uses the basis for comparison suggested by Glassey and Resende(1988a) as described in the earlier section. The primary response of interest is cycle time, as a function of the factor levels. Cycle time is estimated using the mean and the half width of a 95% confidence interval. Half width is defined as the absolute value of the difference between one of the bounds of the confidence

interval and the mean. Another response of interest is the coefficient of variation (CV) of the arrival process. The CV of arrivals for the deterministic and Poisson release policies are 0 and 1, respectively. But, for the CONWIP release policy, CV of arrivals is a function of the departure process and the WIP in the system. The CV of arrivals is estimated from the simulation experiment. The third response of interest is number of jobs processed during the simulation run which is a measure of the throughput obtained from the system.

3.3 Simulation Details

A full factorial experiment is carried out with the factors just described and the responses listed are estimated for each combination of the factors. The full-factorial experiment requires a total of 270 (3×3×5×2×3) simulation runs. The number of combinations is derived from the product of the number of levels of each of the factors considered.

Simulations were run using a C based discrete event simulator - Delphi (Chance(1995)), developed at SEMATECH. Simulations were typically run for 1000 days. Statistics were collected after 200 days in order to correct for initialization bias. The duration of the simulation run, number of replications and the truncation point were determined using recommendations made by Schruben *et al.*(1983).

4 ANALYSIS OF RESULTS

The results of the simulation runs for the design points for the experiment are not reported due to the volume of output data generated. To summarize this data, a modified version of the Bonferroni inequality is used. This is elaborated in the next section. Release policies are evaluated for a combination of system configuration, variability of service time and tool reliability. Among the responses obtained for different throughput rates, the mean cycle time pertaining to comparable throughput rates for two release policies, is used to reach a verdict on the relative performance of these release policies.

Comparison of cycle time for two policies can be made by computing the difference of the respective means. There is a caveat here. For the purpose of the comparison, it is necessary to devise a suitable methodology to test for the statistical significance of the differences and it is not sufficient to observe just a small difference in the mean cycle time irrespective of the width of its confidence interval. The primary reason for this fact is that the cycle time observed is typically a confidence interval and hence one does not have a 100 percent confidence interval for the value of cycle time. Thus, one has to compare two intervals, both of which are not 100% confidence intervals. A modified version of the Bonferroni inequality is used to establish confidence intervals with a certain probability, for the difference in cycle time for two systems.

Suppose that I_s is a $100(1-\alpha_s)$ percent confidence interval for the measure of performance μ_s (where s = 1,2,...,k). This can be stated as follows

$$P(\mu_s \in I_s) = 1 - \alpha_s \quad \text{for } s = 1, 2,..., k$$

Then, Law and Kelton (1992) show that the probability that all k confidence intervals simultaneously contain their respective true measures satisfies

$$P(\mu_s \in I_s \text{ for all } s = 1,2,...,k) \geq 1 - \sum_{s=1}^{k} \alpha_s$$

This result is known as the Bonferroni inequality. The above result is altered in the following manner in order to obtain a confidence interval for the difference between two means. In the case of this study, the mean and the half width of cycle time for two systems under the two release policies for a given system configuration and system variability has been obtained from the simulation experiment.

The primary idea used for comparison of two systems is as follows. If system A has a 95% confidence interval of (5.5, 6.3) for a particular performance measure (say μ_1) and system B has a 95% confidence interval of (8.1,9.5) for a measure (say μ_2) then the difference of the two measures ($\mu_1 - \mu_2$) has *at least* a 90% confidence interval of (8.1-6.3, 9.5-5.5) or (1.8, 4). This result is proved in the following manner.

Claim: If $P(\mu_1 \in (l_1,u_1)) = 1 - \alpha_1$ and
$P(\mu_2 \in (l_2,u_2)) = 1 - \alpha_2$ then
$P(\mu_1 - \mu_2 \in (l_1 - u_2, u_1 - l_2)) \geq 1 - \alpha_1 - \alpha_2$

where l_1 and u_1 are the lower and upper bounds respectively for the confidence interval for μ_1 and l_2 and u_2 are the lower and upper bounds, respectively, for the confidence interval for μ_2.

$P(\mu_1 - \mu_2 \in (l_1 - u_2, u_1 - l_2)) =$
$P(\mu_1 - \mu_2 \in (l_1 - u_2, u_1 - l_2) / (\mu_1 \in (l_1, u_1) \text{ and } (\mu_2 \in (l_2, u_2)))$
+
$P(\mu_1 - \mu_2 \in (l_1 - u_2, u_1 - l_2) / (\mu_1 \in (l_1, u_1) \text{ and } (\mu_2 \notin (l_2, u_2)))$
+
$P(\mu_1 - \mu_2 \in (l_1 - u_2, u_1 - l_2) / (\mu_1 \notin (l_1, u_1) \text{ and } (\mu_2 \in (l_2, u_2)))$
+
$P(\mu_1 - \mu_2 \in (l_1 - u_2, u_1 - l_2) / (\mu_1 \notin (l_1, u_1) \text{ and } (\mu_2 \notin (l_2, u_2)))$

It is apparent that the expression
$P(\mu_1 - \mu_2 \in (l_1 - u_2, u_1 - l_2) / (\mu_1 \in (l_1, u_1) \text{ and } (\mu_2 \in (l_2, u_2)))$

$= P(\mu_1 \in (l_1, u_1) \text{ and } \mu_2 \in (l_2, u_2)) = 1 - \alpha_1 - \alpha_2$

The other three expressions are all greater than or equal to zero as they are all probabilities.

Thus,
$P(\mu_1 - \mu_2 \in (l_1 - u_2, u_1 - l_2)) = 1 - \alpha_1 - \alpha_2 +$ (a positive quantity)
So, $P(\mu_1 - \mu_2 \in (l_1 - u_2, u_1 - l_2)) \geq 1 - \alpha_1 - \alpha_2$

The following numerical example is used to explain this result. Consider two 95% confidence intervals for μ_1 and μ_2.
Let $P(\mu_1 \in (19.5,20.5)=0.95$ and $P(\mu_2 \in (17,19))=0.95$.
Then $P(\mu_1 - \mu_2 \in (19.5-19,20.5-17)) \geq 0.9$ or
$P(\mu_1 - \mu_2 \in (0.5,3.5)) \geq 0.9$

With respect to the results of the experiment, if the cycle time and the corresponding half-width are computed for two systems at a given throughput, then a 90% confidence interval for the difference in cycle time for the two systems can be obtained. If this interval does not contain 0 it implies that the difference between the two means is greater than 0 with at least 90% probability. Then, with at least 90% confidence it can be stated that one system clearly dominates the other in terms of the performance measure. In other words the probability that one system is superior to the other is at least 0.9.

The concepts stated in the earlier paragraph were used to statistically prove the superiority of one release policy over the others under different scenarios. If both the bounds in the confidence interval for the difference in means are negative then CONWIP is superior to deterministic policy for that configuration. If both the bounds are positive then the converse is true. In the following tables, the results of the application of this result to the experiment results are summarized. In Table 5, the number of experiments where the differences indicated the superiority of CONWIP over deterministic policy (C > D) and vice-versa (C < D) and number of instances where the difference was insignificant are reported. Table 6 shows the average 90% confidence lower and upper bounds for the difference in mean for the three instances. Table 7 displays the utilization of the bottleneck tool and Table 8 lists the average CV of arrival under CONWIP for each of the instances.

Table 5 CONWIP vs. Deterministic - comparison count

Name of system	C > D	D = C	C < D
3mac_5step	10	12	0
5mac_13step	4	16	10
9mac_25step	0	2	26

Table 6. Average lower bound, Average upper bound in each of the instances

Name of system	C > D	D = C	C < D
3mac_5step	(-1.6, -0.52)	(-0.49, 0.38)	-
5mac_13step	(-1.7, -1.1)	(-0.2, 0.04)	(0.54, 0.68)
9mac_25step	-	(-0.1, 0.5)	(1.19, 2.88)

Table 7. Average utilization of bottleneck in each of the three instances

Name of system	C > D	D = C	C < D
3mac_5step	0.86	0.67	-
5mac_13step	0.79	0.72	0.83
9mac_25step	-	0.68	0.75

Table 8. Average CV in each of the three instances for CONWIP release policy

Name of system	C > D	D = C	C < D
3mac_5step	0.62	0.51	-
5mac_13step	0.47	0.62	0.65
9mac_25step	-	0.58	0.74

5 DISCUSSION OF RESULTS

The conclusions observed from the summary results are classified on the basis of system configuration. In all the cases, the difference between CONWIP and deterministic release policies tends to be almost statistically insignificant for low values of utilization (for low throughput) of the work stations (from Table 7).

i) The 3 machine, 5 step system

CONWIP policy clearly dominates the deterministic release policy. This result can be inferred from the fact that in 10 of the 22 experiments run with this configuration, both the bounds are negative while none of the instances show the converse result. 12 cases show statistically insignificant difference, generally due to low utilization. These results hold true even when the CV of the arrival process is high (greater than 0.8). The arrival pattern, though highly variable, effectively uses information about the state of the system and results in lower cycle time than the deterministic release policy. Thus, for a small system, the total WIP in the system is a useful measure for controlling release of lots into the system if the objective is the minimization of mean cycle time. Both CONWIP and deterministic release policy dominate the Poisson release policy.

ii) The 5 Machine, 13 step system

This system has lesser number of instances where the CONWIP release policy dominates the deterministic release policy. In 4 experiments, the CONWIP policy behaves superior to the deterministic release, 10 of them show the converse result and 16 experiments resulted in statistically insignificant differences.

Table 8 shows that CONWIP dominates deterministic release for systems with low CV of arrival while the converse result is observed for high CV values. Middle values of CV go with statistically insignificant differences. This implies that for highly variable systems, the total WIP in the system is not a good measure to control release of wafers into the factory if the performance measure of interest is mean cycle time. But for systems which are less variable (resulting from medium level tool reliability or medium variation in service time), one observes smaller values of CV of release under CONWIP release and the differences between the mean cycle time under the two release policies tends to become insignificant. This indicates that the use of WIP in the system as a parameter to control release of wafers into the system neither aids or hinders the performance of the system. For systems with even further reduced variability (low level of tool reliability and deterministic service distribution) the CONWIP release behaves superior to the deterministic release policy. Here, the value of CV of arrival is very small. For such systems, the total WIP in the system is an useful measure to control the release of lots into the system.

Both CONWIP and deterministic release policies dominate the Poisson release policy.

iii) The 9 machine 25 step system

The deterministic release policy dominates CONWIP release policy for this system in most of the experiments. 26 instances result in positive lower and upper bounds and 2 result in statistically insignificant differences. None of the instances resulted in CONWIP dominating the deterministic release policy. The average coefficient of variation of the arrival process for the CONWIP release policy is 0.735.

Thus, the total WIP in the system may not be an useful measure for controlling release of lots into large systems if the performance measure of interest is mean cycle time.

As demonstrated by Spearman and Zazanis(1992) and Wein(1988) the CONWIP release policy dominates the Poisson release policy.

6 ANALYSIS OF ACTUAL SYSTEM

The purpose of this analysis is to quantitatively study tradeoffs between different types of release policies used in semiconductor manufacturing systems. Again, the release policies of interest are the deterministic release, CONWIP and Poisson release policies. The manufacturing system used for the initial simulation results consists of the process flow and tool set defined by process engineers at SEMATECH. The tool set comprises of 45 different work stations each of which consists of parallel identical machines. Machines are highly unreliable as typical reliability levels are around 80%. The process flow comprises of 250 process steps. The number of process steps is much greater than the number of work stations. This means that there is reentrant flow in the system. Assuming no scrap, the start rate is the desired output from the system. It is generally specified as the number of wafers processed in a week. For simulations performed, this value has been perturbed about 4500 wafer starts per week. The number of parallel identical tools at each work station is determined using a spreadsheet calculation by ensuring that the utilization of the work station is less than 1 assuming a start rate of 4500 wafers per week.

The performance measures of interest are the mean and standard deviation of cycle time. The simulations were run using ManSim for 1000 days, with an initial condition of zero work-in-process inventory in the factory. Statistics are collected after a period of 200 days (warm-up period) to account for initialization bias.

The results of the simulation runs are shown below. Tables 9 and 10 provide the mean and standard dev. of cycle time as a function of start rates for the three release policies.

Table 9. Throughput vs. Mean Cycle Time

Throughput	Cycle Time (in hours)		
(wafers per week)	Deterministic	CONWIP	Poisson
4606	562.3	586.3	601.2
4501	497.6	522.4	540.2
4269	418.8	441	458.7
4195	402.1	419.3	433.4
3897	358.3	369.9	379.3
3498	325	334.3	341.6

Table 10. TP vs. Standard Dev. of Cycle Time

Throughput	Standard deviation of Cycle Time (hours)		
wafers/week	Deterministic	CONWIP	Poisson
4606	23.9	17	24.2
4501	19.6	11.9	19.4
4269	17.1	11.4	18.2
4195	16.6	11.2	15.3
3897	15.3	12.6	14.9

The above results show that the mean cycle time using

deterministic release policy is lower than that for Poisson and CONWIP release policy. The standard deviation under CONWIP is lesser than that for the other two release policies.

7 CONCLUSIONS

The results of the previous sections showed that it is not possible to assume that one release performs better than another under all circumstances. A careful study is required for each system to study the relationship between the departure process of the last machine and the utilization of the bottleneck. Based on these initial studies, it appears that the deterministic release policy dominates CONWIP and Poisson policy for larger systems while the CONWIP is the superior policy for small systems. There is also some preliminary evidence on the correlation between superiority of deterministic policy, high utilization and high CV of arrival under CONWIP policy.

For ASIC (Application Specific Integrated Circuits) products, if there is a need to ensure on-time delivery (minimization of both earliness and tardiness), CONWIP seems to be an effective release policy. If minimization of mean cycle time is of greatest importance, then the deterministic release policy is recommended. For make-to-market products like DRAMs, the use of CONWIP will ensure maximum throughput without any WIP explosions in the factory. In other words, it will prevent overloading the factory without explicit consideration for its actual capacity. The yield implications of these release policies is not very obvious from these results. But Srinivasan, Sandell and Brown (1994) have shown that a low mean cycle time translates to smaller yield losses. Hence, the deterministic release policy which results in smaller mean cycle time might lead to smaller yield losses. In reality, the actual time spent at each of the process steps has to be known before such conclusions can be drawn.

REFERENCES

Chance Frank, DELPHI: A C-Based Manufacturing Simulator, Version 8, Level 27.

Glassey C. Roger, and Mauricio G.C. Resende, Closed-Loop Job Release Control for VLSI Circuit Manufacturing, *IEEE Transactions on Semicon-ductor Manufacturing*, Vol. 1, No. 1, pp. 36- 46, 1988a.

Glassey C. Roger, and Mauricio G.C. Resende, A Scheduling Rule for Job Release in Semiconductor Fabrication, *Operations Research Letters*, Vol. 7, No. 5, pp. 213-217, 1988b.

Law M. Averill, and W. David Kelton, *Simulation Modeling and Analysis*, McGraw-Hill Inc., Second Edition, 1991.

Lu C.H Steve, D. Ramaswamy, and P. R. Kumar, Efficient Scheduling Policies to Reduce Mean and Variance of Cycle-Time in Semiconductor Manufacturing Plants, *IEEE Transactions on Semiconductor Manufacturing*, Vol. 7, No. 3, pp. 374-388, 1994.

Schruben L., H. Singh, and L. Tierney, Optimal Tests for Initialization Bias in Simulation Output, *Operations Research*, Vol. 31, pp. 1167-1178, 1983.

Segal M., and W. Whitt, A Queuing Network Analyzer for Manufacturing, *Proceedings of 12th International Teletraffic Congress*, Torino, Italy, 1988.

Spearman L. Mark, and Michael A. Zazanis, Push and Pull Production Systems: Issues and Comparisons, *Operations Research*, Vol. 40, No. 3, pp. 521-532, 1992.

Srinivasan K., R. Sandell and S. Brown, Correlation Between Yield and Waiting Time, *SEMATECH Technology Transfer Report No. 94112635 A-XFR.*

Wein M. Lawrence, Scheduling Semiconductor Wafer Fabrication, *IEEE Transactions on Semiconductor Manufacturing*, Vol. 1, No. 3, pp. 115-129, 1988.

AUTHOR BIOGRAPHIES

RAKA SANDELL is a Manufacturing Systems Engineer at MOS 12 Die Manufacturing at Motorola in Chandler, Arizona, specializing in production planning, real-time dispatching, and modeling, analysis and control of semiconductor manufacturing systems. She has a B.S. degree in Electrical Engineering from the University of Delhi and an M.S. degree in Operations Research from the University of Texas, Austin.

KRISHNA SRINIVASAN is a Senior Modeling Engineer with the Worldwide Enterprise Modeling Group of Motorola Manufacturing Systems at Motorola's Corporate Offices in Schaumburg, Illinois. His areas of specialization include developing solutions for sales, operations and production planning problems, queuing networks and factory modeling. He has a B.S. degree in Mechanical Engineering from the Indian Institute of Technology, Madras and an M.S. degree in Operations Research from the University of Texas, Austin.