行政院國家科學委員會專題研究計畫成果報告

以光罩更換次數為主要評估標準，為黃光區步進式對準機發展一啟發性投料與派工法則

Developing a Heuristic Dispatch Rule for Steppers in Photolithography Area to Minimize the Number of Photomask Changes

計畫編號：NSC 88-2416-H-002-056
執行期限：87 年 8 月 1 日至 88 年 7 月 31 日
主持人：陳靜枝　　臺大資訊管理系所

一、中文摘要

本研究探討晶圓製造生產流程的瓶頸工作區—黃光區—中步進式對準機特殊的生產特性，也就是加工元件具有回流、且必須爭奪同一資源（回到同一機台或是只有一台機台）、機台製程轉換的整備時間耗時等特性，提出了一套啟發式的下線法則，稱為順序投料法（SEQ-SEQ input control policy）。其作法是針對某一機台連續投料 K 個加工元件，且機台必須處理完這 K 個加工元件的同一製程後，才能轉換製程。其目的是藉由減少製程轉換的次數，以減少機台因製程轉換所造成的時間浪費，進而提高機台的使用率而增加產能。

為決定連續投料的批數 K 與控制何時投料的參數，本研究提出了一組合演算法，以最小化加工元件等候時間和機台閒置時間的加權總和作為搜尋控制參數的準則，藉此找到一組投料數 K 與控制何時投料的工作負荷臨界值，使得回流加工元件彼此衝撞的機會降低，且機台的使用率又能維持一定水準之上。

為了評估順序投料法在實際晶圓廠的適用性，和比較順序投料法與其他已知下線法則的優劣，本研究是利用系統模擬的方法來驗證。並且依據台灣某 DRAM 製造廠之生產流程以及真實的生產資料，建構一模擬模型作為測試環境。最後，模擬結果顯示順序投料法在不增長生產週期的情況下，對於減少製程轉換的次數、提升機台的產能均有顯著的改善。所以說順序投料法的確適用於具有回流限制的生產環境，而利用組合演算法所搜尋的投料數與控制下線的臨界值也是有效的。

關鍵詞：黃光區、步進式對準機、晶圓投料法、順序投料法、回流、系統模擬

Abstract

This study proposed an input control policy, called “SEQ-SEQ,” for a wafer fabrication based on the special production characteristics of the steppers in the photolithography area. For special production characteristics of steppers, it means that the production environment possesses reentrant production flows where the same machine should processes the same job when the job returns to this working area. To add the complexity, the setup time of each production operation is time-consuming. The “SEQ-SEQ” input control policy releases k jobs sequentially for a machine when the workload of this machine is lower than a predefined threshold, and the machine wouldn’t be allowed to change production operations until the same operation of k jobs are finished. The purpose of SEQ-SEQ is to reduce the time of setup operations, and to raise the utilization of the machine. An algorithm is proposed to determine the two parameters to minimize the sum of the job waiting time and the machine idle time.

In order to evaluate the suitability of the policy in practice and to compare the performance with other input control policies, the method of system simulation is adopted in this study. The simulation model is built according to the observation of a DRAM wafer fabrication in Taiwan and the data needed in the simulation model is collected from this plant. Finally, as observing from the simulation results , SEQ-SEQ would decrease the time of changing operations and increase the throughput without increasing the cycle time

Keywords: Photolighography area, Input control policy, Stepper, SEQ-SEQ policy, Simulation, Wafer Fabrication, Workload,.

I. Introduction

Due to the complexities of the wafer production processes and the intensities of capital and technical investment, it is important for semiconductor manufacturers to utilize their equipments efficiently. However, the wafer production processes are different from general production processes in such a way that producing one wafer needs hundreds of operations, and takes a few months to complete. Meanwhile, the production characteristics such as wafer reentrance, equipment downtime, random yields, and so on make scheduling and

dispatching in semiconductor wafer fabrication particularly difficult. These difficulties and problems are elaborated in Uzoy, Lee, and Martin-Vega [26] as well as in Hughes [12] and Bai & Gershwin [2].Uzsoy, Lee and Martin-Vega[27] review the research related to shop floor control in the semiconductor industry. They classify the research by techniques used in solution and discuss the pros and cons of these approaches. A thorough description of wafer fabrication and specific modeling and analysis problems are seen in Johri [13]. In additions, Leachman and Hodges [17] conducted on-site interviews with manufacturing personnel in 16 plants to identify factors in different facets correlated strongly with productivity.

From this point of view, to have an effective operation control strategy is urgent for a wafer fabrication plant. An operation control strategy is usually a combined choice of an input control policy and a dispatch rule. In the study of input control methods, Glassey and Resende [9] propose a heuristic approach called a “Starvation Avoidance” input control method similar to the reorder point concept in the traditional EOQ model. They evaluate this approach through simulation and conclude that this approach is more appropriate to a wafer fabrication plant, in terms of reducing WIP and increasing throughput, than the traditional input control method. Lozinski and Glassey [20] further develop a mechanism to determine Bottleneck Starvation Indicators according to the Starvation Avoidance Input Control method. Lawton, Drake, Henderson, Wein, Whitney, and Zuanich[16] propose an input control method, called Workload Regulating Wafer Release Method, that regulate wafer input by workload measurement of the bottleneck machines. That is, the wafers are input into the system when the workload is less than a threshold that is chosen to guarantee a desired level of throughput. They compare this method with Uniform input and Poisson input methods through simulation and conclude that their method is superior among all policies. They provide a way of computing the workload though did not suggest the way of calculating the threshold. Wein [28] use a Brownian network model to approximate a multiclass queueing network model with dynamic control capability which emulates a wafer fabrication plant. His study shows that input control methods are more effective than dispatch rules in terms of reducing the cycle time and WIP.

In the study of dispatch rule, Adachi, Talavage, and Moodie [1] adopt a rule-based control method in a decision support expert system, called a rule-based decision support system (RBDSS). Through simulation, they compare the performance of RBDSS with that of a non-AI system and conclude that the decision support system performs better. Glassey and Petrakian [8] also develop a dispatch rule based on the principle of starvation avoidance. However, their algorithm continuously adjust the priorities of the jobs waiting before work units which results in requirements of large capacity and longer processing time of a computer unit. Lu and Kumar [21], on the other hand, evaluate buffer-based policies vs. due-date-based policies through control theory to develop the bounds of delays for a wafer fabrication plant. They also use simulation in their study to compare these two types of policies. They conclude that due-date-based policies are also stable in the long run. Chang and Liao [3] develop a three-layer real-time scheduling and dispatching control framework for a semiconductor wafer fabrication. They adopt the Lagrangian Relaxation and Network flow techniques to solve this problem. Kumar [15] modified least slack first dispatch rule and proved that the rules are more efficient than other rules such as FIFO, LWNQ, and so on in terms of reducing mean and variance of cycle time. Levitt and Abraham [18] proposed a Just-In-Time method for semiconductor manufacturing which requires bottleneck to be identified and Kanban cards to be consistently adjusted. Chern, Liu, and Shieh [6] propose a dispatch algorithm for Stepper, called Stepper Dispatch Algorithm or SDA, in a wafer manufacturing plant.

They construct a simulation model and proven statistically that SDA is superior to the other dispatch rules such as FIFO, SPRT, and Least Slack. Chern, Hsu, and Shieh[5] further develop a dispatch algorithm for steppers and furnaces, called Time-Performance-Mask Algorithm or TPMA, which adopts some the principles of assigning priorities to wafers from SDA, in a wafer manufacturing plant. They also construct a simulation model and proven statistically that TPMA is superior to the other dispatch rules such as FIFO, SPRT, and Least Slack but not different from SDA. Other discussions of general scheduling and dispatching algorithms could be found in [11], [14], and [23] though they are not directly related to a wafer fabrication process.

In all of the above studies, simulation is used as a tool to either evaluate the algorithms or justify their arguments. Dayhoff and Atherton[7] first utilize discrete event simulation technique to model the process of a semiconductor wafer fab. Miller[22] reviews the experience of working in IBM's General Technology Division. Spence and Welter[25] also adopt a simulation model to evaluate the performance of working units in the photolithography area of a wafer fab. They compute signatures of cycle time and throughput to judge the performance. Instead of simulation, Chen, Harrison, Madelbaum, Ackere, and Wein[4] develop a mathematical model of a wafer fabricating process and derive the formula of average queue length and average waiting time by using theory of queueing network. However, the formula could be used only to a simplified model and the error of approximation could be up to 14%. Lin and Raghavendra[19] study the Join the Shortest Queue (JSQ) policy through queue network model though they assume that all the queues are identical. They develop a method to approximate the performance of the system and show that the method provides accurate estimates of the mean response time.

The basic operation areas for a wafer fabrication are diffusion, thin film, etching, photolithography, and ion implant as seen in Figure 1. The bottleneck of the whole process is photolithography process of which photomask operations performed on the steppers are most complicated and requires high level of exactitude. Among all kinds of wafer fabrications, DRAM wafer is very difficult to produce and requires special treatments different from other types of wafers. Each DRAM wafer has to go through lithographic area more than ten times. Each time a different layer of circuit is implanted on the DRAM wafer through a different photomask operation that requires different processing time as well as a unique mask. Furthermore, the wafer has to be processed on the same stepper for these photomask operations because of precision requirements. However, steppers are very expensive and mask changes on steppers are troublesome and time-consuming. Under the requirement to reduce the number of mask changes, this research is focused on finding a heuristic operation control strategy to maximize the utilization of the steppers. Although many input control method as well as dispatch algorithms have been developed and tested in the above studies, none of them are actually implemented in a wafer fabrication plant. One of the reasons is that they are too complicated to implement. As results of this work, a simple and easy-to-implement input control policy, called sequential-sequential Algorithm or SEQ-SEQ, is constructed to effectively reduce cycle time at a required stepper utilization rate for a wafer fabrication plant. Like the above studies, simulation will be used to evaluate this input control policy since the system is too complicated to be analytically modeled. The data needed in simulation model are collected from a large U.S. and Taiwan wafer fabrication joint venture.

The rest of the paper is organized as follows. Section 2 presents the construction of SEQ-SEQ Algorithm. Section 3 characterizes the simulation model of a wafer fabrication and lists the parameters needed in

simulation. Section 4 displays the results of simulations and performs hypothesis tests through statistical analyses. Finally, section 5 summarizes the results.

II. Problem Description

A typical production cycle of a DRAM wafer starts with non-photolithography processes such as diffusion, thin film, or etching as seen in Figure 1. Once the wafer enters photolithography area, it is coated, photomasked, and developed for the first layer of circuit as seen in Figure 4. It is assumed that the result of photolithography process is either good or rework but not scraped. The wafer is then transferred to non-photolithography areas and stays there for some other processes such as ion implant, diffusion, and etc. The wafer then returns to photolithography area for the second-layer circuit and so on. The whole production cycle of a wafer finishes when all the required layers of circuits are properly produced. It should be noted that SEQ-SEQ is designed for a general DRAM wafer fab which may have only one type of product and whose inventory policy is make-to-stock.

A DRAM wafer usually needs to go through hundreds of different operations of which at least *10* are photomask operations processed on steppers. Two types of photomask operations are usually performed on steppers: critical or non-critical. Critical operations required to be performed on the same steppers for the same wafer while non-critical operations do not have this restriction. Although all steppers have the capabilities of performing both operations, DRAM wafer fab usually separates steppers for critical and non-critical operations. Every photomask operation needs a distinguishing mask while mask changes on steppers are troublesome and time-consuming tasks. In order to minimize the number of mask changes, wafers are released to the steppers sequentially so that they can be processed consecutively without changing the masks too often. However, the time intervals between two critical operations for the same wafer are usually different. Therefore, the wafer might wait for steppers for some operations while the steppers might be idle between some operations when no wafers in queue. For example, two lots of wafers are released to the system sequentially. The first lot finishes the first critical operation and goes to the other areas. The first lot comes back to the same stepper in several minutes only to find that the second lot is still processed the first critical operation on the same stepper. Under this circumstance, the first lot has to wait in queue until the second lot finishes the first critical operation. In another case, the second lot finishes the first operation before the first lot returns. The stepper is idle and has to wait until the first or second lot returns. In any case the resource is wasted.

The objective of this research is to find the optimal input control policy, called SEQ-SEQ, which minimizes the number of mask changes and mean cycle time. In the following section, the parameters needed in SEQ-SEQ are defined and elaborated.

III. SEQ-SEQ Algorithm

An operation control strategy usually consists of two rules: input control and dispatch. An input control rule tells the system when to release new jobs and how many new jobs to release. A dispatch rule tells the machines which job to process when it is idle. This study proposes an input control method, called SEQ-SEQ, which determines the time to release new jobs by comparing the current total workload for a particular stepper with a threshold level as seen in Figure 2(a) and 2(b). If the workload is less than the threshold, then k lots of new jobs are released to the system for this stepper. The SEQ-SEQ algorithm calculates the parameters such as the workload threshold, W, and the batch size of new jobs, k. Before SEQ-SEQ is constructed, some decision variables and state variables are defined as follows first:

k: the number of new lots to be released
n: the number of critical operations
C: the time to change masks
P_{ij}: the processing time of critical operation i for lot j where $1 \le i \le n$ and $1 \le j \le k$
A_{ij}: the time interval between critical operation i and $i+1$ in non-stepper area for lot j where $0 \le i \le n$ and $1 \le j \le k$
T_{ij}: the starting processing time of lot j for critical operation i where $0 \le i \le n$ and $1 \le j \le k$
Ω: the total current workload of the stepper

In additions, several assumptions are imposed before the following computation.

- Simplify the whole wafer process as Figure 2(a).
- Assume that all the processing time, P_{ij}, and the time interval between two critical operations, A_{ij}, are known.
- Assume the time to change mask, C, for every lot and every operation is known and a constant.
- Assume only one stepper in the system since all the wafers are returned to the same steppers in which they are processed for the first critical operation.
- Assume that the system starts from empty.

With the above definitions and assumptions, the starting processing time, T_{ij}, could be computed by the following ways:

(1) $T_{01}=0,\ T_{02}=T_{01}+A_{01},\ T_{03}=T_{02}+A_{02},\ T_{0k}=T_{0(k-1)}+A_{0(k-1)}$
(2) $T_{11} = (T_{01}+A_{01})+P_{11}+C$
(3) $T_{12} = Max(T_{02} + A_{02},\ T_{11}) + P_{12}$
(4) $T_{1k} = Max(T_{0k} + A_{0k},\ T_{1(k-1)}) + P_{1k}$
(5) $T_{21} = Max(T_{11} + A_{11},\ T_{1k}) + P_{21} + C$
(6) $T_{22} = Max(T_{12} + A_{12},\ T_{21}) + P_{22}$
(7) $T_{2k} = Max(T_{1k} + A_{1k},\ T_{2(k-1)}) + P_{2k}$
(8) $T_{n1} = Max(T_{(n-1)1} + A_{(n-1)1},\ T_{(n-1)k}) + P_{n1} + C$
(9) $T_{nk} = Max(T_{(n-1)k} + A_{(n-1)k},\ T_{n(k-1)}) + P_{nk}$

Subscript 0 represents the release of the first new job before the first critical operation. Since the system is empty, the first lot is processed immediately after released into the system. The process starting time of the second and the following $k-1$ lots, on the other hand, have to choose the maximum of the arrival time of itself and the departure time of the previous lot. Therefore the above derivation of T_{ij} can be simplified as follows:

If $j=1$, $T_{ij} = Max(T_{(i-1)j}+A_{(i-1)j},\ T_{(i-1)k}) + P_{ij}+C$.
Else, $T_{ij} = Max(T_{(i-1)j} + A_{(i-1)j},\ T_{i(j-1)}) + P_{ij}$.

The objective is to find the optimal k that minimizes the weighted sum of the wafer waiting time and the stepper idle time as an example seen in Figure 3(a). That is,

$$\mathrm{Min}_k\Big(\sum_{i=1}^{n-1}\sum_{j=2}^{k}|\delta_1(T_{ij} + A_{ij} - T_{(i+1)(j-1)})| + \sum_{i=1}^{n-1}|\delta_1(T_{i1} + A_{i1} - T_{ik})|\Big),$$

If $(T_{ij} + A_{ij} - T_{(i+1)(j-1)}) < 0$ or $(T_{i1} + A_{i1} - T_{ik}) < 0$, then $I=0$
Else $I=1$.

In other words, δ_0 represents the weight of wafer waiting time and δ_1 represents the weight of stepper idle time. SEQ-SEQ algorithm is construction to find the optimal solution based on the following bounds:

- Lower bound of k: 1.
- Upper bound of k: $\lceil \mathrm{Max}\ A_{ij}/\mathrm{Min}\ P_i \rceil$.

The reason for the lower bound of k to be 1 is that at least one lot of wafers has to be released into the system. The reason to justify the upper bound of $k = \lceil \mathrm{Max}\ A_{ij}/\mathrm{Min}\ P_i \rceil$ is that the machine will not be idle when $\lceil \mathrm{Max}\ A_{ij}/\mathrm{Min}\ P_i \rceil$ lots of wafers are released to the system. Therefore, more than $\lceil \mathrm{Max}\ A_{ij}/\mathrm{Min}\ P_i \rceil$ lots of new jobs released into the system will only increase wafers' waiting time.

To determine the releasing time of new jobs, a threshold, W, has to be found first. Whenever a new lot of wafers released into the system, the total future processing time will be added into the total current workload, Ω, for this particular stepper. Whenever a wafer finished the current operation, the processing time is deducted from Ω. The total current workload, Ω, is compared with the threshold, W. If $W \ge \Omega$, k lots of new

wafers is released. Otherwise no action. An example with k=10 as seen in Figure 3(b) could demonstrate the relationship between time and workload. SEQ-SEQ algorithm is constructed to find the optimal solution based on the following bounds:

- Lower bound of W: 0 lots of wafers.
- Upper bound of W: $\lceil \text{Max } A_{ij}/\text{Min } P_i \rceil$ lots of wafers.

The new wafers will be released when the system is empty if W = the workload of 0 lots of wafers. However, if W = the workload of $\lceil \text{Max } A_{ij}/\text{Min } P_i \rceil$ lots of wafers, the system releases new lots when the first operation of the first lot is finished. At this moment, so many wafers already saturate the stepper, which leaves the stepper no idle time. More wafers released only makes the wafers wait longer.

Combining the above argument, SEQ-SEQ algorithm is constructed as follows:

1. Input all the needed information including, n, C, δ_0, δ_1, A_{ij}, and P_{ij}.
2. Initialize k_0, Q_0, and W_0 to be 0.
3. Let M be the total processing time of $\lceil \text{Max } A_{ij}/\text{Min } P_i \rceil * \lceil \text{Max } A_{ij}/\text{Min } P_i \rceil$ lots of wafers.
4. For $k=1$ to $\lceil \text{Max } A_{ij}/\text{Min } P_i \rceil$
5. Let Ω be the total processing time of k lots of wafers.
6. For $W=0$ to $\lceil \text{Max } A_{ij}/\text{Min } P_i \rceil$
7. Initialize Idle_time, Waiting_time, and TNOW to be 0.
8. While TNOW < M
9. For $i = 1$ to n
10. For $j = 1$ to k
11. If machine's mask ≠ job's operation, then change masks, which adds C to TNOW.
12. If (T_{ij} <= TNOW), add (TNOW $-T_{ij}$) to Waiting_time and P_{ij} to TNOW.
    Else add (T_{ij} −TNOW) to Idle_time and update TNOW by ($T_{ij} + P_{ij}$).
13. Update T_{ij} by (TNOW + A_{ij})
14. Subtract P_{ij} from Ω.
15. If $\Omega \le W$, release another k lots of wafers into system and add the total processing time of k lots of wafers into Ω.
16. Endfor
17. Endfor
18. Endwhile
19. Compute S=(δ_0Waiting_time + δ_1 Idle_time)
20. If $Q_0 > S$, then $k_0=k$, $W_0=W$, and $Q_0=S$.
21. Endfor
22. Endfor
23. Output k_0, W_0, and Q_0.

SEQ-SEQ input any set of n, C, δ_0, δ_1, A_{ij}, and P_{ij}, in the first step. SEQ-SEQ initializes the final optimal solution (k_0, Q_0, W_0) to be zero in step 2. The total elapsed time for each combination of (k, Q, W) is computed in step 3. The search range of k is set in step 4. Let Ω be the total workload in the system. Ω is equal to the total processing time of k lots of wafers at first in step 5. The search range of W is set in step 6. The system variables are initialized in step 7. The algorithm computation elapsed time is controlled in step 8. Each policy has to go through all the lots and operations and the computation ranges are set in step 9 and 10. If the operations for consecutive lots are different, additional mask change time is added in step 11. The idle time or waiting time is computed in step 12 according to the previous discussion. Each lot's starting time is updated in step 13. The processing of this lot for the particular operation is subtracted from total workload, Ω, in step 14. The new job release decision is made in step 15 when comparing Ω with W. The weighted sum of the idle time and the waiting time is computed in step 19. The final optimal solution is found in step 20.

The performance of SEQ-SEQ is evaluated through simulation in the next two sections.

IV. Simulation Model

A typical wafer fabrication cycle starts with

non-photolithography processes such as cleaning and diffusion as seen in Figure 1. Once the chips enter photolithography area, five operations are performed as seen in Figure 4. The wafers leave the photolithography area after the first layer of circuit is properly produced. The wafers continue the processes in non-lithography and lithography areas more than ten times until all required circuit are properly produced. SLAM II is used to build the simulation model to evaluate SEQ-SEQ with other input control policies. The data needed in this simulation model is collected from an U.S.-Taiwan joint wafer fabrication located in Hsin-Chu, Taiwan. As seen in Table 1, each wafer has to go through similar processes but takes different lengths of stepper time as well as other machines' time. Assumptions about this model are listed as follows:

1. Only one type of product is considered in this model because most of the time the DRAM wafer plant has only one major DRAM product in production.
2. Each lot consists of 25 wafers.
3. The engineering lot and engineer holding time of steppers are both ignored.
4. The outcome of each finished critical operation is an independent Bernoulli trial with rework probabilities listed in Table 2 and is either good or rework but not scraped. Each reworking lot delays a constant time of *24* hours and a variable time following an exponential distribution with mean = *36* hours.
5. The time between machine failures and repair time for each stepper and other machines are both assumed random variables following certain exponential distributions with means listed in Table 3(a), (b), and (c). Each machine is also scheduled for the regular maintenance as seen in Table 4.
6. The time to prepare each input lot is a constant of *12* minutes.
7. Hot lot percentages are *0.07*.
8. Four photomasks are available for each photomask operation. Two cases are considered in this simulation model, one with mask changing time = *6* minutes and the other with mask changing time = *20* minutes.

To validate the simulation model, suppose that the input control method is uniform input with *27.857* lots each day (*28* lots for *6* days and *27* lot for *1* day) the dispatch rule used for steppers is FCFS. *720* days is chosen to be the total simulation time with a *360*-day warm-up period. Figure 5(a) shows the WIP vs. time Figure 5(b) shows the utilization of steppers vs. time from simulation runs. As seen in both figures, after one year, the system becomes quite stable. Also from the simulation the average throughput is *27.88* lots and average cycle time is *64.93* days. The average WIP from simulation run is *1808.7* lots which is equal to the average WIP is *1808.8* lots from Little's formula. According to this input rate, the utilization rate of each stepper is *0.9437*. The failure and maintenance rates are *0.0109* and *0.0028* respectively. The total stepper busy rate is *0.9546* that is close to *0.9548*, the busy rate from the simulation run. As results, the model is verified and validated to be accurate. Therefore, the observations from this simulation model could be used to compare SEQ-SEQ and other strategies.

V. Computational Analysis

In this section, five input control policies are compared. It should be noted that 98% of steppers' busy rate is assumed. By this assumption, the parameters of these five input control policies are computed and describe as follows:

1. Uniform Input Control (UN): The policy releases new jobs into system in a constant rate every day. In the case of mask changing time = *6* minutes, the rate is *27.857* lots of wafers (*28* lots for *6* days and *27* lot for *1* day) every day. In the case of mask changing time = *20* minutes, the rate is *25.857* lots of wafers (*26* lots for *6* days and *25* lot for *1* day)

every day. The interarrival time between two lots follows Exp(*7* minutes).

2. Fix-WIP Input Control (FW): The policy keeps a constant WIP level of *1805* lots of wafers in the system when mask changing time is *6* minutes while keeps the level of WIP to be *1647* lots of wafers in the system when mask changing time is *20* minutes. Whenever, the WIP is less than this level, new lots of wafers are released into system.
3. Workload Regulation Input Control (WL): The policy keeps a constant workload of each stepper to be *327000* minutes when mask changing time = *6* minutes or *298000* minutes when mask changing time = *20* minutes. The system releases new wafers whenever the workload of each stepper drops below this threshold level.
4. Starvation Avoidance Input Control (SA): This method computes the virtual inventory between new job and the return job for steppers. However, it is assume that the new job arrives at steppers immediately right after releasing. Therefore, virtual inventory in this case is *0*. The system will run by UN policy first for one year, then switch to SA which releases new wafers into system whenever the stepper is idle.
5. SEQ-SEQ Input Control (SS): The policy releases 8 lots of new wafers sequentially into system when the workload of the stepper is less than *32550* minutes for the case with mask changing time = *6* minutes or *30750* minutes for the case with mask changing time = *20* minutes. The lot size and workload is found by the algorithm constructed in the above section when assuming idle rate of each stepper is less than *10*% and $\delta_i = 1$ for all *i*.

The dispatch rule used with all the above input control method is the modified FIFO whose algorithm is elaborated as follows:

(1) A stepper is idle.
(2) Any hot lot? If no, assign the lot to this stepper and go to step (4). Else continue.
(3) Any lots in the queue waiting for this stepper? If no, go to step (6). Else continue.
(4) Any lots that need the same mask? If yes, assign the lots to this stepper and go to step (6). Else continue.
(5) Assign the lot with earliest arrival time to this stepper.
(6) Update the workload for this stepper. Assignment completes.

In order to compare these policies at the same ground, a constant busy rate of each stepper is assumed to be *98*% which includes the utilization, failure, repair, and regular maintenance. The measurement used in this study to evaluate the performance of SEQ-SEQ as well as other input policies is mean cycle time (in hours) which is defined as the average time a wafer stays in the system after it is released and right before it is finished. *10* runs of simulation are done for each input control policy with mask changing time = *6*- and *20*-minute respectively.

In both cases, the cycle time from SA is longer and than other policies as seen in Figures 6(a) and 6(b) while WIP and throughput from SA are very unstable compared with other policies as seen in Figures 7 and 8. It should be noted that SA performs worst among all policies in terms of cycle time because DRAM wafer fabrication process is very long and SA tends to saturate the system with releasing large number of lots at one time. Therefore, SA keeps a larger WIP than other policies as seen in Figure 7(a) and 7(b). However, the number of mask change by SS is less than other policies except SA as seen in Figure 8(a) and 8(b).

V.I One-way ANOVA

For each case of mask changing time, one-way ANOVA is performed to test the average throughput as well as mean cycle time and average number of mask change at *5*% significant level, α. All the test results are displayed in Tables 5and 6. For mean cycle time of the case with mask changing to be *6*

minutes, the null hypothesis is stated as follows.

H_0: The mean cycle time of the case with mask changing to be *6* minutes from these five input control policies are the same.

From Table 5(a), $F=2513.14$ which is greater than the critical value $F_{4,45;\sim=0.05}=2.57$. Therefore, H_0 is rejected and not all input control policies generate the same mean cycle time of the case with mask changing to be *6* minutes. The same conclusion applies to average monthly throughput and number of mask change of the case with mask changing time to be *6* minutes from Tables 5(b) and (c). It is also true that mean cycle time, average weekly throughput and number of mask change of the case with mask changing to be *20* minutes by 5 input control policies are not the same as seen in Tables 6(a)—(c).

V.II *t*-test

Because the results from one-way ANOVA are all rejecting H_0, *t* tests are performed to compare SS with each of the other four input control policies. These *t* tests are conducted in terms of mean cycle time, average monthly throughput, and average number of mask changes for the both cases with masks changing to be *6*- or *20*-minute at *5%* significant level. All these *t* test statistics are shown in Tables 7 and 8. For example, the null hypothesis to compare SS and UN with regard to the mean cycle time of the case with mask changing time to be *6* minutes is as follows:

H_0: The mean cycle time of the case with mask changing time to be *6* minutes from SS is the same as one from INIF.

From Table 7(a), $t=18.39$ which is larger than the critical value $t_{18;\alpha\ =0.05}=2.10$. Therefore, H_0 is rejected which implies that SS generates the smaller mean cycle time of the case with mask changing time to be *6* minutes than UN does. The same conclusion applies to average weekly throughput and average number of mask change of the case with mask changing time to be *6* minutes except compared with SA as seen in Table 7. However, the null hypothesis of cycle time is not rejected when compare SS with UN or FW for the case with masks changing time to be *20* minutes as shown in Table 8. The null hypotheses of average throughput and average number of mask change are rejected of the case with mask changing time to be *6* minutes except compared with SA as seen in Table 8.

It should be noted that the system is most unstable when adopting SA input policy as seen from the variances of throughput and cycle time. Therefore, SA should not be applied in the kind of system studied in this research. On the other hand, SS input control policy generates not only second best but also stable results and thus, could be adopted in a DRAM wafer fabrication plant.

VI. CONCLUSION

This research introduce a heuristic input control algorithm, called sequential-sequential Algorithm or SEQ-SEQ, which adopts special engineer requirements and can be easily implemented to help solving the new job releasing problem for steppers in the photolithography area of a DRAM wafer fabrication plant. Simulation is used to evaluate the performance of SEQ-SEQ as well as to compare SEQ-SEQ with other known input control methods such as Uniform Input, Close-Loop, and etc. An US-Taiwan wafer fabrication joint venture is used as an example in this study. As results of this work, SEQ-SEQ is constructed and proven statistically to be significantly better than other known input control policies in terms of average monthly throughput and mean cycle time. It is of future interest to include other work area into this study and to develop a combined production operation strategy for the whole wafer fabrication.

Reference

1. Adachi, T., J.J. Talavage and C.L. Moodie, “A Rule-based Control

Method for a Multi-loop Production Sytem," Artificial Intelligence in Engineering, Vol. 4, No. 3, pp.115-125, 1989.

2. Bai, X. and S.B. Gershwin, "A Manufacturing Scheduler's Perspective on Semiconductor Fabrication," VLSI Memo, Massachusetts Institute of Technology, pp.89—518, 1990.
3. Chang, S.C. and D.Y. Liao, "Scheduling Flexible Flow Shops with No Setup Effects," IEEE Transactions on Robotics and Automation, Vol. 10, No. 2, pp. 112—122, 1994.
4. Chen, H., J. M. Harrison, A. Mandelbaum, A.V. Ackere and L.M. Wein, "Empirical Evaluation of a Queueing Network Model for Semiconductor Wafer Fabrication," Operation Research, Vol. 36, No. 2, pp.202-215, 1988.
5. Chern, C.C., E. Hsu, and C.C. Shieh, "Time-Performance-Mask Algorithm (TPMA) for Photolithography and Diffusion Areas in a Wafer Manufacturing Plant," Proceeding of the 5th International Conference of the Decision Science Institute, Vol. 2, pp.1680—1682.
6. Chern, C.C., Y.L. Liu, and C.C. Shieh, "Stepper Dispatch Algorithm(SDA) for Photolithography Area in a Wafer Manufacturing Plant," Proceeding of the 1998 Annual Melting of the Decision Science Institute, Vol. 3, pp.1454—1458.
7. Dayhoff, J.E. and R.W. Atherton, "Simulation of VLSI manufacturing areas," VLSI Design, pp.84-92, December 1984.
8. Glassey, C.R. and R.G. Petrakian, "The Use of Bottleneck Starvation Avoidance with Queue Predictions," University of California, Berkeley, November 1989.
9. Glassey, C.R. and M.G.C. Resende, "Closed-Loop Job Release Control for VLSI Circuit Manufacturing," IEEE Transactions on Semiconductor Manufacturing, Vol. 1, No. 1, pp.36-46, February 1988.
10. Goldratt, E.M., "Computerized Shop Floor Scheduling," International Journal of Production Research, Vol. 26, No. 3, pp.443-455, 1988.
11. Graves, S.C., "A Review of Production Scheduling," Operations Research, Vol. 29, No. 4, pp.646-675, 1981.
12. Hughes, R.A. and J.D. Schott, "The Future of Automation for High Volume Wafer Fabrication and ASIC Manufacturing," Proceedings of the IEEE, Vol. 74, No. 12, pp.1775—1793, 1986.
13. Johri, P.K., "Practical Issues in Scheduling and Dispatching in Semiconductor Wafer Fabrication," Journal of Manufacturing Systems, Vol 12, No. 6, pp.474—485, 1992.
14. Koulamas, C., "The Total Tardiness Problem: Review And Extensions," Operations Research, Vol. 42, No. 6, pp. 1025—1041, 1994.
15. Kumar, P.R., "Scheduling Semiconductor Manufacturing Plants," IEEE Control System, pp.33-40, December 1994.
16. Lawton, J.W., L.M. Wein, R. Whitney and D. Zuanich, "Workload Regulating Wafer Release in a GaAs Fab Facility," Int'l Semiconductor Manufacturing Science Symposium, pp.33-38, 1990.
17. Leachman, R.C. and D.A. Hodges, "Benchmarking Semiconductor Manufacturing," IEEE Transactions on Semiconductor Manufacturing, Vol. 9, No. 2, pp.156-169, May 1996.
18. Levitt, M.E. and J.A. Abraham, " Just-In-Time Methods for Semiconductor Manufacturing," IEEE/SEMI Advanced Semiconductor Manufacturing Conference, pp.3-9.
19. Lin, H.C. and C.S. Raghavendra, "Approximating the mean response time of parallel queues with JSQ policy," Computer Ops Res. Vol. 23,

No. 8, pp.733-740, 1996.
20. Lozinski, C. and C.R. Glassey, "Bottleneck Starvation Indicator for Shop Floor," IEEE Transactions on Semiconductor Manufacturing, Vol. 1, No. 4, pp.147—153, November 1988.
21. Lu, S.H. and P.R. Kumar, "Distributed Scheduling Based on Due Dates and Buffer Priorities," IEEE Transactions on Automatic Control, Vol. 36, No. 12, pp.1406-1416, December 1991.
22. Miller, D.J., "Simulation of a Semiconductor Manufacturing Line," Communications of the ACM, Vol. 33, pp.98—108, 1990.
23. Panwalkar, S.S. and W. Iskander, "A Survey of Scheduling Rules," Operations Research, Vol. 25, No. 1, pp.45—64, 1977
24. Schragenheim, E. and B. Ronen, "Drum-Buffer-Rope Shop Floor Control," Production and Inventory Management Journal, 3rd Quarter, pp.18-22, 1990.
25. Spence, A.M. and D.J. Welter, "Capacity Planning of a Photolithography Work Cell in a Wafer Manufacturing Line," Proceedings of the IEEE Conference on Robotics and Automation, pp. 702—708, 1987.
26. Uzsoy, R., C.Y. Lee and L.A. Martin-Vega, "A Review of Production Planning and Scheduling Models in the Semiconductor Industry. Part I: System Characteristics, Performance Evaluation and Production Planning," IIE Transactions, Vol. 24, No. 4, pp.47-60, September 1992.
27. Uzsoy, R., C.Y. Lee and L.A. Martin-Vega, "A Review of Production Planning and Scheduling Models in the Semiconductor Industry. Part II: Shop-Floor Control," IIE Transactions, Vol. 26, No. 5, pp.44-55, September 1994.
28. Wein, L.M., "Scheduling Semiconductor Wafer Fabrication," IEEE Transactions on Semiconductor Manufacturing, Vol. 1, No. 3, pp.202-215, August 1988.

Figure 1: Basic Flow of a DRAM wafer fab

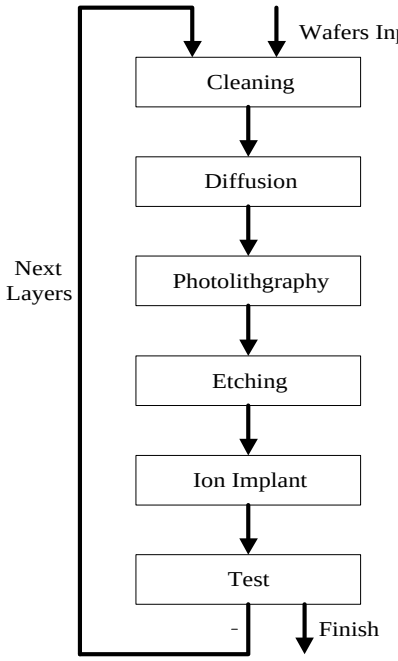


Figure 2(a): Model of SEQ-SEQ for k

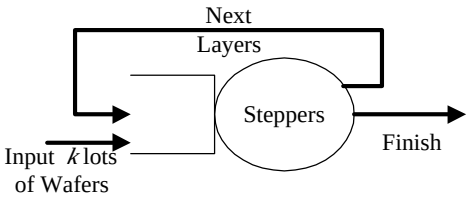


Figure 2(b): Workload Control mechanism of SEQ-SEQ

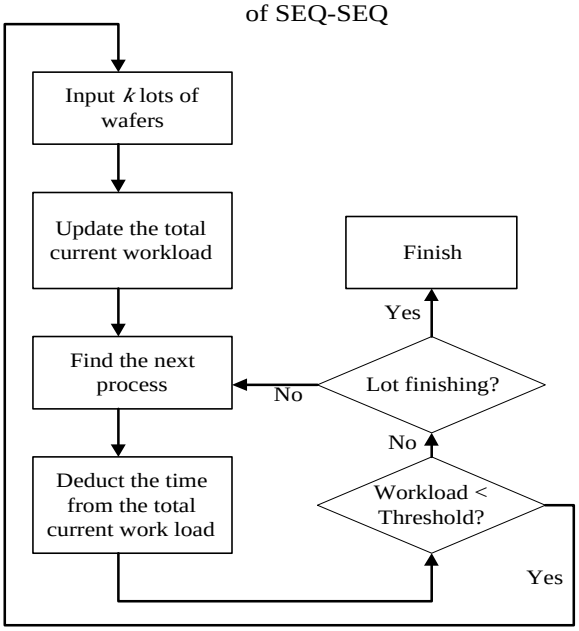


Figure 5(a): Warm-Up Period from WIP

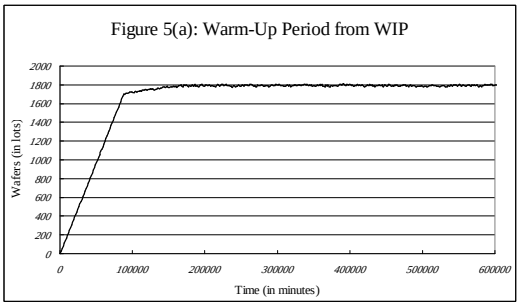


Figure 5(b): Warm-Up Period from Utilization

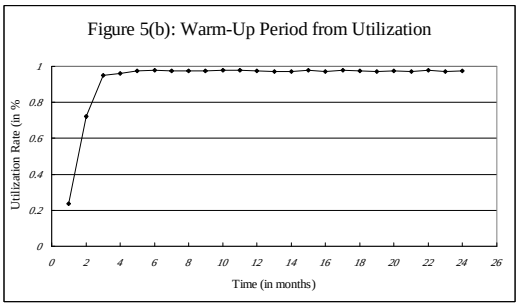


Figure 3(a): The relationship between time and wafer lot sizes

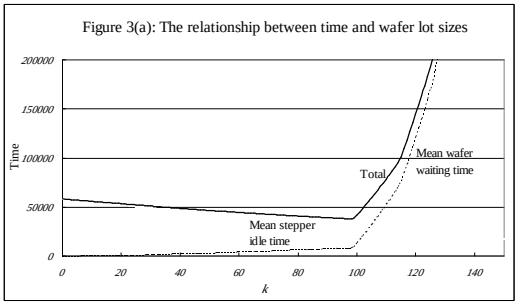


Figure 3(b): Relationship between workload and time (for $k = 10$)

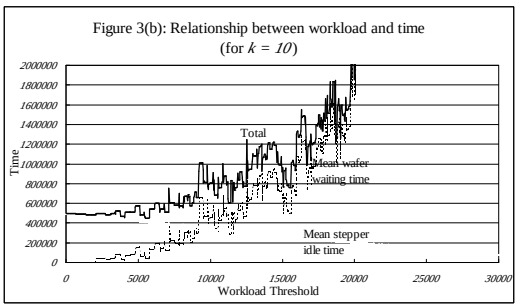


Figure 4: Model Details of the Photolithgraphy Area

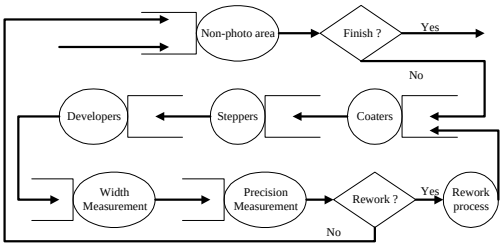


Figure 6(a): Cycle Time for mask changing time = 6 minutes

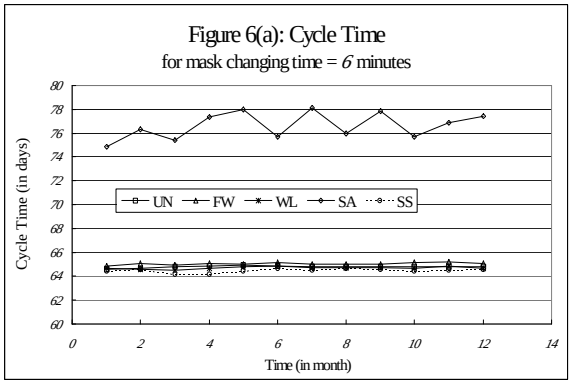
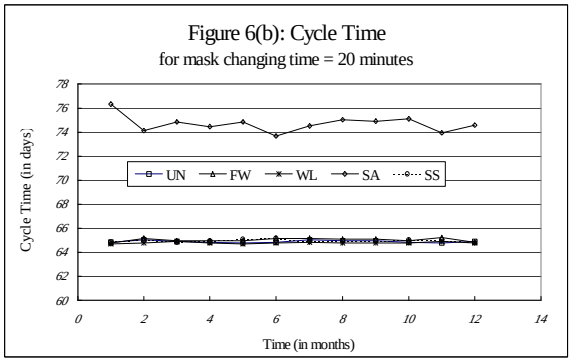
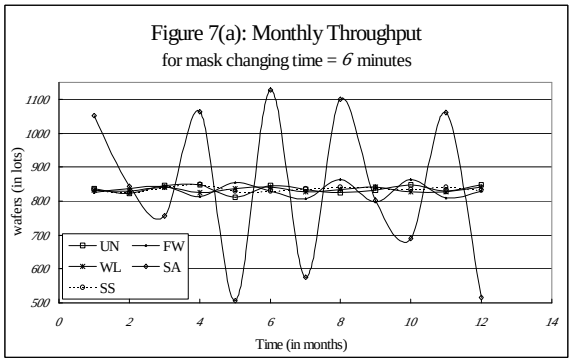
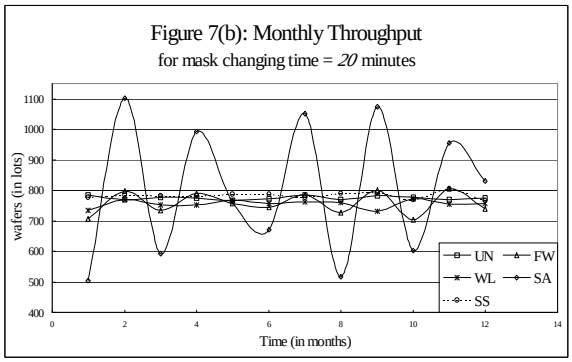
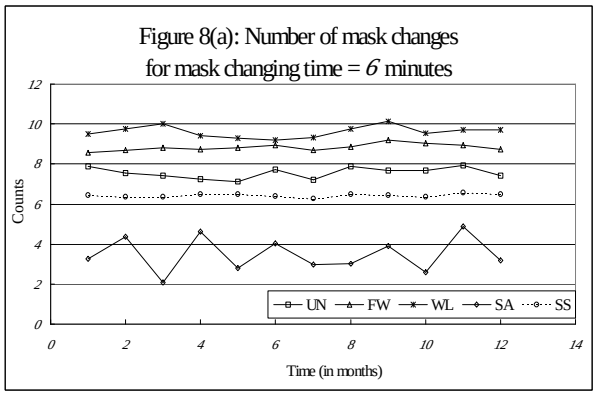


Figure 6(b): Cycle Time for mask changing time = 20 minutes









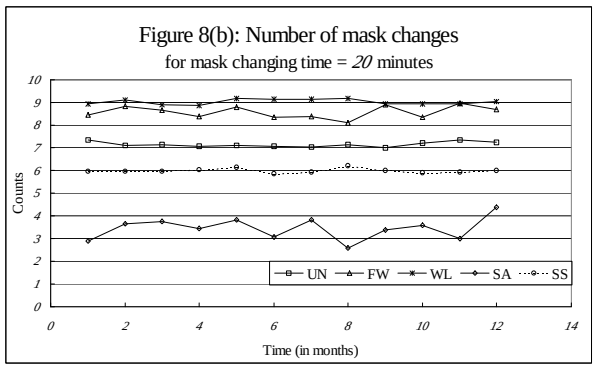


Table 1 : Times of operations i on Stepper and other work centers for all lot (in minutes)

Critical Operation i	Coater	Stepper	Developer	Width Measurement	Precision Measurement	Non-lithography area
0						*14400*
1	*35*	*34*	*36*	*9*	*6*	*7200*
2	*35*	*31*	*36*	*9*	*6*	*7560*
3	*35*	*32*	*36*	*9*	*6*	*10080*
4	*35*	*33*	*36*	*9*	*6*	*10560*
5	*35*	*32*	*36*	*9*	*6*	*3210*
6	*35*	*35*	*36*	*9*	*6*	*5050*
7	*35*	*35*	*36*	*9*	*6*	*3890*
8	*35*	*35*	*36*	*9*	*6*	*5350*
9	*35*	*34*	*36*	*9*	*6*	*4220*
10	*35*	*34*	*36*	*9*	*6*	*14760*

Table 2 :

Rework Rate (in %)

Operation i	Rate %
1	*1.0*
2	*1.2*
3	*3.2*
4	*1.2*
5	*4.0*
6	*2.2*
7	*1.1*
8	*1.6*
9	*0.6*
10	*2.9*

Table 3(a): Machine failures of Steppers (in hours)

Stepper	Mean time between failures	Mean repair time
1	*145*	*1.3*
2	*388*	*0.5*
3	*211*	*0.6*
4	*138*	*0.3*
5	*482*	*0.1*
6	*63*	*2.2*
7	*96*	*0.6*

Table 3(b): Machine failures of Coaters (in hours)

Coater	Mean time between failures	Mean repair time
1	*260*	*1.5*
2	*260*	*1.5*
3	*134*	*1*
4	*197*	*1.6*
5	*210*	*0.9*
6	*329*	*0.7*
7	*211*	*0.1*
8	*480*	*4.6*

Table 3(c): Machine failures of Developers (in hours)

Developer	Mean time between failures	Mean repair time
1	*652*	*0.7*
2	*979*	*0.7*
3	*650*	*0.9*
4	*215*	*0.8*
5	*490*	*0.8*
6	*647*	*0.3*
7	*993*	*0.4*
8	*152*	*0.4*

Table 4 : Regular Maintenance of work units in Photolithograph Area

Work units	Type of regular maintenances	Interval (in days)	Repair Time (in minutes)
Coater	*1*	*60*	*120*
Developer	*1*	*60*	*120*
Measurer for width	*1*	*1*	*10*
	2	*30*	*60*
	3	*180*	*480*
Measurer for precisio	*1*	*7*	*20*
	2	*30*	*60*
Stepper	*1*	*30*	*84*
	2	*90*	*150*
	3	*180*	*240*

Table 5(a) : Mean Cylce Time
(Mask changing time = *6* minutes)

Sources	*df*	SS	MS	F
Input Control Policies	*4*	*997.89*	*249.47*	*2513.14*
Error	*45*	*4.47*	*0.10*	
Total	*49*	*1002.36*		

Table 5(b) : Average Monthly Throughput
(Masking changing time = *6* minutes)

Sources	*df*	SS	MS	F
Input Control Policies	*4*	*894.48*	*223.62*	*13.66*
Error	*45*	*736.51*	*16.37*	
Total	*49*	*1630.99*		

Table 5(c) : Average Number of Mask Changes
(Masking changing time = *6* minutes)

Sources	*df*	SS	MS	F
Input Control Policies	*4*	*243.59*	*60.90*	*6275.16*
Error	*45*	*0.44*	*0.01*	
Total	*49*	*244.03*		

Table 6(a) : Mean Cylce Time
(Mask changing time = *20* minutes)

Sources	*df*	SS	MS	F
Input Control Policies	*4*	*642.75*	*160.69*	*1704.94*
Error	*45*	*4.24*	*0.09*	
Total	*49*	*647.00*		

Table 6(b) : Average Monthly Throughput
(Masking changing time = *20* minutes)

Sources	*df*	SS	MS	F
Input Control Policies	*4*	*21725.69*	*5431.42*	*97.28*
Error	*45*	*2512.45*	*55.83*	
Total	*49*	*24238.14*		

Table 6(c) : Average Number of Mask Changes per day
(Masking changing time = *20* minutes)

Sources	*df*	SS	MS	F
Input Control Policies	*4*	*196.08*	*49.02*	*12177.69*
Error	*45*	*0.18*	*0.004*	
Total	*49*	*196.26*		

ps. Critical F value to reject H_0 is $F^*_{4,45;0.95}=2.57$

Table 7(a) : t test for Mean Cycle time (in days)
(Mask changing time = *6* minutes)
SS: mean=*64.42* and Stdev=*0.044*

Input Control Policies	Mean	Stdev	t value	Reject H_0
UN	*64.86*	*0.06*	*18.39*	Yes
FW	*65.00*	*0.06*	*36.52*	Yes
WL	*64.61*	*0.04*	*9.46*	Yes
SA	*75.88*	*0.70*	*51.03*	Yes

Table 7(b) : t test for Average Monthly Throughput
(Mask changing time = *6* minutes)
SS: mean=838.66 and Stdev=*0.78*

Input Control Policies	Mean	Stdev	t value	Reject H_0
UN	*835.72*	*1.02*	*6.79*	Yes
FW	*833.08*	*1.92*	*7.45*	Yes
WL	*832.26*	*0.75*	*31.54*	Yes
SA	*843.90*	*8.72*	*7.97*	No

Table 7(c) : t test for Average Number of Mask Changes per day (Mask changing time = *6* minutes)
SS: mean=*6.45* and Stdev=0.04

Input Control Policies	Mean	Stdev	t value	Reject H_0
UN	*7.51*	*0.07*	*46.47*	Yes
FW	*8.93*	*0.11*	*72.86*	Yes
WL	*9.85*	*0.11*	*98.30*	Yes
SA	*3.50*	*0.14*	*-66.57*	Yes

Table 8(a) : t test for Mean Cycle Time (in days)
(Mask changing time = *20* minutes)
SS: mean=*64.99* and Stdev=*0.04*

Input Control Policies	Mean	Stdev	t value	Reject H_0
UN	*64.96*	*0.10*	*1.30*	No
FW	*65.01*	*0.04*	*1.47*	No
WL	*64.84*	*0.04*	*9.64*	Yes
SA	*73.91*	*0.68*	*41.43*	Yes

Table 8(b) : t test for Average Monthly Throughput
(Mask changing time = *20* minutes)
SS: mean=784.86 and Stdev=*0.95*

Input Control Policies	Mean	Stdev	t value	Reject H_0
UN	*775.82*	*1.22*	*26.81*	Yes
FW	*759.86*	*2.11*	*33.43*	Yes
WL	*755.75*	*1.17*	*77.76*	Yes
SA	*814.01*	*16.46*	*-5.55*	Yes

Table 8(c) : t test for Average Number of Mask Changes per day (Mask changing time = 20 minutes)
SS: mean=*6.00* and Stdev=0.03

Input Control Policies	Mean	Stdev	t value	Reject H_0
UN	*7.06*	*0.09*	*35.60*	Yes
FW	*8.53*	*0.04*	*148.98*	Yes
WL	*8.98*	*0.04*	*223.19*	Yes
SA	*3.46*	*0.09*	*-91.50*	Yes

ps. Critical t value to reject H_0 is $t^*_{18;0.05}$=*-2.1 or* H_0 *is* $t^*_{18;0.95}$ =*2.1*