Journal of the Korean
OR/ MS Society
Vol. 18, No. 1, April 1993

A Send-ahead Policy for a Semiconductor Wafer Fabrication Process

Ilkyeong Moon*

Abstract

We study a manufacturing process that is quite common in semiconductor wafer fabrication of semiconductor chip production. A machine is used to process a job consisting of J wafers. Each job requires a setup, and the i_{th} setup for a job is successful with probability p_i. The setup is prone to failure, which results in the loss of expensive wafers. Therefore, a trial run is first conducted on a small batch. If the set up is successful, the test is passed and the balance of the job can be processed. If the setup is unsuccessful, the exposed wafers are lost to scrap and the mask is realigned. The process then repeats on the balance of the job. We call this as send-ahead policy and consider general policies in which the number of wafers that are sent ahead depend on the cost of the raw wafer, the sequence of success probabilities, and the balance of the job. We model this process and determine the expected number of good wafers per job, the expected time to process a job, and the long run average throughput. An algorithm to minimize the cost per good wafer subject to a demand constraint is provided.

1. Introduction

This paper is concerned with a manufacturing process that is quite common in semiconductor wafer fabrication of integrated circuit (IC) or chip production. The production of semiconductor chips is accomplished in several stages that begins with raw wafers of silicon. Refer to [1,2,3,5, 6,7,8] for a more detailed description of semiconductor chip manufacturing. Wafers are grouped in lots (or jobs), the members of which travel together in a standard container and are destined for conversion to the same final product. The lot size, usually between 20 and 100 wafers, differs from one product to another within the same facility.

We concentrate on a photolithography process at the first stage of IC production, which is

* Department of Industrial Engineering, Pusan National University

called as wafer fabrication. Coming into the photolithography process, a wafer is coated with a light sensitive material called photoresist, which is then exposed to a ultraviolet light through a mask that reflects the designed circuity to be built on to the wafer. The exposure step is one of the most sensitive and failure prone steps in wafer fabrication, and most of the scraps take place. If the test shows that the designed circuit exposed in the photoresist does not meet design specifications (due to, for example, misalignment of the mask), then the photoresist layer will be stripped away and the above process must be repeated. To reduce the number of scrapped wafers, it is a common practice to conduct trial runs involving small batches. We can describe above process as follows: A machine is used to process a job (or lot) consisting of J wafers. Each job requires a setup that consists of aligning a mask. The i_{th} setup for a job is successful with probability p_i and unsuccessful with probability $1-p_i$. For example, if p_i is increasing in i, it indicates the effect of learning. The setup is prone to failure, which results in the loss of expensive wafers. Therefore, a trial run is first conducted on a small batch (say, one or two units). After the setup, the batch is exposed and developed while the balance of the job waits until a test is performed several stages later. If the set up is successful, the test is passed and the balance of the job is then exposed and developed. If the setup is unsuccessful, the exposed wafers are lost to scrap and the mask is realigned. The process then repeats on the balance of the job. We call this as send-ahead policy. Here we consider general policies in which the number of wafers that are sent ahead depend on the cost of the raw wafer, the sequence of success probabilities, and the balance of the job.

We model this process and determine for every policy the expected number of good wafers per job, the expected time to process a job, and the long run average throughput. Using this information we develop an algorithm to minimize the cost per good wafer subject to a lower bound on the long run average throughput. We show that the form of the optimal policy is data independent and consists of sending ahead one wafer at a time up to the k_{th} setup when the balance of the job $J-k+1$ is exposed and developed. We also consider the case of different job types. We address the problem of maximizing the expected weighted sum of good wafers subject to an upperbound on the expected time to process a weighted sum of jobs.

2. Preliminaries and Optimality of the k-split Policy

Data

J : job size (number of wafers that belong to a job).

s : set-up time for the machine.

t : processing time of the machine per wafer.

p_i : probability of success for the i_{th} setup for a job.

Let $\delta(y, x)$ be the expected waiting time for the balance of a job size y when $x<y$ wafers are sent ahead. Define $\delta(y,y)\equiv 0$. We assume that $\delta(y,x)$ is non-decreasing in $x<y$ and independent of y. Denote $\delta(y, 1)$ by δ for all $y>1$.

For fixed $k(1\leq k\leq J)$, consider a policy that allows k setups. We denote the policy as a k-setup policy. The set of these policies consists of k positive integers x_i, $i=1,\cdots,k$ adding up to J. By a k-split policy we mean a k-setup policy for which $x_i=1$, $i\leq k-1$ and $x_k=J-k+1$. We now compute the expected time to process a job and the expected number of good wafers for k-setup policies. Next we show that the k-split policy minimizes the expected time to process a job among all k-setup policies. We also show that the k-split policy maximizes the expected number of good wafers among all k-setup policies.

Let $T[x_1,\cdots,x_k; J]$ be the time to process a job of size J for a k-setup policy. Similary let $G[x_1,\cdots,x_k; J]$ be the number of good wafers out of J for a k-setup policy. Define $\beta_i=\prod_{j=1}^{i}(1-p_j)$ for $i=1,\cdots,k$.

Property 1. $ET[x_1,\cdots,x_k; J]$ is minimized by the k-split policy.

Proof.

$ET[x_1,\cdots,x_k; J]$

$=E$[time to process a job of size J | 1st setup is fail]P[1st setup is fail]+

E[time to process a job of size j | 1st setup is successful]P[1st setup is successful]

$$=s+tx_1+\delta(J, x_1)+p_1 t(J-x_1)+(1-p_1)ET[x_2,\cdots,x_k; J-x_1]$$

$$=s+tx_1+\delta(J, x_1)+p_1 t(J-x_1)+(1-p_1)[s+tx_2+\delta(J-x_1, x_2)+p_2 t(J-x_1-x_2)]$$

$$+(1-p_1)(1-p_2)[s+tx_3+\delta(J-x_1-x_2, x_3)+p_3 t(J-x_1-x_2-x_3)]+\cdots$$

$$+(1-p_1)\cdots(1-p_{k-2})[s+tx_{k-1}+\delta(J-x_1-\cdots-x_{k-2}, x_{k-1})+p_{k-1} t(J-x_1-\cdots-x_{k-1})]$$

$$+(1-p_1)\cdots(1-p_{k-1})[s+t(J-x_1-\cdots-x_{k-1})]$$

$$=tJ+s[1+(1-p_1)+(1-p_1)(1-p_2)+\cdots(1-p_1)(1-p_2)\cdots(1-p_{k-1})]$$

$$+\delta(J, x_1)+(1-p_1)\delta(J-x_1, x_2)+\cdots+(1-p_1)(1-p_2)\cdots(1-p_{k-2})\delta(J-x_1-\cdots-x_{k-2}, x_{k-1})$$

$$=tJ+s\left(\sum_{i=0}^{k-1}\beta_i\right)+\sum_{i=1}^{k-1}\beta_{i-1}\delta\left(J-\sum_{l=1}^{i-1}x_l, x_i\right).$$

Since $ET[x_1,\cdots,x_k; J]$ is independent of x_k and β_i's are positive, the largest x_k, i.e. $x_k=J-k+1$ minimizes $ET[x_1,\cdots,x_k; J]$. ■

Property 2. $EG[x_1,\cdots,x_k;\ J]$ is maximized by the k-split policy.

Proof.

$$EG[x_1,\cdots,x_k;\ J]=p_1x_1+p_1(J-x_1)+(1-p_1)\ EG[x_2,\cdots,x_k;J-x_1]$$

$$=P_1J+(1-p_1)p_2(J-x_1)+(1-p_1)(1-p_2)p_3(J-x_1-x_2)+\cdots$$

$$+(1-p_1)\cdots(1-p_{k-2})p_{k-1}(J-x_1-\cdots-x_{k-2})+(1-p_1)\cdots(1-p_{k-1})p_k(J-x_1-\cdots-x_{k-1})$$

$$=J[1-\textstyle\prod_{j=1}^{k}(1-p_j)]-x_1(1-p_1)[1-\prod_{j=2}^{k}(1-p_j)]-\cdots$$

$$-x_{k-1}(1-p_1)\cdots(1-p_{k-1})[1-\textstyle\prod_{j=k}^{k}(1-p_j)]$$

$$=J(1-\beta_k)-\sum_{i=1}^{k-1}(1-\frac{\beta_k}{\beta_i})\ \beta_i x_i$$

$$=J-\sum_{i=1}^{k}\beta_i x_i.$$

Since β_i's are decreasing in k, the largest x_k, i.e., $x_k=J-k+1$ maximizes $EG[x_1,\cdots,x_k;\ J]$. ■

Remark. One industrial process that we are familiar with used two units in each trial, it is hence of interest to know how does this compare with the single unit trials (k-split policy) for fixed k. Let $EB[1,\cdots,1,J-k+1;J]$ and $EB[2,\cdots,2,\ J-2k+2;J]$ donote the expected number of bad units under the two policies, respectively, for a given $k\le\frac{J}{2}$. From Property 2, it easy to derive the following equation:

$$EB[2,\cdots,2,\ J-2k+2;J]-J\beta_k=2(EB[1,\cdots,1,\ J-k+1;\ J]-J\beta_k).$$

Since $\beta_k=(1-p_1)(1-p_2)\cdots(1-p_k)$ is typically very small, the above equation implies that the expected number of bad units under the two-unit trials is almost twice as much as that under the k-split policy. ■

Let $ET(J,k)$=ET$[1,\cdots,1,\ J+1-k;J]$ be the expected time to process a job for the k-split policy, $k=1,\cdots,J$. Similarly let $EG(J,\ k)=EG[1,\cdots1,\ J+1-k;\ J]$ be the expected number of good wafers for the k-split policy.

Property 3. $ET(J,\ k)$ is increasing concave.

Proof. For all $1\le k<J$,

$$\triangle ET(J,k) = ET(J,k+1) - ET(J,k)$$
$$= tJ + s\left(\sum_{i=0}^{k} \beta_i\right) + \sum_{i=1}^{k} \beta_{i-1}\, \delta\left(J - \sum_{l=1}^{i-1} 1,\ 1\right)$$
$$- tJ - s\left(\sum_{i=0}^{k-1} \beta_i\right) - \sum_{i=1}^{k-1} \beta_{i-1}\, \delta\left(J - \sum_{l=1}^{i-1} 1,\ 1\right)$$
$$= s\beta_k + \beta_{k-1}\,\delta > 0.$$
$$\triangle^2 ET(J,k) = \triangle ET(J,k+1) - \triangle ET(J,k) = -(s\beta_k p_{k-1} + \delta\beta_{k-1} p_k) < 0. \quad ■$$

Property 4. $EG(J,k)$ is increasing in k. Moreover $EG(J,k)$ is concave if either (1) $p_i = p$ for all i (2) p_i is decreasing (3) $\frac{1}{2} \le p_i$ is increasing.

Proof. For all $1 \le k < J$,

$$\triangle EG(J,k) = EG(J,k+1) - EG(J,k)$$
$$= J - \sum_{i=1}^{k} \beta_i - \beta_{k+1}(J-k) - \left[J - \sum_{i=1}^{k-1} \beta_i - \beta_k(J-k+1)\right]$$
$$= (\beta_k - \beta_{k+1})(J-k) > 0$$

since $J > k$ and β_k's are decreasing in k.

$$\triangle^2 EG(J,k) = \triangle EG(J,k+1) - \triangle EG(J,k)$$
$$= -\beta_k \{[J-k][p_{k+1} - p_{k+2}(1-p_{k-1})] + p_{k+2}(1-p_{k-1})\}.$$

Hence, it is enough to show that

$$p_{k+1} - p_{k+2}(1 - p_{k+1}) \ge 0.$$

This is obvious for (1) and (2) above. For (3) note that

$$p_{k+2}(1-p_{k+1}) \le \frac{1}{2} p_{k+2} \le \frac{1}{2} < p_{k+1}. \quad ■$$

3. Minimizing cost per good wafer subject to a demand constraint

Let G_n be the number of good wafers in the n_{th} job and let T_n be the time to process the n_{th} job. Since G_n, T_n are i.i.d. and $P_r\{G_n>0\} > 0$ and $P_r\{T_n>0\} > 0$, the reward renewal theorem applies and the long run average throughput is $\frac{EG}{ET}$. Here, we want to minimize the cost per good wafer subject to the condition that long run average throughput is large enough to satisfy demand rate, say d. Let c be the cost per wafer For fixed k,

P1

$$Min\ cJ/EG[x_1,\cdots,x_k;J]$$
$$s.t\quad EG[x_1,\cdots,x_k;J] \geq dET[x_1,\cdots,x_k;J].$$

Equivalently the objective in P1 can be replaced by $Max\ EG[x_1,\cdots,x_k;J]$.

From Property 1 and Property 2, we get the following obvious Property.

Property 5. Given k, the k-split policy is optimal for P1 if it is feasible. Else, there is no feasible k-setup policy.

From the above it is clear that the k-split policy is optimal among all send-ahead policies. From Property 3 and Property 5, we get the following algorithm that finds the optimal value of k for P1.

Algorithm

Step 1. Set $k=J$ and $x_i=1$ for $i<k$, $x_k=J-k+1$
If it is feasible, it is overall optimal. Stop.
Else, go to *Step 2.*

Step 2. $k \leftarrow k-1$ and $x_i=1$ for $i<k$, $x_k=J-k+1$.
If it is feasible, stop. Else go to *Step 3.*

Step 3. If $k=1$, stop. This problem is infeasible
Else go to *Step 2.*

Example. J=18, s=5, t=0.5, p=(0.30, 0.32, 0.33, 0.35, 0.36, 0.38, 0.40, 0.40, 0.50, 0.55, 0.56, 0.60, 0.70, 0.75, 0.77, 0.78, 0.79, 0.80), $\delta(y, x)=5+.\cdot$ *if* $x<y$, d=0.38 After applying above algorithm, we get $k^*=8$. Then, the optimal number of trial runs is 7 and $x_i^*=1$ for $i<8$ and $x_8^*=11$, $EG[x_1,\cdots,x_8;J]=15.491$ and the long run average throughput becomes 0.38. ■

4. Maximizing expected weighted sum of good wafers subject to a processing time constraint.

Now we consider the multiple job types. Consider the program

P2

$$Max_{(k_1, \ldots, k_N)} \sum_{i=1}^{N} \pi_i EG_i(J_i, k_i)$$

$$s.t \sum_{i=1}^{N} b_i ET_i(J_i, k_i) \leq M$$

For example, if management has decided to process b_i jobs of type i and the profit per good wafer is π_i, then above program maximizes the expected profit subject to a restriction on the total expected time to process all the jobs. We provide a heuristic algorithm based on the marginal allocation algorithm (see [4] for the details of the algorithm). Since $ET_i(J_i, k_i)$'s are concave function, the algorithm cannot guarantee to produce an optimal solution. See Fox [4] for the conditions for which the marginal allocation algorithm generates an optimal solution.

Marginal Allocation Heuristic

Step 1. Start with $\underline{k}^0 = \underline{e}$. Set $j=1$.

Step 2. $\underline{k}^j = \underline{k}^{j-1} + \underline{e}_i$, where $\underline{e}_i$, is the i_{th} unit vector and i is the index for which $\pi_i \triangle EG_i(J_i, k_i^{j-1})/b_i \triangle ET_i(J_i, k_i^{j-1})$ is maximum.

Step 3. If $\sum_i b_i ET_i(J_i, k_i^j) > M$, terminate. We get a solution $\underline{k}^{j-1}$. Else $j \leftarrow j+1$ and go to *Step 2.*

5. Conclusion

We have studied a photolithography process in the semiconductor chip production. To reduce the number of scrapped wafers, it is a common practice to conduct trial runs involving small batches. The policy has been called as send-ahead policy. We have derived the expected number of good wafers per job, the expected time to process a job, and the long run average throughput for the policy. The k-split policy is shown to be an optimal policy among all send-ahead policies. We have developed an algorithm to minimize the cost per good wafer subject to a lower bound on the long run average throughput. The problem of maximizing the expected weighted sum of good wafers subject to a processing time constraint has been also addressed.

There are several interesting areas related to semiconductor manufacturing processes in which operations research techniques can be applied. Production planning models for analyzing the system characteristics of various wafer fabrication processes, scheduling the electronic fabrication facilities, and wafer design problems arising from the fabrication process of printed circuit boards and integrated circuits are some of the areas which need intensive investigations. Refer to [1,2,3,5,6,7,8] for some the research in this field.

References

[1] Bitran, G. and D. Tirupati, "Planning and Scheduling for Epitaxial Wafer Production Facilities," *Operations Research*, Vol. 36, No. 1 (1988), pp.34-49.

[2] Bitran, G. and D. Tirupati, "Development and Implementation of a Scheduling System for a Wafer Fabrication Facility," *Operations Research*, Vol. 36, No. 3 (1988), pp.377-395.

[3] Chen, H., M. Harrison, A. Mandelbaum, A. Van Ackere, and L. Wein, "Emipirical Evaluation of a Queueing Network Model for Semiconductor Wafer Fabrication," *Operations Research*, Vol. 36, No. 2 (1988), pp.202-215.

[4] Fox, B., "Discrete Optimization Via Marginal Anaysis," *Management Science*, Vol. 13, No. 3 (1966), pp.210-216.

[5] Lee, C., R. Uzsoy, and L. Martin-Vega, "Efficient Algorthms for Scheduling Semiconductor Burn-In Operations," *Operations Research*, Vol. 40, No. 4 (1992), pp.764-775.

[6] Tang, C., "Composing Batches with Yield Uncertainty," Working Paper, Graduate School of Management, UCLA (1990).

[7] Uzsoy,R., C.Lee, and L.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(1992), pp.47-60.

[8] Wein L., "Scheduling Semiconductor Wafer Fabrication," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 1, No. 1 (1988), pp.115-129.