Simulation Analysis of Scheduling Rules in MEMS Manufacturing

LIXIN WANG
Gintic Institute of Manufacturing Technology
Nanyang Technological University
71 Nanyang Drive, Singapore 638075
SINGAPORE

Abstract: - This paper focuses on the production scheduling in MEMS (microelectromechanical system) manufacturing. The whole MEMS production process can be organized as three sub-processes, that is, the front-end process, the wafer cap process and the back-end process. As the total processing time of the front-end process is significantly longer that that of the wafer cap process, the coordination between the releases of wafers to these two sub-processes is an important issue. In this paper, three synchronization rules are developed and they are evaluated together with six dispatching rules under Poisson input. The performance measures considered are cycle times of wafers. A visual interactive simulation model is constructed to imitate the production line. The simulation results indicate that both synchronization rules and dispatching rules have significant impacts on the performance of MEMS manufacturing.

Key-Words: - MEMS, Scheduling, Synchronization, Dispatching, Simulation, Cycle Time

1 Introduction

MEMS (Micro-Electro Mechanical Systems) are integrated micro devices or systems combining electrical and mechanical components fabricated using integrated circuit (IC) compatible batch processing techniques. The MEMS fabrication process studied in this paper is based on a commercial SCREAM (single crystal reactive etching and metallization) micromachining technology. This technology uses reactive ion etching both to define and release structures[1]. The production process can be organized as three sub-processes, namely, the front-end process, the wafer cap process and the back-end process (see Fig.1). Raw wafers are initially released to the two parallel sub-processes, the front-end process and the wafer cap process, with a batch size of eighteen wafers. Then they are processed in these two sub-processes with this batch size concurrently. Every wafer that has undergone all the operations in the front-end process will be bonded with a wafer which has completed all the operations in the wafer cap process at the bonding machine. The bonded wafer then continues its processing in the back-end process individually.

In the wafer front-end process, there are 106 steps and the sum of processing time is 63.2 hours. While in the wafer cap process, they are only 24 steps and 14.6 hours respectively. Moreover, there are several workstations which are shared by both the front-end process and the wafer cap process. Since the output wafers from these two sub-processes will go through the bonding workstation, one of the critical issues is how to synchronize the releases of these two sub-processes. If the synchronization problem is not properly managed, the cycle time (the time from the release of the raw

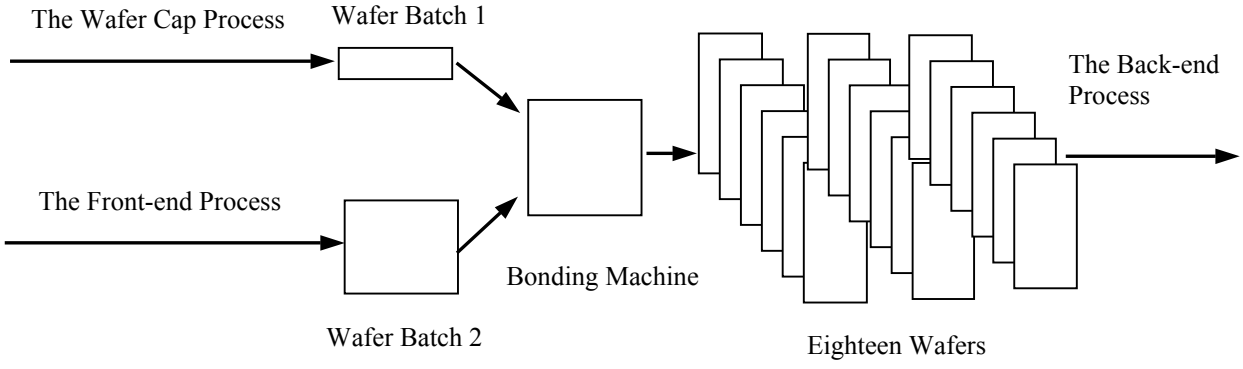

Figure 1 Schematic View of the Whole MEMS Process

material to the production line until it comes out) of the product will become longer.

According to the authors' knowledge, there is currently no publication considering the synchronization problems in MEMS manufacturing. However, we can find some studies on the similar scheduling problems. As described above, MEMS fabrication processes combine IC and micromaching processes. Consequently, the process flows and the equipment used in MEMS fabrication are very similar to those for semiconductor manufacturing. Studies point out that semiconductor manufacturing is one of the world's most complicated manufacturing processes [2][3]. Scheduling in semiconductor manufacturing is always a very tough issue. There have been a lot of studies in this area [4][5][6][7][8]. Due to the similarities between MEMS manufacturing and semiconductor manufacturing, some of the scheduling rules and research methods can also be applied in MEMS manufacturing

The purpose of this study is to develop scheduling rules to reduce cycle times of the wafers. Three synchronization rules are presented and they are used with six dispatching rules to evaluate the performance of the MEMS manufacturing under Poisson input rule. Since the production flow in MEMS is very complicated, a discrete event simulation model is built to imitate its process flow

The remainder of this paper is organized as follows. Section 2 describes synchronization rules and dispatching rules used in the MEMS manufacturing and the simulation experiments are provided in section 3. In section 4, the results of the simulation study are presented and discussed, and the conclusions of the study are contained in section 5.

2 Scheduling Rules

Owing to its complexity, three types of rules are considered for production scheduling and control in this MEMS manufacturing system. They are release rules (also known as input rules), synchronization rules, and dispatching rules. Release rules dictate the release of raw wafers to the process. Synchronization rules are new rules developed in this paper to release raw wafers to the process in coordination with release rules. Since the front-end process is the main part of the whole MEMS process and it takes longer time to complete the process, it will be more manageable and simpler to apply synchronization rules to control the release mechanism of the wafer cap process.

In this paper, Poisson input is applied to release new wafers into the front-end process which is also one of the most extensively studied input rule in semiconductor manufacturing systems.

2.1 Dispatching Rules

The First In First Out (FIFO) dispatching rule and Shortest Remaining Processing Time (SRPT) rule are used in this study because they are widely used in practice. In addition, the author developed another four rules, namely, CAPFIFO, FRONTFIFO, CAPSRPT and FRONTSRPT by giving processing priority to the wafers in the front-end process or those in the wafer cap process at the workstations shared by these two sub-processes. E.g., in CAPFIFO rule, FIFO is utilized for all the workstations except the workstations shared by both the front-end and wafer cap process in which cases priority is given to the wafers in the wafer cap process. Totally, six dispatching rules are studied.

2.2 Synchronization Rules

Synchronization rules control the release of a raw wafer batch to the wafer cap process after a new batch has been released to the front-end process. Their descriptions are as follows:

SIMPLESYN (simple synchronization)**.** A wafer batch will be released into the wafer cap process at the same time as a batch is released to the front-end process.

LINESYN (line balancing synchronization). The idea for this rule is to balance the WIP (work-in-process) for the front-end process and the wafer cap process. Since the total processing time for these two sub-processes are different, the ideal WIP level for these two processes should also be different. A simple approximation would be the ratio of the WIP of the two sub-processes should be equal to the ratio of the total processing time in the two sub-processes.

Let WIP_1 *and* WIP_2 be the mean number of waiting wafers in the front-end and the wafer cap process respectively. Similarly, let T_1 and T_2 denote the total processing time for these two sub-processes. A procedure for applying the LINESYN rule is given below.

Procedure (LINESYN)

Step1. Calculate WIP inventory (WIP_1 *and* WIP_2)

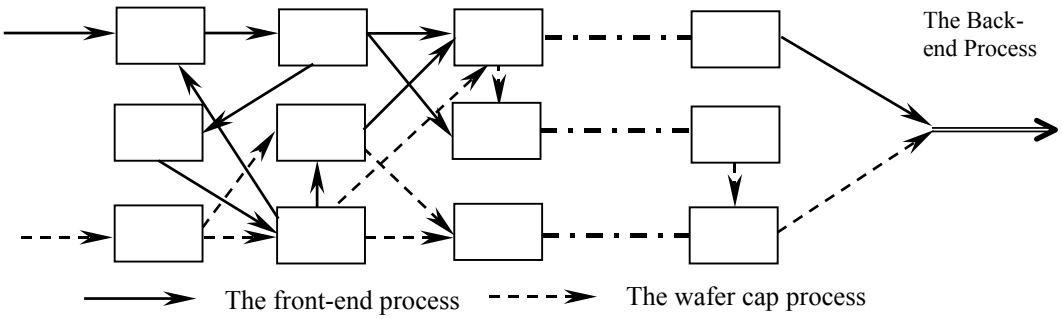

Figure 2 Schematic Representation of the Front-end Process and the Wafer Cap Process

Step2. If $WIP_1 > WIP_2 \times (T_1 / T_2)$, then release one wafer batch to the wafer cap process, else do not release.
Step 3. Go back to step 1.

QUEUESYN (queueing network models based synchronization). In this rule, we try to apply analytical queuing network models to solve this synchronization problem. As described previously, the MEMS production line is too complicated to be analyzed intuitively (see Fig.2). Since the input rule is Poisson input and the processing time of each machine is with exponential distribution (see Section 3), each machine in the MEMS manufacturing can be simplified as an M/M/1 queuing model. The expected waiting time (including processing time) at the machine can be easily calculated [9].

However, for those machines where one wafer has to visit more than once or both wafers processed in the front-end process and those in the wafer cap process have to visit, the queueing models are much more complicated (see Fig.3). Under these dispatching rules (excluding FIFO), they are actually the non-preemptive priority-discipline queueing models which assume that there are N priority classes (class 1 has the highest priority and class N has the lowest). Let W_k be the steady-state waiting time at the machine (including service time) for a wafer of priority class k. According to [9], then

$$W_k = \frac{1}{AB_{k-1}\ B_k} + \frac{1}{\mu}, \text{ for } k = 1, 2, \ldots, N, \qquad (1)$$

Where $A = \dfrac{\mu^2}{\lambda}$,

$B_0 = 1$,

$B_k = 1 - \dfrac{\sum_{i=1}^{k} \lambda_i}{\mu}$,

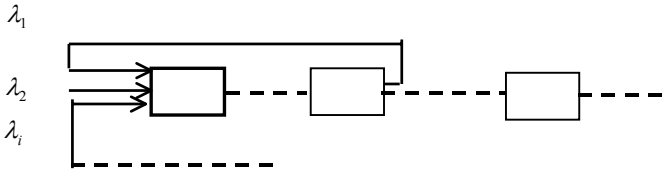

Figure 3 The Machine with Multiple Entrances

μ = mean service rate of the machine,
λ_i = mean arrival rate for priority class i, for $i = 1, 2, \ldots, N$,

$$\lambda = \sum_{i=1}^{N} \lambda_i .$$

Based on Equation 1, we can calculate the expected waiting time of each wafer at each machine under different dispatching rules. After extending the front-end process and the wafer cap process into two virtual flow shops, each virtual flow shop is actually a queueing network with a series of infinite queues (see Fig.4). The expected total waiting time of each wafer in the whole or one portion of the virtual flow shop, W_t, can be obtained merely summing the corresponding quantities obtained at the respected machines, i.e.,

$$W_t = \sum_{i=1}^{M} W_i ,$$

Where W_i is the expected waiting time of each wafer at the machine i, for $i = 1, 2, \ldots, M$,

Therefore, we can find the portion of the virtual flow shop of the front-end process which has the total expected waiting time equal to (or similar to) the total expected waiting time of the wafer cap process (see Fig.4). A procedure for applying the QUEUESYN rule is given below.

Procedure (QUEUESYN)
Step 1. Identify the last portion of the front-end process with similar or same total expected waiting time of the whole wafer cap process,

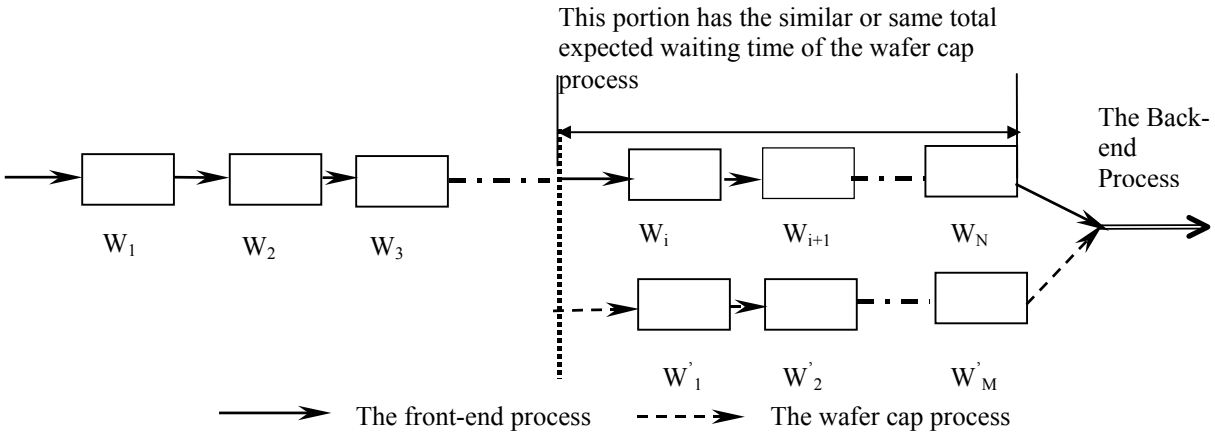

Figure 4 Schematic View of the QUEUESYN Rule and Two Virtual Flow Shops

Step 2. If one wafer batch enters the portion in the front-end process, release one wafer batch into the wafer cap process, else do not release.

Step 3. Go back to step 2.

3 Simulation Experiments

To compare all the rules introduced in this study, a series of simulation experiments is performed. Two performance measures are used for the comparison: FRONTCT, the time between the release of the raw wafer to the front-end process until it comes out from the bonding machine, CAPCT, the time from the release of the raw wafer to the wafer cap process until it departures from the bonding machine. Since the back-end process does not have the synchronization problem and it is the smallest part of the whole MEMS process, it is not considered when calculating the performance values.

The MEMS production line considered in this study consists of 37 single-server or multiserver workstations. Processing time for a wafer at one station is generated with an exponential distribution. We consider only a single MEMS product to be produced in the production line. In this study, Poisson input with three synchronization rules and six dispatching rules which resulted in 18 (1×3×6) combinations are investigated. The release rate under Poisson input is 0.0775 batch/hour, or 1.3950 wafers/hour. With this release rate, the percent utilization is around 92% for the single bottleneck station.

In the simulation experiments, each rule combination underwent 20 replications (runs) and each simulation run was carried out for a simulation time of 25,000 hours. Different random seeds were used for the 20 runs, and each run was started with an empty line. To obtain system performance in a steady state, statistics of the initial transient period (1,5000 hours) of each run were excluded from analysis. The simulation models were built using *EXTEND (version 4.01),* a simulation software developed by Imagine That Inc. and the simulation tests were carried out on a personal computer with a Pentium II (400 MHz) processor.

4 Simulation Results and Discussions

Results of the tests are summarized in Table 1 and Fig.5 and Fig.6. In Table 1, both the average values and the confident intervals with a significance level of 0.05 (α=0.05) are listed.

Fig.5 and Fig.6 indicate that synchronization rules have much more significant impact on CAPCT than they do on FRONTCT. The reason is simple because the synchronization rules only control the raw wafer release in the wafer cap process. Among the three synchronization rules, QUEUESYN performs the best and gives the shortest CAPCT. The reason is that QUEUESYN is derived from analytical queuing network models and is able to reduce the time difference of the wafers from different sub-processes arriving at the bonding machine more accurately. LINESYN performs worse because it is a heuristic rule based on the very rough estimation of the WIP ratio between the front-end process and the wafer cap process. Simplesyn rule performs the worst since it does not intend to coordinate the releases.

It should be noted here that QUEUESYN is supposed to achieve ideal solution as this rule try to make wafers output from the front-end process arrive at the bonding machine at the same time as those output from the wafer cap process do. Table 1

Table 1 Simulation Results on FRONTCT and CAPCT (hours)

Cycle time	Synchronization rules	Dispatching rules					
		CAPFIFO	CAPSRPT	FIFO	FRONTFIFO	FRONTSRPT	SRPT
FRONTCT	SIMPLESYN	209.80 (±17.5)	212.75 (±13.8)	194.44 (±11.6)	171.88 (±10.5)	160.86 (±10.2)	216.99 (±15.5)
	LINESYN	213.32 (±19.4)	194.92 (±9.58)	189.31 (±7.20)	203.75 (±10.2)	181.66 (±9.92)	206.79 (±15.3)
	QUEUESYN	210.43 (±16.8)	209.12 (±18.8)	203.66 (±15.8)	191.18 (±11.7)	175.43 (±13.8)	200.45 (±9.33)
	No Bonding	209.80 (±17.5)	212.75 (±13.8)	194.44 (±11.6)	168.83 (±9.72)	160.45 (±10.1)	216.99 (±15.5)
CAPCT	SIMPLESYN	209.80 (±17.5)	212.75 (±13.8)	194.44 (±11.6)	171.88 (±10.5)	160.86 (±10.2)	216.99 (±15.5)
	LINESYN	68.707 (±4.40)	64.771 (±1.98)	61.266 (±1.59)	57.933 (±2.32)	53.777 (±2.10)	67.132 (±3.23)
	QUEUESYN	29.994 (±2.56)	24.007 (±0.355)	39.685 (±2.78)	89.755 (±7.43)	69.517 (±4.77)	25.661 (±0.251)
	No Bonding	21.174 (±0.212)	20.978 (±0.246)	36.230 (±1.65)	60.987 (±7.94)	58.392 (±5.16)	22.425 (±0.16)

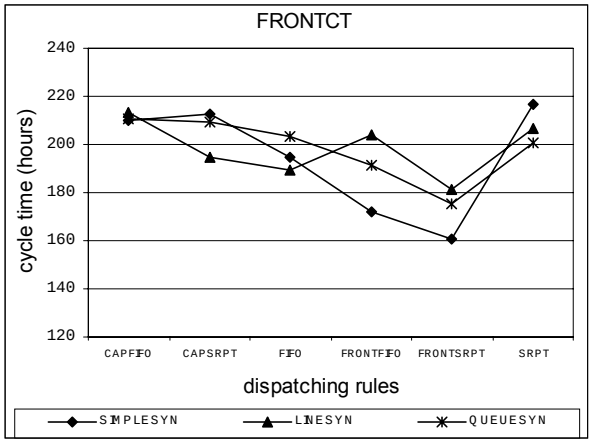

Figure 5 Simulation Results on FRONTCT

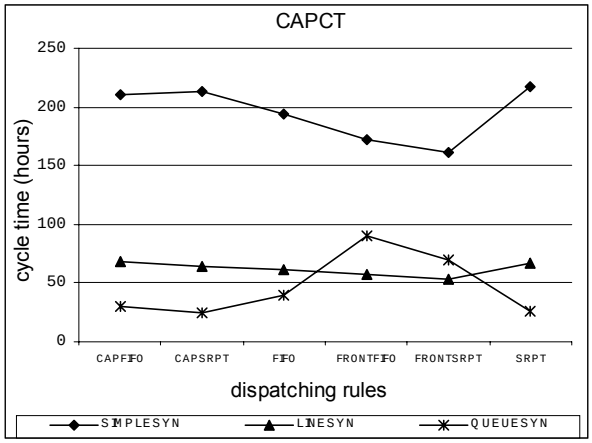

Figure 6 Simulation Results on CAPCT

also lists FRONTCT and CAPCT when each wafer either output from the front-end process or from the wafer cap process can immediately go through the bonding machine without any delay. In these cases, the values of FRONTCT and CAPCT can be regarded as the ideal solutions. It can be found from Table 1 that results under QUEUESYN rule are very near the ideal solutions, while there are still some difference between them. The reason is that QUEUESYN is only able to reduce the difference of the expected arrival time of the wafers from different sub-processes at the bonding machine. However, due to the stochastic nature of the manufacturing system, there are inevitably some variations of the expected arrival times for the wafers.

It can also be seen from Fig.5 and Fig.6 that dispatching rules also have significant impact on the performance of MEMS manufacturing. FRONTSRPT and FRONTFIFO yield the best performance of the six dispatching rules. In these cases, wafers in the wafer cap process arrive at the bonding station earlier than those in the front-end process and FRONTSRPT and FRONTFIFO which give priority to the wafers in the front-end process will make them move ahead faster, and hence help reducing the cycle time. However, under QUEUSYN rule, the performance of FRONTSRPT and FRONTFIFO is not good, because in these cases, the release to the wafer cap process is delayed too much so that the wafers in the front-end process might arrive at the bonding station earlier than those in the wafer cap process and FRONTSRPT and FRONTFIFO which give priority to the wafers in the front-end process will delay the completion of the wafer from the wafer cap process and hence incur the longer waiting time.

5 Conclusions

MEMS manufacturing is perhaps the most complicated manufacturing process. Discrete event simulation which is used in this problem, enables one to evaluate the process at a fraction of the cost and time actually needed for physical production. We compare 18 combinations of three synchronization rules and six dispatching rules in the MEMS manufacturing system. The results show that synchronization rules and dispatching rules have significant impact on the performance of MEMS manufacturing. From the rules studied, QUEUESYN and FRONTSRPT perform the best.

In this paper, we apply analytical queuing network models to solve the synchronization problem in the MEMS manufacturing system. The results show that this method is much better than simple heuristic rule and can achieve almost ideal solutions. However, it must be noted that only under M/M/s/FIFO (or priority discipline) queuing models, we can get the product form solutions of the expected waiting times. If under other scheduling rules, e.g., under CONWIP (constant work-in-process) or WR (workload regulation) input, no single product form solution is available. How to derive better and more accurate synchronization rules in these cases will be exactly a challenging research work.

References:

[1] M. J. Madou, *Fundamentals of Microfabrication,* Boca Raton and New York, CRC Press LLC, 1997.

[2] P.K. Johri, Practical Issues in Scheduling and Dispatching in Semiconductor Wafer Fabrication, *Journal of Manufacturing Systems,* Vol.12, No.6, 1993, pp. 474-485.

[3] I. Duenyas, J. W. Fowler, and L. W. Schruben, Planning and Scheduling in Japanese Semiconductor Manufacturing, *Journal of Manufacturing Systems*, Vol.13, No.5, 1994, pp.323-332.

[4] L. Wein, Scheduling Semiconductor Wafer Fabrication, *IEEE Transactions on Semiconductor Manufacturing*, Vo.1, No.3, 1988, pp.115-130.

[5] Y. D. Kim, D. H. Lee, J. U. Kim, and H. K. Roh, A Simulation Study on Lot Release Control, Mask Scheduling, and Batch Scheduling in Semiconductor Wafer Fabrication Facilities, *Journal of Manufacturing Systems*, Vol.17, No.2, 1998, pp. 107-117.

[6] R. Uzsoy, 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, 1992, pp. 47-60.

[7] R. Uzsoy, 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, 1994, pp. 44-55.

[8] J. W. Fowler, J. K. Cochran, and S. Horng, A Group-technology Coded Literature Review of Semiconductor Manufacturing Operations Publications: the MASMLAB Bibliography Web Site, *IEEE Transactions on Semiconductor Manufacturing*, Vol.12, No.2, 1999, pp. 259-263.

[9] F. S. Hillier and G. J. Lieberman, *Introduction to Operations Research*. Singapore, McGraw-Hill, 1995, Chap. 15, pp.679-713.