Short-Term Bottleneck Detection for Process Planning in a FAB

Karam Yang, Yongho Chung, and Sang C. Park

***Abstract*—This paper presents a detection process for bottlenecks in a wafer FAB (fabrication) using a simulation approach. In a semiconductor manufacturing industry, a Wafer FAB facility includes various equipment and wafer products. The wafer FAB has many characteristics, such as re-entrant processing flow and batch tools. The performance of a complex manufacturing system (i. e. semiconductor wafer FAB) is mainly decided by a bottleneck. In this paper, we define the problem of a bottleneck process and propose a simulation based framework for short-term (3 shift) bottleneck detection. By the proposed framework of this paper, the performance of a wafer FAB is improved in on-time delivery and the mean of minimum of cycle time.**

***Index Terms*—Bottleneck detection, semiconductor, simulation, wafer fab.**

I. Introduction

A semiconductor chip is the highly miniaturized and integrated electronic circuits that consist of thousands of components. All semiconductor manufacturing processes start with raw wafers (thin discs made of silicon). A lot is a moving entity in a semiconductor manufacturing system. Each job contains a fixed number of wafers. Several thousand identical chips can be manufactured using a single wafer by building up the multi-layer of electronic circuits in a wafer fabrication facility [1].

A wafer FAB (wafer fabrication) for manufacturing semiconductor products has complex and various characteristics, such as hundreds of production equipment, dozens kinds of wafer products, batch tools, re-entrant processing flow, and sequence dependent setups. Each kind of products has a special technologic process flow which involves many processing steps. A wafer fabrication process takes 70~80 percent of total wafer cycle time; the identification of workloads, which consider the total capacity and effective management of the physical distribution of semiconductor wafers, is important for total work and WIP amount [2].

In the result of existing research, this research area regarding a wafer FAB mCanufacturing industry requires unique and challenging features in comparison to others is determined. Bottleneck detection is an important performance indicator in a wafer fab. Due to the characteristics mentioned above, bottleneck detection frequently occurs. Because the breakdown of overloaded work-centers could cause an extreme bottleneck step in a fab, this bottleneck detection has great impact on target achievement, on-time delivery and cycle time.

Over the past few years, a large number of WIP control approaches has been developed and industrial researchers (i. e. Minimum Inventory Variability Scheduler (MIVS)) has proposed global Line Balance (LB) algorithms [3].

Both MIVS and LB consider a WIP balance from the viewpoint of operations, while some researchers addressed a WIP balance from the viewpoint of work centers. Although WIP control approaches look at a WIP balance at different viewpoints, they all have a same objective to make sure that a fab is run in a smooth way and each lot goes through the fab at the right pace [3]. Bottleneck detection studies have been used to estimate WIP levels. Most of these studies have focused on the determination of an appropriate WIP level.

The bottleneck approach, begun from Theory of Constraints (TOC), was applied through the use of workload regulation [4]. These methods aim to avoid bottleneck starvation by lot release control and clarify that a bottleneck output is not steeply reduce by the uncertainty of a wafer fab. Tsai *et al*. (2003) proposed the four categories of dispatching rules for the bottleneck and non-bottleneck machines. This method uses the pre-defined upper control limit and lower control limit of WIP to determine a bottleneck that is in a starved state and a non-bottleneck that is in a crowded state.

In this paper, bottleneck is not the viewpoint of a machine, but the viewpoint of a step with the highest WIP in its upstream buffer and severe fluctuation. We, firstly, focus on the classification of bottleneck steps and then verify the steps are not in a starvation state in last, regardless of dispatching rules. Bottlenecks can be classified based on the duration of detection. There are short-term bottlenecks and long-term bottleneck. Based on the short-term bottleneck detection method, the objective is to improve a target move achievement ratio. However, the objective of long-term bottleneck detection methods is proposed to improve the on-time delivery.

This paper is structured as following. The process of fabrication is described in Section II, and the simulation model and simulation engine are described in Section III. The bottleneck detection framework in short-term is provided in Section IV. The conclusion of this paper is given in the final Section.

II. The Process of Fabrication

The processes in semiconductor fabrication lines can be summarized as 8 main processes. The typical work areas in a wafer fab are shown Fig. 1. Mask production,

Manuscript received January 20, 2014; revised April 28, 2014.

The authors are with the Department of Industrial Engineering, Ajou University, San 5, Woncheon-Dong, Yeongtong-Gu, Suwon 443-749, Korea (e-mail: {vision0429, scpark} @ajou.ac.kr, yongho1230@gmail.com).

DOI: 10.7763/IJCCE.2014.V3.365

Photolithography, Etch process, Diffusion, Thin film and CVD (Chemical Vapor Deposition), Implantation, CMP (Chemical Mechanical Polishing) and Clean processes are those eight.

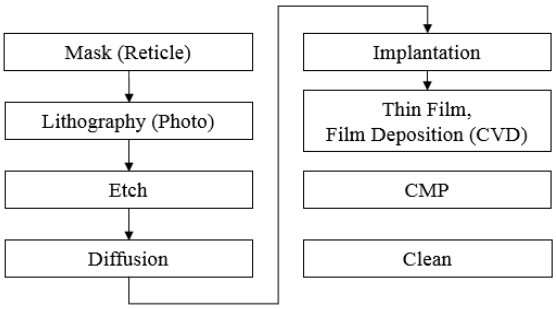

Fig. 1. The basic process of the FAB.

A mask is a series of electronic data that define geometry for the photolithography steps of semiconductor fabrication. Each of the physical masks generated from data is called a reticle. A mask set for a modem process typically contains as many as twenty or more mask, each of which defines a specific photolithographic step in the semiconductor fabrication process. The process except for photolithography regarded as non-bottleneck processes.

Photolithography is a process used in microfabrication to pattern parts of a thin film or the bulk of a substrate. It uses light to transfer a geometric pattern from a photo mask. Coating, exposure, developing, and process control are the main steps of the photolithography process. In the first step, the wafer is coated with a thin film of a photosensitive polymer, called photoresist strip. Patterns are produced on the wafers surface when an ID pattern is transferred from a photo mask (reticle) onto the photosensitive polymer, which replicates the pattern in the underlying layer. Every wafer passes through the photolithography area up to 60 times because of the different layers. The photolithography work area is a typical example of a bottleneck process because steppers are very expensive.

Etching is used microfabrication to chemically remove layers from the surface of a wafer during manufacturing. Etching is a critically important process module, and every wafer undergoes many etching steps before it is complete. We differentiate between wet and dry etching. In the first case, liquids are used, whereas gases are necessary for the latter case.

Diffusion process forces the materials to diffuse by offering high temperature environment. Diffusion is a current in a semiconductor caused by the diffusion of charge carriers. The diffusion current can be in the same or opposite direction of a drift current that is formed due to the electric field in the semiconductor.

A thin film is a layer of material ranging from factions of a nanometer to several microfabrication in thickness. Electronic semiconductor devices and optical coatings are the main applications benefiting from thin film construction.

Chemical vapor deposition (CVD) is a chemical process used to produce high performance solid materials. The process is often used in the semiconductor industry to produce thin films. Ion implantation is a materials engineering process by which ions of a material are accelerated in an electrical field and impacted into a solid. This process is used to change the physical, chemical, or electrical properties of the solid.

Chemical Mechanical Polishing (CMP) is a process if smoothing surfaces with the combination of chemical and mechanical forces. It can be thought of as a hybrid of chemical etching and free abrasive polishing. Cleaning process is used microfabrication to remove the impurity on the layers from the surface of a wafer.

III. Simulation Model

A. Simulation Model

The wafer FAB dataset MIMAC 6 from Measurement and Improvement of Manufacturing Capacities (MIMAC) is used to framework out ideas. MIMAC 6 is chosen to be tested with the proposed method, because MIMAC 6 is a typical complex wafer FAB model including multiple products, various production facilities, and process flows with hundreds of steps, which is rather close to a real 200 mm wafer FAB [4]. Table I shows the characteristic of MIMAC 6. The FAB consists for produces 9 products having process steps with each other. Step is process step. Processing Time means Lot or Batch processing time. Each machine group is composed of multiple tools. A lot is made up 24 wafers. The dataset includes 104 tool groups, 228 tools, and maximum 355 process steps.

TABLE I: Basic Characteristic of MIMAC 6

Product	Steps	Tool Groups	Tools	Processing Time
1	331	67	1190	1510058 sec
2	355	68	1164	1429932 sec
3	234	56	850	931595 sec
4	322	62	1167	1294167 sec
5	247	61	899	1018322 sec
6	266	62	967	1161121 sec
7	287	60	973	1281664 sec
8	252	60	859	1136358 sec
9	247	67	886	1104458 sec

B. Simulation Software

SEEPLAN® generates two plans (In plan: production plan, Out plan: release plan) from target production quantity and current WIP. The backward pegging engine determines the FAB in profile from a given FAB out target taking into account the capacity of the FAB. The forward loading simulation engine generates loading schedule of each equipment in the FAB and production plan by a discrete event simulation with dispatching rule by Park *et al.* (2008) [5]. Fig. 2 shows how the SEEPLAN® engine generates loading schedule for each tool in the FAB [6]. The loading simulation engine generates loading schedule and lot history on the basis of master data. The master data are bill of process (BOP) model, resource model, and dispatching rule. It specifies product type, quantity, and release sequence for each equipment of the first step resource group. The simulation result can be analyzed to see key performance indices (KPI)

including resource utilization, productivity, and WIP fluctuations. Typical examples of dispatching rules are first-in-first-out (FIFO), earliest-due-date (EDD), critical-ratio (CR), and operation-due-date (ODD).

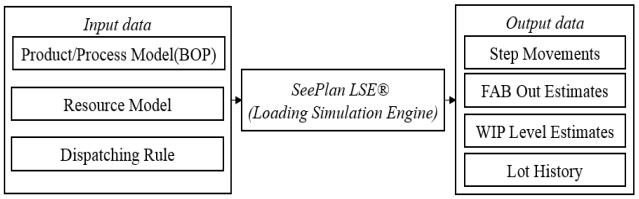

Fig. 2. FAB simulation engine.

IV. Short-Term Bottleneck Detection Framework

This paper describes a method to detect and analyze the bottleneck in short term. Bottleneck can be classified based on the duration of bottleneck step in process. Bottleneck can be classified as short term bottlenecks and long term bottlenecks. Short term bottleneck steps are steps which impede the system performance for a day. Long term bottleneck steps are steps which imped the system performance for long duration of time. Long term bottleneck is a step which has high WIP fluctuation and tardiness in the system throughput the analysis duration. Short term bottleneck steps are steps which impede the achievement ratio of the target rate. This framework identifies the bottleneck step by determining the target achievement ratio. This framework focuses on target achievement ratio.

Fig. 3 shows a flow chart of the bottleneck detection procedure, consisting of three factors (inflow, target move, move) Fig. 4. Inflow of step k calculates the amounts of wafers from step k-1. So inflow amounts are move rate of step k-1. Move can be determined by simulation from the number of production wafers in step k. Move is the total number of operation in step k. Target move using by pegging. Pegging is a process of labeling WIP lots for a target order which is specified by the due date, quantity, and product specifications including customer information [6]. It decides target for each step, gives current WIP position for each demand, and calculates the latest possible start time (LPST) for each lot. As a result, we can get the release plan to meet the out target considering current WIP and machine status.

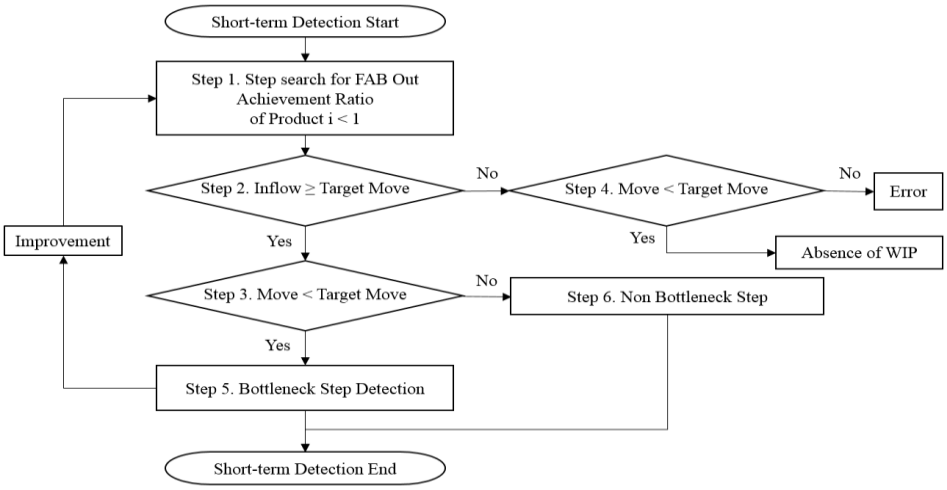

Fig. 3. Short-Term bottleneck detection.

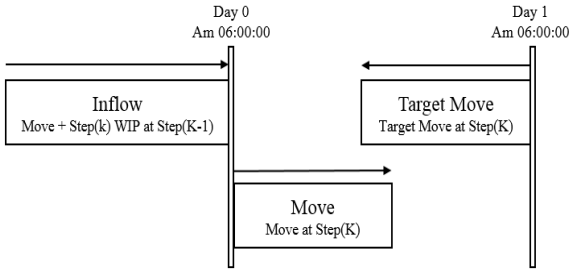

Fig. 4. The diagram of factors for short-term (Step K).

To detect the momentary bottleneck, the status of each step is identified. The status of each step can be classified into active and inactive states. Active state is the state in which the step is processing a part and inactive state includes waiting time and failure time of the step. The step with longest active duration at any instant is considered as the momentary bottleneck at that instant. The step with lowest target achievement ratio becomes the bottleneck step.

The flow chart for the short term bottleneck detection is shown in Fig. 3. First, this framework finds the process causes of tardiness product. This method of identification assumes that any step that is active will be classified as a bottleneck step, even if that step is nor lowering the system performance or impeding other steps in the system. This method of identification is in contrast to bottleneck definition that a bottleneck step is one that slows down the processing of the complete production system or impedes the system performance. The short term bottleneck detection framework is adopted to detect the bottleneck in this research. We measure the target achievement ration of different steps. The target achievement ratio is (1).

$$\text{Target Achievement ratio} = \text{Move}/\text{Target Move} \quad (1)$$

This framework the analysis of bottleneck characteristics.

Detailed Algorithm:

Step 1. Apply the bottleneck detection method to find the process of each production line, the step with the lowest target production ratio is considered as tardiness.

Step 2. Compare the Inflow of tardiness process step with

the Target Move, if the Inflow is lower than the Target Move, then go to step 4, otherwise go to step 3.

Step 3. Compare the Move with Target Move, if the Move is lower than the Target Move, then go to step 5, otherwise go to step 6. The step is potential for that to improvement.

Step 4. There is a lack of WIP. Go back to the earlier steps and to find bottleneck step again.

Step 5. It is bottleneck step. The step of bottleneck has to improvement.

Step 6. It is non bottleneck step. There is no need to improve them.

The cases classifications are shown in Table II.

TABLE II: CASES CLASSIFICATIONS

Inflow and Target Move	Move and Target Move	Classification
Inflow ≥ Target Move	Move ≥ Target Move	Step 6
	Move < Target Move	Step 5
Inflow < Target Move	Move ≥ Target Move	The error
	Move < Target Move	Step 4

V. RESULT

In this paper, we suggest a short term bottleneck detection framework increasing the target achievement and production rate. The short term bottleneck detection framework as presented in this paper is a very flexible tool and can be used for a wide range of production systems containing a wide range of entities as for steps. The method is easy to apply. The internal structure of the simulation is not needed, merely a history of the step activities. As the short term is a day. Knowing the likelihood of each step to be the bottleneck aids the manager in making a tradeoff between the benefits of improved throughput. This research work can be divided into bottleneck steps and non-bottleneck steps in the semiconductor fabrication line. If the processing time of different steps in nearly equal to each other, the position of bottleneck will vary with the bottleneck detection methods. Furthermore, the bottleneck will change during transient process. So, bottleneck detection is not an easy task. However, it is very essential to improve the performance of semiconductor manufacturing line.

Research is in progress to adapt the short term method to detect the bottleneck, allowing the monitoring of the bottleneck as is a day between different steps over time, and to improve the throughput of the short term bottleneck steps. This gives the manager of the production system in order to improve the overall system performance.

The study for future research to pursue is the research for other special cases. Expansion the field of bottleneck detection to whole semiconductor manufacturing is also challenging task.

ACKNOWLEDGMENT

This work was partially supported by *the AURI (C0003579)* funded Korea.

REFERENCES

[1] L. Moench, J. W. Fowler, S. Dauz`ere-P ́er`es, S. J. Mason, and O. Rose, "A survey of problems, solution techniques, and future challenges in scheduling semiconductor manufacturing operations," *Journal of Scheduling*, pp. 1–17, 2011.

[2] T. Y. Yang, "A study on the production control policies considering WIP balance and setup time in a semiconductor fabrication line," KAIST, 2005.

[3] Z. G. Zhou and O. Ros, "WIP control and calibration in a wafer FAB," in *Proc. 2012 Winter Simulation Conference*, 2012, pp. 2007-2018.

[4] Z. G. Zhou and O. Rose, "A bottleneck detection and dynamic dispatching strategy for semiconductor wafer fabrication facilities," in *Proc. Winter Simulation Conference*, 2009, pp. 1646-1656.

[5] B. K. Choi and J. C. Seo, "Capacity-Filtering algorithm for finite capacity planning of a flexible flow line," *International Journal of Production Research*, vol. 47, pp. 3363-3386, 2008.

[6] K. Ko, B. C. Park, S. K. Yoo, E. S. Park, and B. H. Kim," Simulation based FAB scheduling: SEEPLAN®," in *Proc. 2010 Winter Simulation Conference*, 2010.

Karam Yang is studying in the Ajou University, Korea. She is a MS candidate student in Modeling & Simulation Lab, the Department of IE. Her research interests are in defense modeling & simulation, digital manufacturing system, discrete event system simulation.

Yongho Chung was born in Korea, and is with the Ajou University of Suwon, where he studies industrial information system engineering. He is a Ph.D. candidate student in Modeling & Simulation Lab, the Department of IE. His research interests include digital manufacturing system, CAD&CAM, discrete event system simulation, semiconductor process simulation.

Sang C. Park is a professor in the Department of Industrial Engineering at Ajou University. Before joining Ajou, he worked for DaimlerChrysler Corp. and CubickTek Co., developing commercial and in-house CAD/CAM/CAPP/Simulation software systems. He received his BS, MS and Ph.D. degrees from KAIST in 1994, 1996 and 2000, respectively, all in industrial engineering. His research interests include geometric algorithms in CAD/CAM, process planning, engineering knowledge management, and discrete event system simulation.