Computers & Operations Research 38 (2011) 666–673

Contents lists available at ScienceDirect

Computers & Operations Research

journal homepage: www.elsevier.com/locate/caor

Scheduling algorithms for a semiconductor probing facility

June-Young Bang [a], Yeong-Dae Kim [b,*]

[a] System Engineering Team, Device Solution Business, Samsung Electronics Co., Ltd., Hwasung, Gyeonggi-Do 445-701, Republic of Korea
[b] Department of Industrial Engineering, Korea Advanced Institute of Science and Technology, Yuseong-gu, Daejeon 305-701, Republic of Korea

ARTICLE INFO

Available online 26 August 2010

Keywords:
Semiconductor probing facility
Scheduling
Bottleneck-focused scheduling
Hybrid flowshop

ABSTRACT

This paper focuses on a scheduling problem in a semiconductor wafer probing facility. In the probing facility, wafer lots with distinct ready times are processed on a series of workstations, each with identical parallel machines. We develop a heuristic algorithm for the problem with the objective of minimizing total tardiness of orders. The algorithm employs a bottleneck-focused scheduling method, in which a schedule at the bottleneck workstation is constructed first and then schedules for other workstations are constructed based on the schedule at the bottleneck. For scheduling wafer lots at the bottleneck workstation, we consider prospective tardiness of the lots as well as sequence-dependent setup times required between different types of wafer lots. We also present a rolling horizon method for implementation of the scheduling method on a dynamic situation. The suggested methods are evaluated through a series of computational experiments and results show that the methods work better than existing heuristic methods.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

In this paper, we consider a production scheduling problem in a semiconductor wafer probing facility in which wafers with multiple product types are tested. In general, manufacturing processes for semiconductor chips are composed of the following four major processing stages: wafer fabrication, wafer probing, assembly, and test. Among these processes, wafer probing process is relatively simple compared to other processes. However, scheduling wafer lots in the probing facility may be very important, since orders are often issued for wafers that have been completed at the probing facility. Note that even small companies can design their own semiconductor products and produce them without wafer fabrication and probing facilities but with only assembly and test facilities. For such small companies (customers), the wafer probing process may be considered as the final process for the products (wafers). In addition, in semiconductor manufacturing companies with fabrication and probing facilities, assembly/test facilities are often located relatively remotely from fabrication and probing facilities, and these two pairs of facilities are operated relatively independently in many cases. Therefore, meeting orders of the customers (small companies as well as assembly/test facilities) in terms of quantity, quality, and delivery schedules is very important for the probing facility.

In the case considered in this study and many other cases as well, orders for wafers are issued to both fabrication and probing facilities. An order is specified by the due date, product (wafer) type, and the number of wafers to be produced. In these facilities, wafers are usually processed in lots of 25 wafers, and there are one or more lots for each order. Here, a wafer lot is the basic processing and transfer unit, that is, it denotes a set of wafers that are processed and moved together. Since the wafer fabrication process is a very complex process with very long production lead times, completion times of lots, which become the ready times of the lots at the probing facility, may be distributed in wide ranges, even in cases where lots associated with the same order are released to the fab at the same time. The difference of ready times of the first lot and the last lot associated with the same order usually ranges from a few hours to several days.

The manufacturing process in the wafer probing facility is performed in four major steps: direct current (DC) test, wafer burn-in and probing (wafer test), ink marking, and back grinding, each performed on a workstation with multiple parallel machines. Therefore, the probing facility may be considered as a *hybrid flow shop*, in which there are multiple parallel machines in each processing stage. Among workstations for the four stages, the wafer test workstation is the bottleneck workstation in the probing facility considered in this study. Note that the processing time on the wafer test workstation is longer than that on the other workstations and testers in the test workstation are the most expensive and critical resources in the wafer probing facility.

At the wafer test workstation, sequence-dependent setup times are required between different types of wafers, since wafers should be tested under a configuration which is specified

* Corresponding author. Tel.: +82 42 350 3120; fax: +82 42 350 3110.
E-mail address: ydkim@kaist.ac.kr (Y.-D. Kim).

0305-0548/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cor.2010.08.010

for each wafer type. If the configuration for the next wafer is different from that of the current wafer, a setup operation for the test should be performed. There are three tasks to be done for a setup operation on the test machines: downloading software for the test, changing tools (load boards and probe cards) for the test, and changing temperature inside the machines. The times required for downloading the test software and changing the tools may be considered constant, but the times required for changing temperature are not constant but it can vary depending on the wafer types processed before and after the setup. The changeover time between different wafer types can range from several minutes to several hours.

Scheduling problems for wafer probing facilities, or hybrid flow shops, have been studied by many researchers including Ovacik and Uzsoy [1], Chen et al. [2], Chen and Hsia [3], Lee [4], and Ahmadizar et al. [5] Also, Liu and Chang [6] propose a Lagrangian relaxation method for scheduling problems with the objective of minimizing inventory cost and backorder cost in hybrid flow shops in which the effect of sequence-dependent setup times is very significant. Scheduling problems at the wafer test workstation in the probing facility are considered in a few studies as well [7–11]. Pearn et al. [7,8] and Yang et al. [9] consider parallel-machine scheduling problems associated with the wafer test workstation and transform the scheduling problems into vehicle routing problems with time windows and present algorithms for the objective of minimizing total setup times. In addition, Pearn et al. [10] modify heuristics for vehicle-routing problems to solve parallel-machine scheduling problems with the objective of minimizing total setup times, and Ellis et al. [11] present heuristic algorithms for the scheduling problems with the objective of minimizing makespan.

There are a number of studies on scheduling problems with sequence-dependent setup times, as surveyed in Allahverdi et al. [12]. However, there are not many studies on the problems with due-date related performance measures, and in most of these studies metaheuristic methods are developed. Tamimi and Rajan [13] and Fowler et al. [14] propose a hybrid genetic algorithm for the objective of total weighted tardiness, and Park et al. [15] use a neural network method. Recently, Jungwattanakit et al. [16] propose scheduling algorithms based on tabu search and simulated annealing for the objective of minimizing makespan and number of tardy jobs, and Tavakkoli-Moghaddam et al. [17] develop a genetic algorithm for the objective of minimizing the number of tardy jobs and total completion time. In other studies, metaheuristics are used in multiple-stage solution methods. For example, tabu search methods are used in multi-stage heuristic algorithms for the scheduling problems with the objective of minimizing total weighted tardiness in Kim et al. [18], Bilge et al. [19], and Eom et al. [20]. Also, Behnamian et al. [21] present hybrid metaheuristic algorithms composed of simulated annealing, variable neighborhood search, ant colony optimization methods for the problems with the objective of minimizing total earliness and tardiness. On the other hand, Luh et al. [22] use a Lagrangian relaxation method to generate job sequences and then use a greedy search method to form schedules for parallel machines from the sequences, and Demirkol and Uzsoy [23] develop a decomposition method which is composed of bottleneck identification and re-optimization procedures.

One of the scheduling methods most commonly employed in practice is the one based on priority dispatching rules. Kim et al. [24,25] suggest priority-rule-based algorithms for lot release control and lot scheduling, and Kim et al. [26] develop a real-time scheduling method, for the objective of minimizing tardiness of orders in wafer fabs. For scheduling jobs with sequence-dependent setup times on parallel machines in a wafer probing facility, Lee and Pinedo [27] develop a priority rule called apparent tardiness cost with setup (ATCS) rule to find an initial schedule, and then use a simulated annealing algorithm to improve the solution. Moreover, Mason et al. [28] modify the ATCS rule for scheduling jobs on batch-processing machines considering jobs to arrive in near future as well as jobs currently available, while Pfund et al. [29] modify the ATCS rule to consider ready times of the jobs and develop a new rule named ATCSR, and show that ATCSR outperform other existing priority rules. In addition, Lee et al. [30] develop a real time dispatching method based on Petri nets for multi-criterion objectives such as those related to work-in-process inventory, tardiness, and profit.

In this paper, we present a heuristic algorithm for the lot scheduling problem with the objective of minimizing total tardiness of orders in a wafer probing facility in which multiple types of products are produced. Here, the tardiness of an order is defined as $\max\{C_o - D_o, 0\}$, where C_o and D_o are the completion time and due date of the order, respectively. The completion time of an order is the time when all wafers for the order have been completed in the probing facility. A bottleneck-focused scheduling approach is employed in the heuristic algorithm, and priority-rule-based algorithms are used for scheduling at the workstations including the bottleneck workstation. Also, we suggest a rolling horizon method to cope with the dynamic nature of the probing facility, i.e., the fact that lots arrive at the facility dynamically, since the bottleneck-focused method, which is a static scheduling method, may not be directly applied to such a situation.

2. Bottleneck-focused scheduling method

The bottleneck-focused scheduling algorithm, which is suggested by Lee et al. [31], is an iterative scheduling method for multiple-stage scheduling problems. The bottleneck-focused algorithm is similar to the shifting bottleneck heuristic suggested by Adams et al. [32] in that a bottleneck machine is considered with a higher priority when scheduling decisions are made. However, in the shifting bottleneck heuristic, which is originally developed for job shop problems, a (new) bottleneck machine is identified in each step and the schedule at the bottleneck machine is obtained and then fixed during the procedure of the heuristic, while in the bottleneck-focused algorithm, which is used for flow shop problems, the bottleneck machine is fixed throughout the procedure and schedules on the machines including the bottleneck are determined with an iterative procedure.

In this research, we extend the bottleneck-focused algorithm to consider special characteristics of the probing facility, which are distinct (non-zero) ready times of lots and sequence-dependent setup times on the wafer test workstation, i.e., the bottleneck of the probing facility. We develop a priority-rule-based algorithm called the *progress-based lot scheduling algorithm* to schedule wafer lots on parallel machines in the bottleneck workstation considering sequence-dependent setup times. Since the bottleneck workstation is not the first workstation of the probing facility, ready times of lots at the bottleneck need to be estimated and set with a certain method, and these ready times are updated iteratively in the procedure of the bottleneck-focused method.

In the following, we present the iterative procedure of the bottleneck-focused scheduling method, and a scheduling algorithm for the parallel-machine scheduling problem, which is to be used in the bottleneck workstation. The following notation is used in this paper.

- i index for the lots
- k index for the workstations, $k=1,\ldots,K$
- b index for the bottleneck workstation, i.e., the wafer test workstation

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات