Journal of Manufacturing Systems 49 (2018) 107–120

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

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

Setting daily production targets with novel approximation of target tracking operations for semiconductor manufacturing

Yu-Ting Kao[a,*], Shi-Chung Chang[a,b]

[a] Institute of Industrial Engineering, National Taiwan University, Taipei, Taiwan, ROC
[b] Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan, ROC

ARTICLE INFO

Keywords:
Daily target setting
Re-entrant line
Target-induced variability
Bernoulli trial approximation
Hybrid tandem queue

ABSTRACT

Daily production target setting based on machine capacities and available wafer-in-process (WIP) is an important practice in a semiconductor fab characterized by re-entrant process steps sharing individual machine group capacities. Operations of individual machine groups will track their respective target guidance through detailed machine allocation and lot dispatching (MALD), inducing variations of wafer flow. Existing approaches do not explicitly address such target-induced variations (TIV), and the resultant target setting may incur throughput loss and prolonged cycle times. A novel design of target-setting algorithm considering TIV called TaTIV is proposed to take TIV into account for setting daily targets systematically. TaTIV integrates three innovations: i) a Bernoulli trial model for approximating TIV of MALD at a process step under a given target, ii) a hybrid and recursive tandem queue approximation of multiple-step target-induced wafer flows and flow times given initial WIPs and machine group capacities of the day, and iii) a fixed-point iteration between target setting and target-induced wafer flow estimation. Analyses and simulations exploiting fab data show that TaTIV sets targets closer to what are actually achieved at individual steps than those set without considering TIV, reduces inter-step variations of machine allocation, and improves fab throughput and cycle times.

1. Introduction

In a semiconductor fab, each type of silicon wafers requires a serial flow of process steps, which may involve hundreds of steps as well as delicate and expensive machines. The process flow of a wafer is highly re-entrant because the flow requires multiple visits to each machine group as layers of circuitry are successively added onto the wafer. This re-entrant feature is different from both traditional flow shops and job shops, posing a unique challenge to production operation management [1]. Wafers of different products as well as similar products at different layers of fabrication compete for the finite capacity of a machine group [2]. Dynamic and uncertain factors (e.g., batching requirements, waiting time constraints, machine capability, reticle availability, machine qualification, and availability) and production priorities further complicate fab operation management [3]. Increasing fab production efficiency through scheduling and control have been common fab practices that decompose and coordinate operation management in multiple continuous improvements, which are critical to achieving high return on investment [4–10]. In particular, hierarchical production planning, scheduling, and control decompose and coordinate operation management in multiple time scales and decision levels to manage the extremely complex processes and operations in a fab for efficiency and productivity [11–13].

Fig. 1 depicts the three-level hierarchy adopted by many fabs to manage the complexity of semiconductor fabrication operations: i) master production schedule (MPS), ii) daily target setting (DTS), and iii) detailed machine allocation and wafer lot dispatching (MALD). In the top level, the MPS aims to control the WIP level and cycle time of the entire fab while satisfying delivery requirements of customer orders. Frequently, MPS considers fab in an aggregated and coarse granularity capacity model with simplified production flow dynamics. MPS schedules the quantities of wafer release and output for a fab using day or week as a time unit over a few months to one year. In the bottom level, the detailed machine allocation determines the allocation of available machine capacity in each machine group to various production steps that require the same machine group once every few hours and based on fab production states. Real-time dispatching executes priority rules and dispatches individual wafer lots to the appropriate and available machines for processing. Lot dispatching is local to individual machines and steps and responds to operation status in real time.

In the middle level, DTS provides a critical linkage between MPS

* Corresponding author.
E-mail address: f97546010@ntu.edu.tw (Y.-T. Kao).

https://doi.org/10.1016/j.jmsy.2018.09.003
Received 14 June 2018; Received in revised form 12 August 2018; Accepted 12 September 2018

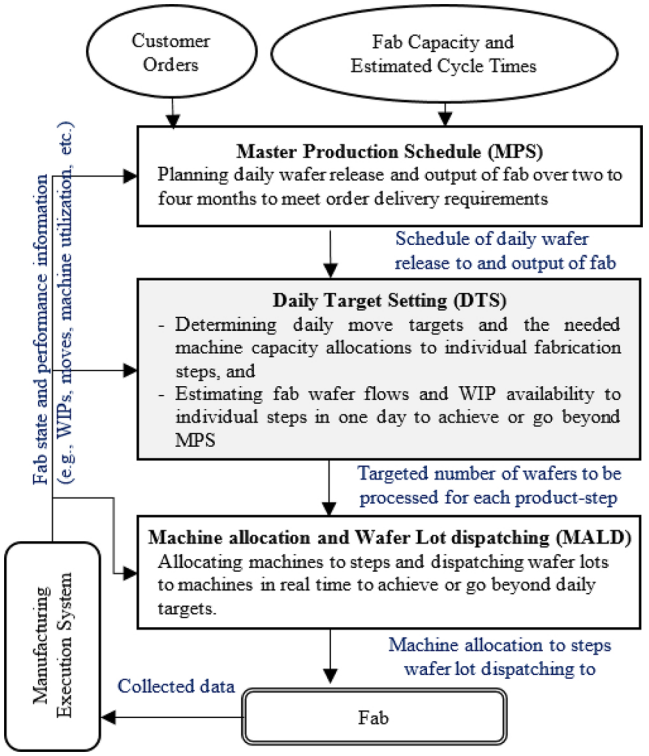

Fig. 1. Functional hierarchy for fab operation management.

and MALD by planners and shop floor supervisors respectively. DTS determines the targeted number of wafers to be processed for each product type at each process step and the estimated machine capacity allocation to the step daily. In practice, DTS is one of the key review items in the daily production meeting of a fab. Comparing the target number of wafers processed at each product step in a day and the actually achieved ones (called moves) helps a fab manager grasp and control the progress of production [14]. Effective target setting has made significant impacts on fab performance and facilitated quick response, on time delivery and high throughputs [15]. DTS is the focal problem in this paper.

There have been various studies of the DTS problem. The author of [16] provided a detailed problem description. Govind et al. [5] adopted a linear programming-based optimization engine to convert fab requirements into targets for individual machine groups. Such integration of target, near real-time scheduling and dispatching has resulted in significant improvements in output and cycle time in the lithography area. In [16], a daily target-setting system (TSS) is proposed to approximate, under given targets, WIPs flowing into individual steps in a day and then a fixed point iteration to solve the chicken-and-egg problem between wafer-flow and proportional-to-workload target setting. Field applications demonstrated that quality DTS leads to more than 20% increase in daily moves and more than 8% decrease in the wafers-in-process (WIPs) of a foundry fab case.

Despite the successful results, all the aforementioned target-setting schemes are mean value-based approaches. In specific, the estimated wafer flows and targets generated by TSS may significantly differ from those actually accomplished in one day as evidenced by a memory fab case depicted in Fig. 2. The horizontal axis shows the step index, while the vertical axis shows the achievable flow quantity percentage. The differences between the estimated flows (blue line) and the actually achieved flows (red line) are obvious, even up to 40%. The production targets are higher than the actual flows at certain steps, especially before step 121. It is because TSS assumes constant machine allocation over time, that is, various production steps sharing the same machine group have the same wafer flow times. However, the machines are allocated to different steps dynamically within time by shop floor supervisors to track the targets, which generates the variations of wafer flows to each step and unreachable targets.

The role of variations in a fab has gained increasing attention from researchers in recent years for better control and fab management [17]. Wu identified the importance of manufacturing variability in [18] due to the combined effects of machine allocation, lot dispatching, rework, scraps, setup, preventative maintenance, and machine breakdown. High variability leads to accumulated WIP while low predictability results in on-time delivery [19]. The analysis of Kingman [20] indicates that the WIP of a single server queue is proportional to the variation coefficients of the arrival and service processes. Increase in service and inter-arrival variability to individual steps lead to an increase in waiting times, thereby increasing cycle times among the process steps.

Many studies have demonstrated that the incorporation of variations into production flow helps to reduce the operation variations and to improve fab performance. The authors of [19] proposed a last-station-pace strategy to reduce inter-departure variability and performed simulation validation. The study of [21] reported that a fab accomplished 50% cycle time decrease and 35% throughput increase by lowering its operation variability through proper scheduling. Kacar et al. [22] showed that in a fab with high variability, production planning with cycle time models that consider processing variability, such as engineering holds and yield excursions, performs better than models without. Hassoun and Rabinowitz also showed that high utilization variability of machine capacity is critical to predicting WIP bubbles with accuracy [23]. Such studies have indicated the potential that including variations into DTS may help provide reasonable approximations of wafer flows and then improves fab performance.

In a fab, one major objective of MALD is to track daily production targets set by DTS; however, poor targets misguide MALD and induce unnecessary variations [24]. In this paper, we investigate how daily targets should be set to achieve desirable fab performance by considering the chicken-and-egg relation that target setting induces

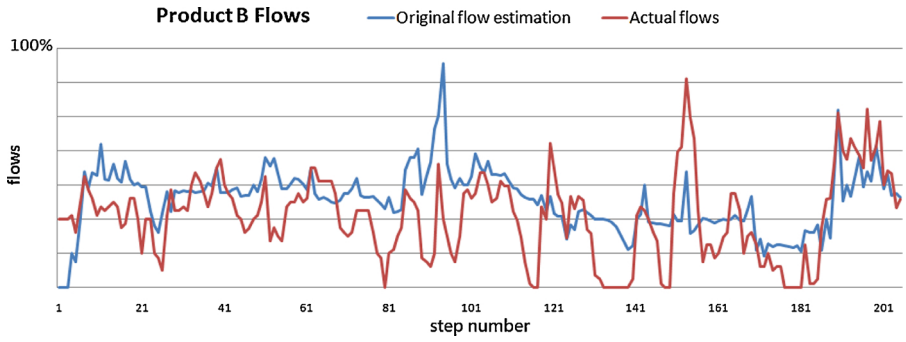

Fig. 2. Comparison of estimated flows from TSS and actual flows.