Dynamic Dispatching Rule Considering Finite Workforce for Semiconductor Wafer Fabrication Process

Po-Hsun Chen and Chuang-Chun Chiou*
Department of Industrial Engineering and Enterprise Information, TungHai University, Taiwan
**Corresponding Author: cjchiou@thu.edu.tw*

Received 25 February 2020; received in revised form 11 May 2020; accepted 3 June 2020

Abstract

Job dispatching plays an important role in semiconductor wafer fabrication processes, the goal is to decrease work-in-process (WIP), have a better yield rate, and to satisfy the due date. Most researches focus on how line balancing and WIP controlling effect the decision flow while dispatching, however, not every wafer fab process are fully automated. Factories that required operators to handle materials or works are considered semi-automated, and the finite workforce may cause materials waiting for operators to move or to monitor. As a result, a dynamic dispatching rule considering finite workforce is proposed for semi-automated wafer fab line. A case study is demonstrated by simulation to present the conclusion. The results show that finite workforce will decrease throughput amount and increase cycle time. By the dynamic dispatching proposed in this study, both throughput and cycle time have significant improvements.

Keywords: Dynamic dispatching, finite workforce, throughput, cycle time

1. Introduction

The manufacturing process of semiconductor wafers can be separated into two sections, wafer fabrication (fab) and probing as "Frontend"; assembling and testing as "Backend". Most re-entrant flows happen in the Frontend processes, where hundreds of procedures are repeated in different layers due to the variety of products. Managers seek to deal with variables under this condition, variables including: (1) Equipment specific factors, such as, machine breakdowns, batching machines, and different setup requirements; (2) Different kinds of products that lead to unfixed product mix; (3) Operational factors, such as, dispatching rule, work-in-process control, workforce allocation, and maintenance policy; also (4) Tool redundancy, tool dedication, re-entry flow, and etc (Dequeant et al., 2016). Recent researches focus on different factors to shorten cycle time and meet the due date. This study focuses on operational factors, including dispatching, work balancing and workforce allocation.

A good scheduling system can increase throughput by reducing work-in-process, preventing bottleneck starvation and process failures. Studies have been finding methods to optimize production flow through scheduling in decades, sorting from how dispatching rules affect production (Li et al., 2004) (Chen et al., 2012) (Bergmann et al., 2015); how line balancing smooths the reentrant flow (Yoon et al., 2018); and recipe arrangements to increase target production quantity (Park et al., 2013) However, considering not every wafer fab factories are fully automated, operators need to load, transport, and monitor to maintain workflow. In a semi-automated fab production line, workforce allocation effects how materials flow in fab lines, insufficient workforce will cause a bottleneck as machines has to wait for operators to perform the next process.

This study provides a dynamic dispatching rule considering finite workforce and human-related parameters. It proofs how workforce effects dispatching, scheduling, and load balancing. The first section provides an introduction of wafer fab production and the problem in semi-automated facilities; the second section reviews dispatching and scheduling issues considering different parameters with different methods; the third section shows the methodology of this study and provides an actual case in section four.

2. Literature Review

Dispatching is to decide the priority and order of the product after they reached a certain process. The purpose is to control how items flow considering due date, cycle time, even wip volume in a production line. By dispatching, ones that are urgent to meet the due date have a higher priority while the others can be queued in the storage. With an ideal dispatching rule, operators can meet the indicators under different production status.

When a machine is available, a dispatching rule will prioritize working items in the queue. Indicators such as cycle time, due date, are concerned during this period of time. Operators or machines will begin the process under the order. Common dispatching rule including First-Come First-Serve (FCFS), Earliest Due Date (EDD), Shortest Processing Time (SPT), Longest Processing Time (LPT), Critical Ratio (CR). In

Table 1. It shows some common rules for dispatching and their description

Table 1: Common Dispatching Rule and Descriptions

Dispatching Rule	Rule Description
FCFS	Jobs entering the queue first will be processed first.
EDD	Jobs with earliest due date will have the highest priority.
SPT	Jobs taking the shortest operation time will be processed first.
LPT	Jobs taking the longest operation time will be processed first.
CR	The priority is decided by its current workstation, due date and the remaining process time.

Researchers have developed different dispatching rules since 2000s, most studies focus on on-time delivery, cycle time, and throughput. Li et al. (2004) provides a dispatching rule for improving on-time delivery by considering bottleneck machines, non-bottleneck machines, and batching machines. The solution is prior than other dispatching rules on throughput, cycle time and on-time delivery. Chen et al. (2012) proposed a dynamic dispatching rule for semiconductor assembly production line. The study focuses on batch processors and provides an optimal batch forming rule. Also, Meiji and Li (2018) construct a load balancing prediction model, and provide a decision flow to compute the degree of jobs after recognizing the characteristics such as: whether the machine is a batch machine? Even considering whether the product is a hot job. These studies compare their methods with first-in, first out (FIFO), earliest due date (EDD), critical ratios, and other common dispatching rules. The solution shows that adopting a dynamic dispatching rule enables manufacturers to have more controls on machines, wip, and resources.

Besides constructing an algorithm, or decision flow, Bergmann et al. (2015) investigate the suitability of different dispatching rules by data mining method, then emulate scheduling decisions with data gathered form production data. The study demonstrated a scenario with a combination of data mining methods with effective data transformation for approximation of decision rules. Cho et al. (2017) presented a dispatching rule to achieve on-time delivery and throughput for fabs with dedication constraint. They conduct a simulation experimentation applying the operational due date to achieve the goal.

Quoting from “Feel the force of flexible manpower” (2006), the general principle of semi-automatic lines is similar to automatic lines but certain operations are difficult to be handled automatically, such as transferring, loading, and preparing. Considering infinite workforce, operators are available anytime to control workflow; however, if workforce is finite, transferring and loading can be a bottleneck. Thus, manual operations will increase cycle time once the finite workforce cannot handle that much available working items.

Cheong et al. (2007) provide a way to improve labor productivity by minimizing time wastage caused by non-value added activities, such as monitoring, transportation, inefficient operation. About 25.7% of daily tasks are non-value added operations in a 200mm fab, in other words, the human interventions mentioned above will cause too much time waste in fabs.

We found that most studies of fab dispatching rules provide algorithms and decision-making strategies under different circumstances, somehow, they consider workforce infinite and most of the examples are automatic fab production lines. This study takes an unconventional method considering both infinite and finite workforce into a dispatching algorithm which provides priority information when more than one batch of wafer stay in the queue and helps operators on dispatching-related decision making. Eventually, we test the algorithm by constructing a simulation model.

3. Modelling

Wafer fab production processes are done by layers after layers, which leads to the re-entrant situation that batching machines can simultaneously processing on two batches of wafers from different layers. By the fact, we make following assumptions and confirmations:

1. Production information is collected from the production system, or so called Manufacturing Execution System. Operation time of each workstations, wip in queue, and scheduling are obtained by data collection.
2. Considering only one layer of fab production, which consists of some important procedures such as lithography, etching, and even chemical deposition.
3. Defects and productive maintenance are not considered since they not only cause breakdown also but also re-entry.
4. Once a process starts, no other work items can interrupt the current one until it is done.
5. We decrease the kinds of products and combine the one with similar recipes.

Then we focus on which workstations and parameters will be considered in the algorithm.

a. Workstation Selection

We divide the whole layer into two basic stages, and the recipes consist in the stages are shown in table 2.

Table 2: Detail Information for Selected Stages

Stages	Detail procedures
Photolithography	From cleaning, preparing, photoresist application, exposure and developing, to hard baking.
Etching Procedure	A semiconductor process that material is removed from the surface of a wafer, including wet etching and dry etching.

b. Parameter Notation

Table 3: Parameter Notation for the Algorithms

Parameters	Parameters meaning
T_i	The time that product i is in the production line.
$TimeOut_{i(j-1)}$	The time that product i leaves workstation $(j-1)$
RT_{ij}	Residence time of product i in the queue of workstation j
$WIP_j(t)$	The WIP amount of workstation j in time period t.
PT_j	The production time of the workstation j.
TWT_j	Total available working time of workstation j.
α	The weight for due-date parameter.
β	The weight for workload parameter.
PPT_i	The predicted production time of product i.
$PT_i(t)$	Current production time of product i in time period t.
OPT_{ij}	Operator process time of workstation j for product i.

We construct two different decision model under two different considerations.

a. Considering infinite workforce

Considering the workforce infinite, the factors that influence the work flow are how and when the materials enter a machine. First, we need to determine whether the machine is a batch production machine. Most batch production machine have long process time, and the batch size is 4-5 times more than a lot. Batch production machines will interfere the downstream workstations if the jobs batched cannot feed the available machines downstream. If the workstation is not a batch production machine, we will decrease scraps by avoiding overtime queuing. Most chemical recipe requires the job to be pushed into the next stage in a short period of time. We calculate the queue time by Equation (1). Satisfying this condition can make sure that every lot will be processed unless machine break down encounters.

$$MQT_{ij}(t) = T_i - TimeOut_{i(j-1)} \leq MaximumQueueTime_j \quad (1)$$

Next we consider whether hot lot exists and raise its priority to satisfy the urgent need, by equation (2). If work item i is a hot lot, then we set the HT_i value to its residence time, to raise the selection priority value S_{ij}.

$$HT_i = \begin{cases} 0, & if\ i \notin hot\ lot \\ RT_{ij}, & if\ i \notin hot\ lot \end{cases} \quad (2)$$

Next, quoting from recent research (Li et al, 2018), we concern the workload of every machine. If the manufacturing time for WIP in machine j is nearly close to the maximum working time, then the upstream machine should decrease its feeding into machine j. The workload value is calculated by equation (3). Finally, we calculate the selection priority by due date, workload and hot lot parameter by equation (4). The decision flow is showed in Figure 1,

$$\tau_j^n(t) = \frac{WIP_j(t) \times PT_j}{TWT_j} \quad (3)$$

$$S_{ij}(t) = \alpha_1 \frac{PPT_i - PT_i(t)}{DueDate_i - ReleaseTime_i} - \beta_1\, \tau_j^n(t) + HT_i \quad (4)$$

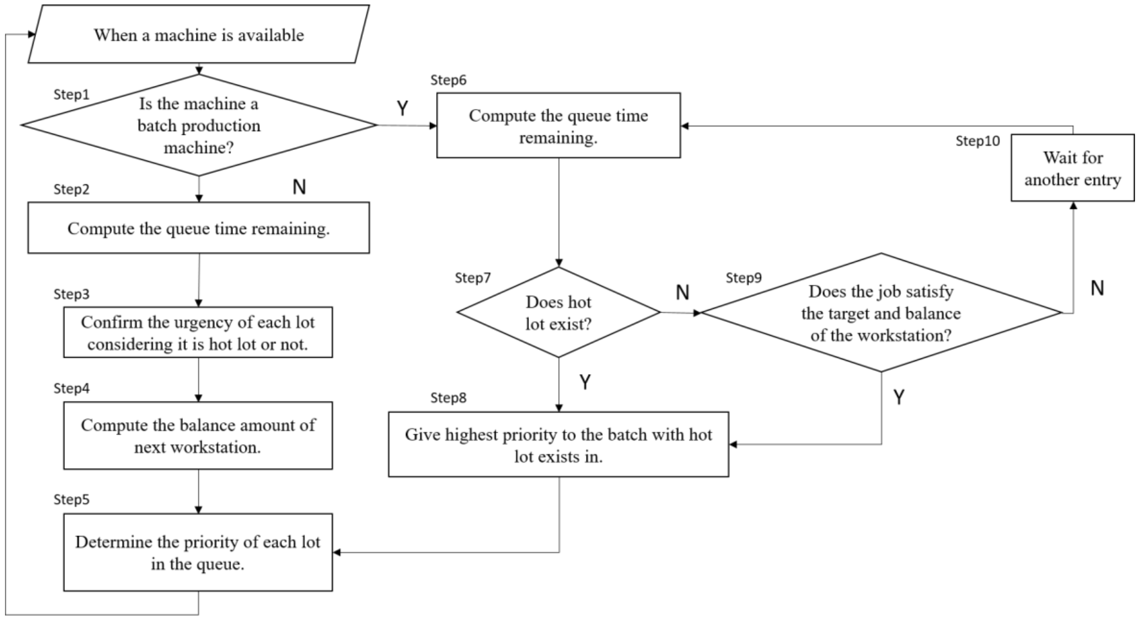

Figure 1: Decision Flow Considering Infinite Workflow

b. Considering finite workforce

How operators are involved in material flow is shown in figure 2. Basic workflow starts when materials are released, items enter a queue then start looping in different machines, finally stocked in warehouse after all jobs are finished. If we consider operations such as transporting, preparing, and feeding, which are operated by man, we must consider whether an operator is available before any materials are available for feeding. In other words, transporting and preparing are separated from other operations due to the fact that only operators and work items are involved. Since workforce are involved before any machines, how operators are allocated should be considered.

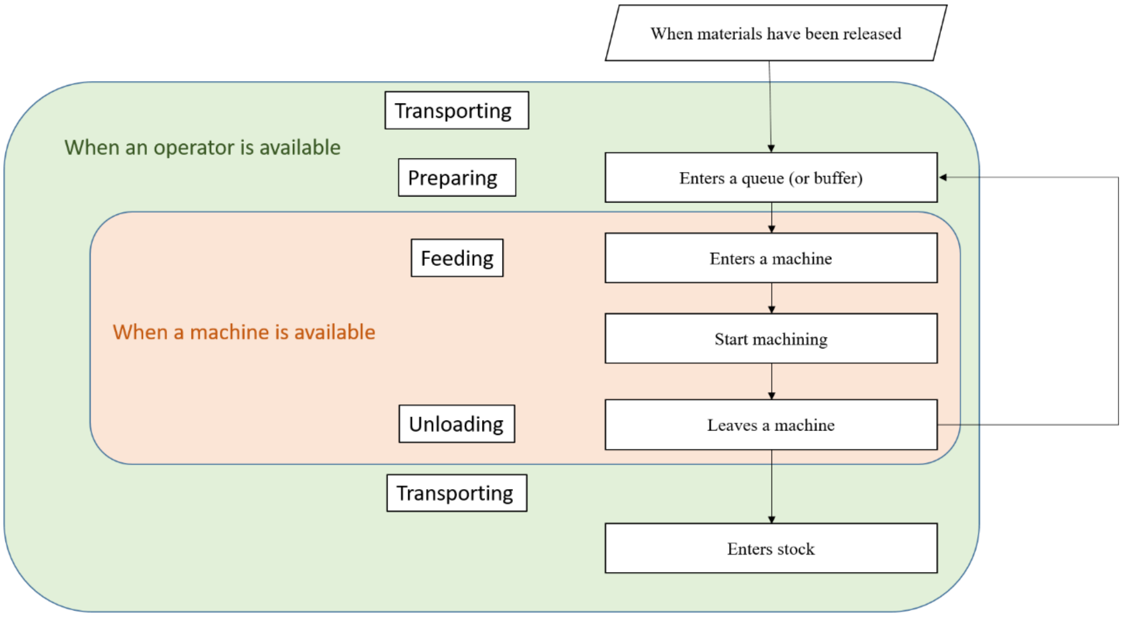

Figure 2: Workflow Considering Operator's Processes

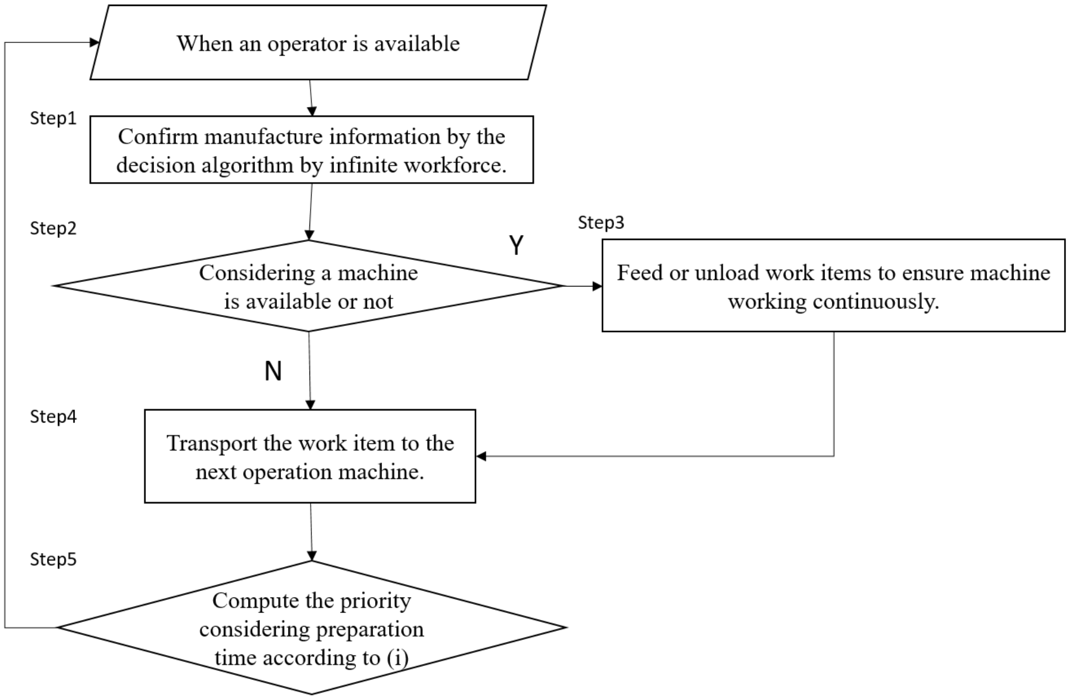

Figure 3: Decision Flow Considering Finite Workforce

After recognizing how operators are involved, we re-design the decision flow with concerns. First, when an operator is available, we provide product information to acknowledge which product has the higher priority according to the algorithm and decision flow proposed in the previous part. Then we consider whether any machine is available or not. The purpose for this step is to increase the utilization of every machine by making sure that every machine is processing. If any machine is available, operators should feed new lot of wafer into the machine for its next procedure; else the operator should focus on transportation or preparation considering these actions will take long period of time. By the conditions above, we recalculate the selection priority by (i). After finishing any process, the operator will be considered as available again.

$$HS_{ij}(t) = \alpha_2 \frac{PPT_i - PT_i(t) + OPT_{ij}}{DueDate_i - ReleaseTime_i} - \beta_2 \tau_i^n(t) + HT_i \quad \text{(i)}$$

4. Case Study

In this section, we will take actual data to show the difference between each dispatching rules. For the simulation model, we took five different categories of wafer products and its actual production schedule to simulate in the process of the proposed layer. With an average release amount of wafer 1300 pieces per day, five kinds of product combination and reasonable due-date, we simulate the production line in simulation time as five days. We took FCFS, EDD, and our algorithm to show how different these dispatching rules affect average cycle time, on-time delivery rate, and throughput rate.

We set the dispatching rule to every workstation, including EDD, FIFO and the method we proposed. The results for infinite workforce are shown in table 4 and finite in table 5.

Table 4: Infinite Workforce Simulation Results

Dispatching Rule	Throughput(Pieces)	Cycle Time	Improvement Throughput(Pieces)	Improvement Cycle Time
FCFS	6000	828.79	-	-
EDD	6000	808.38	-	2.40%
Dynamic Dispatching Rule	6000	801.41	-	3.30%

Table 5: Finite Workforce Simulation Results

Dispatching Rule	Throughput(Pieces)	Cycle Time	Improvement Throughput(Pieces)	Improvement Cycle Time
FCFS	5150	1665.49	-	-
EDD	5200	1624.19	0.90%	2.50%
Dynamic Dispatching Rule	5250	1602.53	2%	3.70%

For the simulation results of infinite workforce, the capacity of each workstation is much greater than the production schedule. The three dispatching rules, EDD, FIFO, and dynamic dispatching rule, came up with the same throughput volume, but we can show some improvements from the cycle time. For the results of finite workforce, we set three operators for simulation and consider their operation time. Throughput volume has decreased about 800 pieces, and the average cycle time has increased one time greater than the results that only workstations are involved.

5. Conclusion

This study addressed the problem how operators or workers will influence work flow performance by different dispatching rules in the production line. We understand that when an operator is involved during production, and adjust their work assignment by considering different rules or priorities. By considering workforce into a dynamic dispatching rule, we can have a better efficiency then taking first-come first-serve (FCFS) or Earliest Due Date (EDD) methods. For infinite workforce, it assumes the capacity of each workstation accommodate the demand, it generates greater throughput than the real output with a dynamic dispatching rules. However, for finite workforce, we need to provide a practical procedure for operators to understand their working order and priority. By taking a semi-automatic fab line as an example, we investigate a fab line when the operators are involved. Through a dynamic dispatching rule and decision flow, a certain layer of fab line may decrease nearly 3 percent of cycle time.

Regarding the linkages between this research and the previous study, we understand that Li et al (2018) provides a dispatching rule with decision flow considering workload, other research provides different perspective on different indicators. Furthermore, we present to provide a compound rule with the basis of different articles. Besides considering due date and machine workload, we also took queue time and workforce as indicators and conducted comparisons by simulation. Our research is based on Li's decision flow, re-design a flow chart considering queue time, production lot priority, due date, and work balance. Besides, this research considers how workforce are involved. We fill the gap for fab lines that are facing transformation from industry 3.0 to 4.0, even before fully automation.

For the future research, it will be valuable to expand the production layer to a more complicated production line, and we are looking forward to find a more significant result to prove the importance of dispatching rules considering finite workforce.

References

Balle, F. & Balle, M. (2006). Feel the force of flexible manpower. *Manufacturing Engineer.*, *84*(6), 20-25.

Bergmann, S., Feldkamp, N., & Strassburger, S. (2015). Approximation of dispatching rules for manufacturing simulation using data mining methods. *2015 Winter Simulation Conference* (pp.2329-2340.), Huntington Beach, CA.

Chen, H. Z., Liu, C., Rong, Y., & Zhu, J. (2012). Research on dynamic dispatching rule for semiconductor assembly production line. *2012 IEEE International Conference on Industrial Engineering and Engineering Management* (pp.493-497). Hong Kong.

Cheong, W. K., Look, S. O., & Mohamed, H. (2007). Improving direct labour productivity through minimizing time wastage approach. *2007 International Symposium on Semiconductor Manufacturing* (pp.1-4). Santa Clara, CA.

Cho, K. H., Chung, Y. H., Choung, Y. I., Park, S. C., & Kim, B. H. (2017). Two boundary based dispatching rule for on-time delivery and throughput of wafer fabs with dedication constraints. *2017 Winter Simulation Conference* (pp.3555-3564). Las Vegas, NV.

Cui, M. & Li, L. (2018). A closed loop dynamic scheduling method based on load balancing for semiconductor wafer fabrication facility. *2018 IEEE International Conference on Smart Manufacturing, Industrial & Logistics Engineering (SMILE)* (pp.1-6). Hsinchu.

Dequeant, K., Vialletelle, P., Lemaire, P., & Espinouse, M. L., (2016, December). A literature review on variability in semiconductor manufacturing: The next forward leap to industry 4.0. *2016 Winter Simulation Conference* (pp.2598-2609). Washington, DC.

Li, L., Qiao, F., Jiang, H., & Wu, Q. (2004). The research on dispatching rule for improving on-time delivery for semiconductor wafer fab. *ICARCV 2004 8th Control, Automation, Robotics and Vision Conference* Vol. 1. (pp.494-498). Kunming, China.

Park, S. C., Kim, B. H., Ahn, E., Chung, Y., & Yang, K. (2013). Fab simulation with recipe arrangement of tools. *2013 Winter Simulation Conference* (pp.3840-3849). Washington, DC.

Yoon, S., & Jeong, S. (2018). Line balancing strategy for re-entrant manufacturing. *IEEE Transaction on Semiconductor Manufacturing, 31*(1), 42-51.

About Authors

Po-Hsun Chen is a master student in the department of Industrial Engineering and Enterprise Information at Tunghai University Taichung, Taiwan ROC.

Chuang-Chun Chiou is a professor in the department of Industrial Engineering and Enterprise Information, at Tunghai University, Taichung, Taiwan ROC. His research interests include supply chain management, healthcare logistics, lean management in service and manufacturing, and data analytics in operations management. His works have been published in European Journal of Operational Research, Annals of Operations Research, Journal of Chinese Institute of Industrial Engineering, Journal of Quality, Journal of Operations Research Society of Japan, and International journal of Production Economics、International Journal of Operational Research and many other refereed journals.