International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 1 (2019), pp. 50-57

Control Policies of Batch Processors in Semiconductor Wafer Fabrication

Pyung-Hoi Koo[1]*
[1] *Department of Systems Management and Engineering, Pukyong National University, Yongsoro 45, Namgu, Busan, South Korea (ORCID: 0000-0002-1695-3039)*

Dug Hee Moon[2]
[2] *Department of Industrial and Systems Engineering, Changwon National University, Changwon, Korea.*

Won-Chul Seo[3]
[3] *Graduate School of Management of Technology, Pukyong National University, Busan, South Korea.*

Abstract

Wafer fabrication (wafer fab) is a manufacturing procedure composed of many different processes to build electrical circuits on wafers layer by layer. Among the processes, there exist batch processes (BPs) where multiple wafer lots are processed on a BP machine simultaneously as a batch. The batch processing leads to a non-smooth product flow due to frequent batching (before BP processing step) and splitting (after BP processing step). In addition, the processing time at the BP station is very large compared to the processing time at discrete processing stations. These characteristics make the BP stations have a critical effect on the performance of the system-wide wafer fab. In most real-world wafer fabs, the production facilities are controlled in realtime due to such uncertainties as machine failure and quality problems. This paper examines control strategies for batch processors in semiconductor manufacturing. We come up with some BP control issues that are important but mostly ignored in existing studies. A few preliminary experimental results are presented to support our findings.

Keywords: batch processors, realtime control strategies, dispatching, batch loading, wafer fabrication system

1. INTRODUCTION

In semiconductor wafer fabrication systems (wafer fabs), electronic circuit layers are repeatedly built on wafers by a variety of processes. Wafers move through the processes in lots each of which generally consists of 20~25 individual wafers. Each wafer lot visits (reenters) the same workstation several times up to forty times. In general, a wafer visits 300-700 process steps on hundreds of different machines in the wafer fab. Because of the long sequences of operations required for the wafers, most wafer fabs suffer from high work-in-process (WIP) inventory and long lead times (about one month). In addition, the cost of building a wafer fab is enormous, often requiring more than ten billion dollars [1]. The complexity involved in the wafer fab and the large investment required make efficient scheduling and control strategies very important for higher fab performance.

The wafer fabs consist of batch processors (BPs) as well as discrete processors (DPs). The BPs can process several wafer lots simultaneously as a batch while the DPs process wafers one at a time. In many wafer fab, more than one-third of the wafer fab operations are performed on batch processors [2]. This paper address the control problems in the BP stations. Diffusion furnaces are a typical example of the batch processors where a number of wafer lots are placed in a reactor, which is then sealed, heated and filled with a carrier gas for changes of their electrical and chemical characteristics [3]. Due to the chemical nature of the process, it may be impossible to process jobs with different recipes together in the same batch. The wafer lots with the same recipe can be viewed as a job family and all the wafer lots in the same job family have the same processing time for a process step. Fig 1 shows a schematic representation of a BP. The wafer lots (often of different types) arriving at the BP station is formed as a batch before being served by a processor. Because of the machine or process constraints, there is a limitation to the number of lots that can be included in a batch. Processing time at the BP stations is long compared to the processing time at the DP stations. Since the wafer lots in a batch are processed together and released at the same time, they must be batched or split regularly during their fabrication processes, which leads to non-smooth product flow. With these characteristics, the BPs has a large effect on system performance in terms of throughput, WIP inventory, cycle time and on-time delivery.

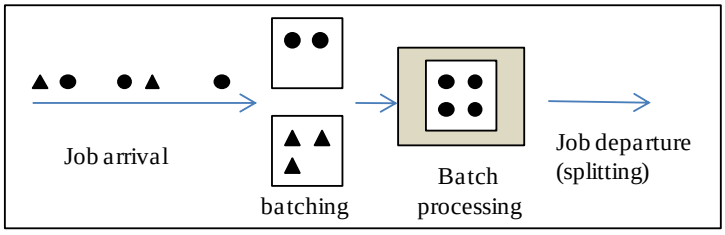


Fig. 1. Schematic representation of batch operations.

Two decision approaches are available to assign jobs to production resources: scheduling and control decisions. Scheduling decisions allocate jobs to resources over time with a horizon of one shift or one day. The scheduling problems are in general static in that decisions are made with data related to jobs and machines which is known in advance. A number of researchers attempt to develop scheduling models to find the optimal schedule [4,5]. It is known that finding an optimal schedule is difficult in general. For example, the optimal job scheduling for a simple production system with only two identical machines with the objective of minimizing makespan is NP-hard [6]. Because of its computation complexity, scheduling literature often focuses on developing a near-optimal schedule by using heuristics [7,8]. Existing works on the BP scheduling are comprehensively reviewed in [3,9,10].

One of the characteristics in semiconductor manufacturing is high uncertainty. Common uncertain events include machine failure, quality problems, urgent orders and so on. The uncertainty reduces the performance of the static scheduling decisions or sometimes makes the schedules infeasible. Therefore, in most real-world wafer fabs, the production facilities are controlled in realtime with only local information or near-future system status information. The operational control decisions are usually made on an event basis. When a BP (resource) becomes available and finds at least one wafer lot waiting in the queue, a BP control decision involves determining (1) the job family to be loaded next (job family selection), (2) the jobs to be put in the next batch (job selection), and (3) the time to be loaded next (loading). The job family selection takes place when incompatible job families are waiting at the queue. The wafer lots in different re-entrant loops in the wafer fab may be considered as different job families because the recipes may depend on the re-entrant loops. The job selection involves selecting the jobs among many jobs of the same job family for the next batch to be processed. The job selection decisions often consider the prioritization of the lots. One simple rule for job selection decision is first-come-first-served (FCFS) policy where the lots arrived earlier have a higher priority to be loaded next. When a due date is related to each job, the urgent jobs may have higher priorities. The loading decision takes place when there are jobs waiting at the queue less than the BP batch capacity. In this case, the BP may wait for more lots to arrive to increase the batch size. There is a trade-off between immediate loading and delayed loading: The immediate loading may underutilize the furnace batch capacity while delayed loading increases the waiting time of the lots that are currently at the queue.

This paper examines the BP control policies developed in semiconductor wafer fabs and attempts to come up with the problems to be addressed to improve system-wide performance. A number of BP control rules have been developed. Earlier work on BP dispatching has been reviewed in [11], [12] and [13]. Van der Zee et al [11] focus on their review on look-ahead BP dispatching policies while Sarin et al. [12] describe BP dispatching problems as a part of wafer fab control problems. This paper extends the previous work of Koo and Moon [13] with recent findings and identifies potential research issues that are important but ignored by the current literature.

BP control policies may be classified into three based on the amount of information considered in control decisions: (1) current information about WIP (work in process) in the BP queue, (2) limited knowledge of future job arrivals, and (3) limited knowledge of system status. For the first category, information currently available about WIP in the BP station queue is considered in the BP control decisions. In the second category, near-future job arrivals are predicted and this information is incorporated in control decisions. Look-ahead control is a term commonly used for this category. For the third category, system status of upstream and downstream stations is considered in the control decisions (We will call this type of control as look-around control). Today, in most wafer fabs, the production information systems such as MES (manufacturing execution system) and MCS (material control system) are installed in the plant to collect shop floor data in real time, which makes decisions for the second and third categories possible.

The remainder of this paper is organized as follows. In the following three sections, research work is reviewed according to BP control policies with only current local WIP information (Section 2), with limited knowledge of future job arrivals (Section 3), and with limited knowledge of system status (Section 4), respectively. Section 5 discusses some BP control issues that are important but ignored by current literature.

2. MBS-BASED BP CONTROL

Minimum batch size (MBS) policy introduced by Neuts [15] is a typical BP control rule where only WIP levels at the BP station are considered. The MBS-based BP control policy works as follows: Let q be the number of jobs waiting in the BP queue, B be the predetermined value, and C be the capacity of a batch processor. Suppose a BP becomes available and is ready to load a new batch. If $q<B$, the loading is delayed. If $B \le q \le C$, the q lots are loaded immediately. If $q>C$, the C lots are selected and loaded immediately. Neut and Nadarajan [16] extend the work of Neuts [15] by considering multiple batch processors. The determination of the optimal MBS value is the main research topic in this strategy. Deb and Serfozo [17] provide a stochastic dynamic program to determine the optimal MBS value under Poisson arrival assumption with the objective of minimizing costs consisting of service costs and waiting costs. They prove that the MBS-based BP control policy is optimal when product arrivals follow the Poisson process and the batch service times are independent and identically distributed (IID). Makis [18] presents an MBS-based BP control policy in which customers waiting times cannot exceed a given constant value. Chandra and Gupta [19] relate batching problem and lot release problem together. They first examine a number of batch sizes to determine the optimal threshold value and then use this value for determining release quantities of wafer lots onto the shop floor so that the total lead time is minimized. Avramidis et al. [4] present an algorithm to determine the optimal threshold value for minimizing the expected number of customers in the system subject to Poisson job arrivals. They

insist that at high traffic intensities system performance is relatively insensitive to the threshold value, and at lower traffic intensities, it is better to set the threshold value low than high.

Kim et al. [20] study the control problems for parallel batch processors with multiple product types of different due dates and different process flows. Two threshold control policies have been presented, MMBS(modified MBS) and PUCH (processing urgency classification heuristic). MMBS is modified from MBS by considering slack time for each product. Here, the job family is selected first based on the average slack time, and then loading decision is made based on the MBS value. In PUCH policy, information on urgencies of lots and the number of waiting lots in the queue is used to select families to be processed first. Here, very urgent lots can be processed as a batch although the number of lots may be smaller than the MBS value for its family. Leachman et al. [21] describe their case study performed for a real-world wafer fab and introduce an MBS-based algorithm implemented for the batch processors. They argue that since the BPs are not bottleneck machines in the wafer fab, operational decisions at the batch processors should be made based on the realtime local information, not based on the off-line detailed schedules.

The MBS-based BP control policies are easy to implement on the shop floor because they require minimal computation with only local WIP information. Hence, it is known that the MBS-based control policies are most widely used in real-world wafer fabs.

3. BP CONTROL WITH LOOK-AHEAD POLICY

Look-ahead control strategies are based on the premise that near-future knowledge about job arrivals will result in a better control decision at the batch processors. We classify the look-ahead control policies into two in terms of the control objectives, lead time minimization and throughput maximization.

3.1 Look-ahead BP control policies for lead time minimization

Lead time is one of the critical performance measures in semiconductor manufacturing. Little's law describes a direct relationship between WIP inventory, throughput rate and lead time: $L = \lambda W$ where L is the WIP inventory, λ is the throughput rate and W is the lead time [22]. Given, throughput rate, minimizing mean production lead time is equivalent to minimizing mean WIP levels. The production lead time is largely composed of time spent processing, transport time, and time spent waiting in the queue. Among these, waiting time accounts for most of lead time. Hence, minimizing waiting time is often a primary performance measure in BP dispatching decisions. The lead time also affects the yield rate of the wafers [21, 23]. Given a constant WIP level, the Little's law indicates that decreasing lead time is equivalent to increasing throughput rate. Hence, the lead time can be used as a surrogate performance measure for throughput rate.

Glassey and Weng [24] is known to be the first to utilize near-future job arrival information for realtime BP control. They present dynamic batching heuristic (DBH) for a batch processing station with a single batch processor and a single job family with processing time T. DBH is activated when a batch processor becomes idle and there are $q \geq 1$ wafer lots waiting in the queue. At time epoch t_o (current decision time), DBH first examines q against the capacity of the batch processor, C. If $q \geq C$, C lots are loaded and processed immediately. If $q < C$, DBH makes batching decision to minimize the total delay time, given the forecasted job arrivals, L, during planning horizon, T. The amount of delay time for products in queue that are waiting for future arrivals is compared to the amount of delay time that can be saved for the future arrivals by waiting until the arrivals occur. Fig 2 shows a waiting time change when we wait for one incoming lot and load $q+1$ lots at time epoch t_1 where t_i represents the arriving time epoch of the next i^{th} lot after t_o. The wait time saving with the delayed loading can be calculated as $\text{Area}_2(t_1) - \text{Area}_1(t_1)$. Here, $\text{Area}_1(t_1) = q(t_1 - t_o)$ and $\text{Area}_2(t_1) = T + t_o - t_1$. The saving time is calculated for all the time epochs of L and the one with the largest positive value is selected for the loading time. If all the wait time savings are negative, the machine starts processing the current WIP immediately. Simulation results show that DBH outperforms MBS-based control policies even with some errors in the arrival times.

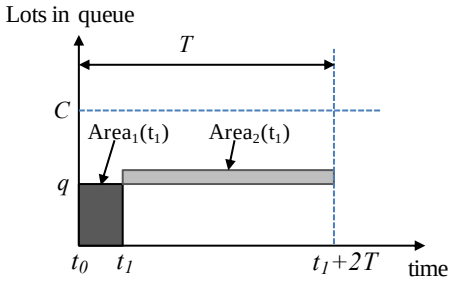


Fig. 2. Wait time saving when the lots are loaded at time t_1.

Fowler et al. [25] present a control heuristic named NACH (Next Arrival Control Heuristic) where the only first future job arrival is additionally considered in the loading decision. They argue that NACH requires fewer data compared to DBH policy and it is robust even with arrival prediction errors. Fowler et al. [26] extend their work for multiple BP cases by presenting NACHM heuristic. Later a number of BP control policies have been presented with some modifications of DBH and NACH. These include Minimum Cost Rate (MCR) heuristic [27] with varying look-ahead horizon, Rolling Horizon Cost Rate (RHCR) heuristic [28] which is a combination of NACH and MCR, and Dynamic Job Assignment Heuristic (DJAH) introduced by Van der Zee et al. [29] where setup cost is considered.

A few works of literature address BP control problems for two-stage manufacturing systems. Cigolini et al. [30] present a BP control heuristic based on the wait-no-longer-than-time (WNLTT) with the objective of minimizing the flow time.

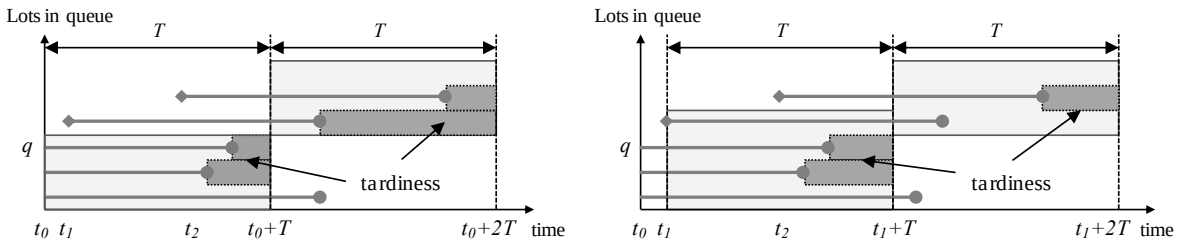
The performance of the WNLTT is examined by using a simulation model with a re-entrant tandem $\beta \rightarrow \delta$ network. They insist that the WNLTT performs well especially when the BP station is a bottleneck resource. Tajan [31] presents a BP dispatching policy based on model predictive control (MPCH) for a tandem $\delta \rightarrow \beta$ network. The MPCH optimally solves a deterministic problem with the truncated horizon at each decision point, and only the first loading batch is implemented and the rest are discarded. Cerekci and Banerjee [32] propose a rolling horizon look-ahead batch control strategy called NARCH (Next Arrival Re-sequencing based Control Heuristic) for a tandem $\delta \rightarrow \beta$ network which incorporates the sequence decisions on the upstream processor through a re-sequencing approach to improve the mean lead time performance of the batch processor. Through simulation experiments, they insist that re-sequencing approach improves the lead time performance of the tandem system as compared to the NACH and MBS-based policies, especially when the number of product types is large and the batch processor traffic intensity is low or moderate.

3.2 Look-ahead BP control policies for tardiness minimization

Recently, job-related performance measure including due-date satisfaction and on-time delivery has received more attention. In a wafer fab where non-memory ASIC (application specific integrated circuit) chips are produced, customer orders with specific due dates often initiate manufacturing. In this case, tardiness is the most commonly used performance measure for on-time delivery. Fig 3 shows the tardiness change for an immediate loading case and a delayed loading case, respectively (imported from [33]) for the situation in which a batch processor is available at time t_0 and finds three products waiting in the queue. Additional products are expected to arrive at times t_1 and t_2. The arrival time and due date for each product are expressed as the start (with a diamond) and the end (with a circle) of the line. The tardiness value for each product is expressed in the shade for each case. The expected tardiness is calculated for the immediate loading and the delayed loading, respectively, and then the loading time with the least expected tardiness is selected.

Kim et al. [20] address BP dispatching problems for the BP stations with parallel machines producing multiple job families with different due dates and different product flow. They present MDBH, a modified version of DBH [24] where due dates are considered in decision making. Here, a job family with the least average slack time is selected to be processed next on an available machine. Once the job is selected, the time to load a batch is determined in such a way that total weighted waiting time in minimized like in the DBH heuristic. Monch and Habenicht [34] propose two BP dispatching heuristics, SBDH (static batch dispatching heuristic) and DBDH (dynamic batch dispatching heuristics) to minimize total weighted tardiness. Their decision processes are based on the ATC (apparent tardiness cost) dispatching rule developed by Vepsalainen and Morton [35]. The DBDH considers future job arrivals while the SBDH does not. Slack times are considered to select the lots to form a batch in both heuristics. Gupta et al. [36] develop a BP dispatching rule, LAB (Look Ahead Batching), for a single batch processor system in which decisions are made considering the arrival epoch and due dates of incoming lots. A conjunctive simulated approach is used to minimize both the average and variation of the tardiness of jobs in a batch. Gupta et al. [37] improve the LAB by considering not only tardiness but also lateness. Sha et al. [38] develop a look-ahead batch control rule (LBCR) where the jobs with urgent due dates have the high priority to process. The LBCR combines with the methodologies of critical ratio(CR) based dispatching rule and the existing NACH policy to increase the rate of on-time deliveries and decrease the job's waiting time. Simulation experiments show that the LBCR decreases the tardiness without significant deterioration of system performance such as flow time.

Cerekci and Banerjee [39] propose a look-ahead BP control strategy for a tandem $\delta \rightarrow \beta$ network for minimizing the mean tardiness. Here, a re-sequencing is employed in BP control decisions to improve the BP control decisions. The idea is that an urgent product is pulled to the front in process sequence at the upstream DP station if there is a benefit in doing so. Mansoer and Koo [33] present a look-ahead control strategy for tardiness minimization on multiple product types. Their model considers a fixed number of future arrivals in decision making with which a fair comparison is made between any decision epochs. They show their strategy outperforms the previous models including the LAB. Later, Koo and Mansoer [40] extend their previous work by considering parallel batch processors. In their model, unlike the other control policies, even when the number of products in the queue is greater than capacity, loading can be delayed for more urgent products to be expected to arrive shortly.



(a) Loading at time t_0 (b) Loading at time t_1

Fig. 3. Total tardiness when the loading time is at t_0 (a) and t_1 (b)

4. BP CONTROL WITH LOOK-AROUND POLICY

The stations in a wafer fab are connected to each other in a physical or logical way. There are often several discrete processing steps between any two batch processing steps in wafer fabs. It is also seen that the BP stations are located back to back along the product flow. The decisions at the BP stations may have a significant effect on the performance of the connected stations. A few research work has used information of upstream and downstream stations to develop BP dispatching policies. Robinson et al. [28] present RHCR-S with the objective to minimize the average lead time for a $\beta \rightarrow \delta$ network where both upstream and downstream information on the current and expected states of a wafer fab is exploited and used for BP control decisions. Experimental results show that the use of upstream and downstream information can lead to performance improvements at the light to moderate traffic intensity, but the improvement vanishes under high traffic conditions. Kim et al. [41] present a rule called BFQL (back and front queues leveling) in which control decisions are made in such a way that the workloads of batch processors and their downstream machines are well balanced. In this rule, batches are classified by the workstation to which they are to be transferred after completion at the BP station. If an immediate downstream station is expected to starve, then the batch required by the station that is expected to begin starving earliest is scheduled next in the BP station.

Neale and Duenyas [42] present a BP dispatching heuristic, TCLH(Two Control Limit Heuristic) for a $\delta \rightarrow \beta \rightarrow \delta$ network, in which the states of upstream and downstream machines are considered. They insist through simulation experiments that the benefit of utilizing information about the state of an upstream discrete machine appears to be an order of magnitude larger than that of utilizing information about the state of a downstream discrete machine. For a $\beta \rightarrow \delta$ system, simply operating the batch processor as if it were stand-alone results in an average number of jobs that is within 1% of best experimental results. Their results are consistent with those of Robinson et al. [28]. They conclude that it is more critical to consider the current state of an upstream discrete machine when controlling a batch processor which is the part of a larger manufacturing network, and utilizing the state of a downstream discrete processor is not that important. Solomon et al. [43] develop an extension of NACH policy for a $\beta \rightarrow \delta$ system that incorporates knowledge about future arrivals and the status of critical machines in downstream processing into the batch processing decision process. The idea is to balance the time for the lots to spend waiting at a BP station with the time spent in the setup at a downstream DP station, thus improve the overall lead time. Koo et al. [44] present a BP dispatching policy where the status of the downstream bottleneck machines are considered in the decision. In their policy, the loading decision is made with an MBS rule in an ordinary situation. However, if the downstream bottleneck machine is expected to be idle shortly, the wafer lots are loaded with higher priority, even with batch size less than MBS. Through simulation experiments, they show that the BP control decisions have a greater effect on the system performance than DP control decisions.

5. DISCUSSIONS AND CONCULSIONS

In the previous sections, we examine existing literature on BP control problems in semiconductor wafer fabs. It is found that a number of research works have been performed in various directions. Throughout the investigation, we have had questions about what the main objective of the BP control problems is. It is believed that the major goal of the BP control decisions is to realize high fab-wide performance. From the system's perspective, we come up with some important issues that have been ignored in existing literature for the wafer fabs to be more productive. Here, we present four potential directions for future research.

First, most research works on BP control has focused on the performance of the stand-alone BP station with the objective of minimizing waiting time reduction or on-time delivery. We believe that the BP dispatching decisions need to be made from a system's perspective. The batch processors account for only a portion of the entire semiconductor manufacturing systems. Due to the re-entrant feature of product flow, the products return to the same processing equipment several times. Most of the control rules developed so far address the isolated problem of optimizing the local performance of batch processors. Their application in multi-stage production systems may lead to less optimal results as far as overall system performance is concerned. Sometimes, locally optimized decisions may deteriorate the system-wide performance. For example, lead time minimization at the batch machines (This is the objective of most previous research works.) may deteriorate the system-wide performance. It may lead to unbalanced and excessive WIP level throughout the manufacturing system, resulting in increased system lead time. To examine this argument, we have run simulation experiments for a wafer fab with 24 workstations, obtained from Wein [45]. Each lot has a process flow with 172 operations at 24 different workstations among which there are four batch processors. Fig 4 shows a local performance of BPs and system-wide performance when two different BP control strategies, MBS and MBS-BM, are applied. The MBS-BM is a modified version of MBS with some additional consideration of the state of downstream bottleneck station. It is seen that MBS-BM provides better performance than MBS in terms of system-wide lead time while it gives worse performance in terms of local lead time at BP station. The result says that the locally good control policy does not mean globally good policy. Identifying the relationship between local performance measures at batch processors and the system-wide performance measures is an interesting topic for future works.

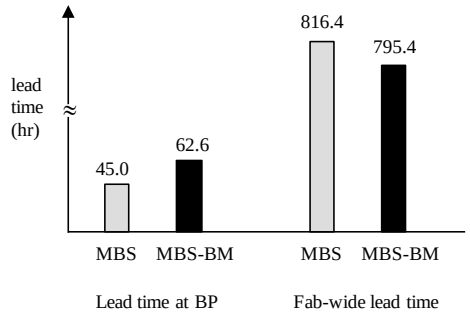


Fig. 4. Performance comparison between local performance and global performance

Secondly, for controlling a wafer fab, there are a variety of shop floor control issues in addition to BP dispatching problems, including lot releasing and DP dispatching for non-bottleneck and bottleneck stations. Identifying the relationship among these shop floor operational decisions has been ignored in the literature. To examine the relationship between lot release rules and BP control rules, we run simulation experiments with two lot release policies and two BP control policies. The lot release policies include an open-loop control policy, CONTIME where wafer lots are released to the system at a constant time interval, and a closed-loop control policy, WR (workload regulation) in which wafer lots are released when a workload for the bottleneck stations falls below a predefined level. The constant inter-arrival time of the CONTIME rule is chosen to have 91.9% utilization for the bottleneck station. The parameters for the WR policy are chosen through preliminary experiments so that the average throughput rate is the same as the average throughput rate of the CONTIME lot release case. Fig 5 shows that the performance of the BP control policies depends on the lot release policies. When the CONTIME lot release policy is used, the MBS-BM provides less lead time than the MBS policy. However, when the WR rule is used, the two BP control rules provide similar performance. Extensive research works are called for to identify the relationship among various operational control schemes.

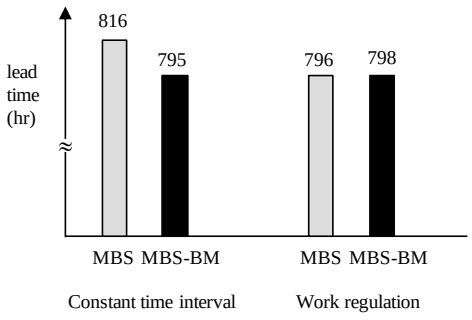


Fig. 5. Relationship between lot release rules and BP control rules

Thirdly, it is important to select an appropriate performance measure for BP control decisions. Most previous research works on BP control decisions attempt to minimize flow-related performance measures including lead time and WIP levels. Only a few pieces of recent researches consider due-date related performance. We insist that the due-date related performances may be more important than the flow-related performance measures because of following two reasons. First, customer satisfaction may be the highest competitive strength in today's fiercely competitive business environment. Delivering products to customers on time is critical for the customer's satisfaction. The second reason is related to the re-entrant product flows. In general, the lithography operation is the bottleneck process that determines the throughput rate of the entire system. Hence, the lithography equipment must be utilized in full capacity. The lost time at the bottleneck machine means the lost time of the whole wafer fab. Not much work has been done about BP control decisions incorporated into the bottleneck station operations. As emphasized in TOC (theory of constraints) philosophy [46], in order to fully utilize the bottleneck machine, all the other stations including BP stations should feed the products in a timely manner to the bottleneck stations smoothly to prevent the bottleneck stations from being idle due to starvation. In general, production schedules are constructed in detail for the bottleneck machines. Since too much WIP in front of the bottleneck machine is not preferable, delivering products too early is not also recommended. Therefore, both tardiness and earliness may be considered in control decision making.

Finally, even though the look-ahead policies mostly provide better performance than the threshold policies, most real-world wafer fabs use MBS-based BP control policies [21]. This is because the threshold policies are easy to implement and future information used in look-ahead control policies is often incorrect due to unpredictable problems such as equipment malfunctions, product quality problems, urgent orders, and so on. It is worthwhile to answer the question how much we can gain from using more complex control rules. More efforts should be made to develop simple and robust look-ahead control rules for real-world applications.

ACKNOWLEDGEMENT

This work was supported by a Research Grant of Pukyong National University(2017 year).

REFERENCES

[1] Ham, M., 2012, Integer programming-based real-time dispatching (i-RTD) for wet-etch station at wafer fabrication, International Journal of Production Research, 50(10), pp. 2809-2822.

[2] Monch, L., Fowler, J., Mason, S., 2013, Production Planning and Control for Semiconductor Wafer Fabrication Facilities: Modeling, Analysis and Systems, Springer, New York, U.S.A.

[3] Mathirajan, M., and Sivakumar, A., 2006, A literature review, classification and simple meta-analysis on scheduling of batch processors in

semiconductor, International Journal of Advanced Manufacturing Technology, 26, pp. 990-1001.

[4] Avramidis, A.N., Healy, K.J. and Uzsoy, R., 1998, Control of a batch-processing machine: A computational approach', International Journal of Production Research, 36(11), pp. 3167-3181.

[5] Fowler, J.W., Phojanamongkolkij, N., Cochran, J.K. and Montgomery, D.C., 2002, Optimal batching in a wafer fabrication facility using a multiproduct G/G/c model with batch processing, International Journal of Production Research, 40(2), pp. 275-292.

[6] Liu, L.L., NG, C.T. and Cheng T.C.E., 2007, Scheduling jobs with agreeable processing times and due dates on a single batch processing machine, Theoretical Computer Science, 374, pp. 159-169.

[7] Jia, Z.H. and Leung, J.Y.T., 2015, A meta-heuristic to minimize makespan for parallel batch machines with arbitrary job sizes, European Journal of Operational Research, 240(3), pp. 649-665

[8] Monch L. and Roob, S., 2018, A metaheuristic framework for batch machine scheduling with incompatible job families and regular sum objective, Applied Soft Computing, 68, pp. 835-846.

[9] Potts, C., and Kovalyov, M., 2000, Scheduling with batching: A review, European Journal of Operational Research, 120, (2), pp. 228-249.

[10] Monch, L. and Fowler, J., Dauzère-Pérès, S. Scott J., Mason, S. and Rose, O., 2011, A survey of problems, solution techniques, and future challenges in scheduling semiconductor manufacturing operations, Journal of Scheduling, 14, pp. 583-599.

[11] Van der Zee, D., 2003, Look-ahead strategies for controlling batch operations in industry-an overview, Proceedings of Winter Simulation Conference, pp. 1480-1487.

[12] Sarin, S. C., Varadarajan, A. and Wang, L., 2011, A survey of dispatching rules for operational control in wafer fabrication, Production Planning & Control, 22(1), pp. 4-24.

[13] Koo, P.H and Moon, D.H., 2013, A Review on Control Strategies of Batch Processing Machines in Semiconductor Manufacturing, IFAC-PapersOnLine, Series Title: Manufacturing Modelling, Management, and Control, 7(1), pp. 1690-1695.

[14] Neale, J. and Duenyas, I., 2003, Control of a batch processing machine serving compatible job families, IIE Transactions, 35, pp. 699-710.

[15] Neuts, M.F., 1967, A general class of bulk queues with Poisson input, Annals of Mathematical Statistics, 38(3), pp. 759-770.

[16] Neuts, M.F. and Nadarajan, R., 1982, A multiserver queue with thresholds for the acceptance of customers into service, Operations Research, 30(5), pp. 948-960.

[17] Deb, R. K. and Serfozo, R. F., 1973, Optimal control of batch service queues, Advances in Applied Probability, 5, pp. 340-361.

[18] Makis, V., 1985, Optimal control of batch service queueing system with bounded waiting time, Kybernetiks, 27, pp. 262-271.

[19] Chandra, P. and Gupta, S., 1997, Managing batch processors to reduce lead time in a semiconductor packaging line, International Journal of Production Research, 35(3), pp. 611-633.

[20] Kim, Y.D., Kim, J.G., Choi, B. and Kim, H.U., 2001, Production scheduling in a semiconductor wafer fabrication facility producing multiple product types with distinct due dates, IEEE Transaction on Robotics and Automation, 17(5), pp. 589-598.

[21] Leachman, R.C., Kang, J. and Lin, V., 2002, SLIM: Short cycle time and low inventory in manufacturing at Samsung Electronics, Interfaces, 32(1), pp. 61-77.

[22] Hopp, W.J. and Spearman, M.I., 2001, Factory Physics: Foundation of Manufacturing Management, Irwin McGraw-Hill, New York, USA.

[23] Kenyon, G, Canel, C. and Neureuther, B.D., 2005, The impact of lot-sizing on net profits and cycle times in the n-job, m-machine job shop with both discrete and batch processing, International Journal of Production Economics, 95(3), pp. 263-278.

[24] Glassey, C. and Weng, W., 1991, Dynamic batching heuristic for simultaneous processing, IEEE Transactions on Semiconductor Manufacturing, 4(2), pp. 77-82.

[25] Fowler, J.W., Phillips, D.T. and Hogg, G.L., 1992,, Real-Time Control of Multiproduct Bulk-Service Semiconductor Manufacturing Processes, IEEE Transactions on Semiconductor Manufacturing, 5(2), pp. 158-163.

[26] Fowler, J.W. Hogg, G.L. and Phillips, D.T., 2000, Control of multiproduct bulk service diffusion/oxidation processes. part 2: Multiple servers," IIE Transactions, 32, pp. 167-176.

[27] Weng, W. and Leachman, R.C., 1993, An improved methodology for real-time production decisions at batch-process work stations, IEEE Transactions on Semiconductor Manufacturing, 6(3), pp. 219-225.

[28] Robinson, J.K. Fowler, J.W., and Bard, J.F., 1995, The use of upstream and downstream information in scheduling semiconductor batch operations, International Journal of Production Research, 33(7), pp. 1849-1869.

[29] Van der Zee, D., 2001, Harten, A. and Schuur, P.C. On-line scheduling of multi-server batch operations, IIE Transactions, 33, pp. 569-586.

[30] Cigolini, R, Perona, M., Portioli, A. and Zambelli, T., 2002, A new dynamic look-ahead scheduling procedure for batching machines, Journal of Scheduling, 5, pp. 185–204.

[31] Tajan, J., Sivakumar, A., and Gershwin, S., 2011, Control of single batch processor with incompatible job families and future job arrivals, IEEE Transactions on Semiconductor Manufacturing, 24(2), pp. 208-222.

[32] Cerekci, A. and Banerjee, A., 2015, Effect of upstream re-sequencing in controlling cycle time performance of batch processors, Computers & Industrial Engineering, 88, pp. 206-216.

[33] Mansoer, P. and Koo, P.H., 2015, Real-time control of the batch processor for tardiness minimization, ICIC Express Letters, 9(5), pp. 1449-1456.

[34] Monch, L. and Habenicht, I., 2003, Simulation-based assessment of batching heuristics in semiconductor manufacturing, Proceedings of Winter Simulation Conference, pp. 1338-1345.

[35] Vepsalainen, A. and Morton, T.E., 1987, Priority rules and lead time estimate for job shop scheduling with weighted tardiness costs, Management Science, 33(8), pp. 1036-1047.

[36] Gupta, A., and Sivakumar, A., 2006, Optimization of due-date objectives in scheduling semiconductor batch manufacturing, International Journal of Machine Tools and Manufacture, 46, pp. 1671-1679.

[37] Gupta, A., and Sivakumar, A., 2007, Controlling delivery performance in semiconductor manufacturing using look-ahead batching, International Journal of Production Research, 45(3), pp. 591-613.

[38] Sha, D., Hsu, S., and Lai, X., 2007, Design of due-date oriented look-ahead batching rule in wafer fabrication, International Journal of Advanced Manufacturing Technology, 35, pp. 596-609.

[39] Cerekci, A. and Banerjee, A., 2010, Dynamic control of the batch processor in a serial-batch processor system with mean tardiness performance, International Journal of Production Research, 48(5), pp. 1339-1359.

[40] Koo, P.H. and Mansoer, P., 2015, A look-ahead control strategy at parallel batch processors with multiple product types, ICIC Express Letters Part B: Applications, 6(12), pp. 2185-2190.

[41] Kim, Y.D., Lee, D.H., Kim, J.U., 1998, A simulation study on lot release control, mask scheduling, and batch scheduling in semiconductor wafer fabrication facilities, Journal of Manufacturing Systems, 17(2), pp. 107-117.

[42] Neale, J. and Duenyas, I., 2000, Control of manufacturing networks which contain a batch processing machine, IIE Transactions, 32, pp. 1027–1041.

[43] Solomon, L., Fowler, J. W., Pfund, M., and Jensen, P. H., 2002, The inclusion of future arrivals and downstream setups into wafer fabrication batch processing decisions. Journal of Electronics Manufacturing, 11, pp. 149–159.

[44] Koo, P.H., Koh S.G. and Park M.J., 2017, Performance analysis of lot release and dispatching decisions in wafer fabrication, International Journal of Applied Engineering Research, 12(16): 2017, pp. 6300-6305.

[45] Wein, L.W., 1988, Scheduling semiconductor wafer fabrication, IEEE Transactions on Semiconductor Manufacturing, 1(3), pp. 115-130.

[46] Goldratt, E.M., and Cox, J., 1984, The Goal, NewYork: North River Press, USA.