TWO-STAGE LOT SCHEDULING WITH WAITING TIME CONSTRAINTS AND DUE DATES

Tae-Sun Yu
Hyun-Jung Kim
Chanhwi Jung
Tae-Eog Lee

Department of Industrial and Systems Engineering
Korea Advanced Institute of Science and Technology (KAIST)
Daejeon 305-701, REPUBLIC OF KOREA

ABSTRACT

We examine a two-stage lot scheduling problem with waiting time constraints and distinct due dates. Wafer lots in diffusion or etch processes generally have due dates specified for each process stage. Some lots even have more strict time constraints that their waiting times between two or multiple stages should not exceed specified limits. We also wish to minimize the variation of the waiting times at the intermediate buffer, which is detrimental to wafer quality variability. To solve such a scheduling problem, we develop a mixed integer programming model for small problems. Also, we suggest an efficient solution procedure for large problems by adopting the earliest due date policy and propose a timing control strategy.

1   INTRODUCTION

In the semiconductor industry, a wafer fabrication procedure is composed of 300-400 steps, and each process step is performed in designated tools. All wafers pass each step, and their lead time is about two months. In a fab, wafers can be processed one by one in a tool, or several lots, each of which contains 25 wafers, can also be operated together. Moreover, the sequence of machines in which each lot is processed is different, and process requirements are also various with wafer types.

In a fab, it is very important to improve wafer quality and satisfy customer demands, and hence there have been many efforts to increase the yield rate and strictly keep the due date. For the wafer quality, each lot or even each wafer is strictly controlled at every process. The residency time constraint, which limits the sojourn time of wafers in a tool, is a typical example of wafer control for improving the quality. There have been many studies on scheduling automated manufacturing systems, such as a cluster tool, a hoist, and a robotic cell with residency time constraints or delay controls (Lee 2008; Kim, Lee, Lee, and Park 2003; Lee and Park 2005; Rostami and Hamidzadeh 2002; Wu, Chu, Chu, and Zhou 2008).

The residency time constraints are also applied to some wafer lots waiting in a stocker or a buffer especially after etching or diffusion is completed. This is because if the wafer surface is exposed to an atmospheric condition for over a certain period of time, some impurities can be produced on the surface, and hence the wafers can suffer serious quality problems. Then those wafers are eventually discarded. Thus, the waiting time constraint is a deadline in a way which should be strictly followed. Besides satisfying the time constraints, it is an essential factor to keep the waiting time deviation among lots constant for fab operators these days. That is because the large variation of waiting times directly results in quality fluctuation. Some particular process steps involving reactive chemical treatments such as oxidation are much more sensitive to the deviation of the waiting times in a buffer.

Even though we stated above the restriction of waiting times with two consecutive processes, more precisely, it can be placed in a set of process steps. In other words, the waiting time can be controlled between a single pair of two steps or among multiple process steps as illustrated in Figure 1. Typical lot

flows in a fab where multiple process machines and various lot types are related are also depicted in Figure 1. The shaded regions are used to represent the existence of waiting time control policy. For example, we can see that the sojourn time between the completion time in the machine m_i and the starting time of the machine m_k should be controlled not to exceed the waiting time constraint. In addition, some machines, such as m_y, may need to process both lots with waiting time constraints and without these constraints, respectively.
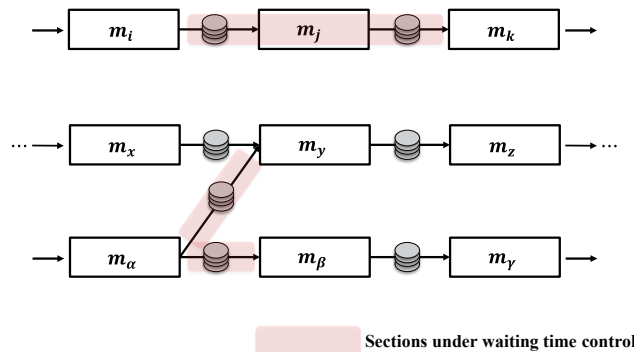

Figure 1: Wafer flow patterns with waiting time constraints.

All wafer lots, in general, have distinct due dates for satisfying customer demands and also have a specific due date at each process step based on the release date. When compared to waiting time constraints, due dates are less strict, so that violations can be tolerated to some reasonable extent. The existence of initial work-in-process in each stage makes the problem more complicated. In other words, the lot scheduling problem in the real industry does not provide an initial state in which the intermediate buffers are empty even though it is a common problem setting or an assumption in machine scheduling studies. Rather, there exist lots already in the intermediate buffers in each process stage at the beginning of scheduling points. To conclude, we investigate the problem involving due dates, waiting time constraints, and waiting time variations together.

Scheduling problems with distinct due dates have been extensively studied in operation research area for various problem structures (Baker and Scudder 1990, Du and Leung 1990). Machine scheduling problems relating to deadlines or maximal time lags are also widely studied with various objective measures (Chen, Potts, and Woeginger 1998). Although the waiting time constraints can be regarded as a special type of maximal time lags, most studies regarding maximal time lags only concern production efficiency measures such as the duration of projects (Herroelen, Reyck, and Demeulemeester 1998). There are only few studies emphasizing the impact of the waiting time constraints. The branch-and-bound algorithm has been developed for minimizing the makespan of two-machine flow shop problems with limited waiting time constraints (Joo and Kim 2009, Yang and Chern 1995). Also, a heuristic algorithm and a mixed integer programming (MIP) model have been proposed for the makespan objective in hybrid two-stage flow shop problems with a batch processor at stage 1 (Su 2003). The flow shop scheduling problems with time constraints between consecutive process steps are examined by developing decomposition approaches based on an MIP model (Klemmt and Mönch 2012). It is known that the optimal schedule for a single machine problem minimizing the waiting time variance has a V-shaped (i.e jobs are arranged so that the job with the smallest process time is located in the middle of the sequence, and the remaining jobs are arranged so that those who precede the shortest job have descending order and the others have ascending order of process times) structure (Eilon and Chowdhury 1977). The relationship between the variance of completion times and the variance of waiting times has been analyzed (Zhou and Cai 1996). It is known that the completion time variance minimization problem is NP-hard even for a single machine problem (Kubiak 1993).

The above problem requirements have been well studied in the conventional machine scheduling area or project scheduling communities. However, to the best of our knowledge, there has been no discussion

for the problems with those requirements together; therefore, minimization of the waiting time variation with problems having time constraints has not been studied yet. These two issues are prevalent in the real industry these days. They are becoming even more significant as the quantity batch production increases in a fab. Thus, our problem is strongly oriented to the latest real needs from the semiconductor manufacturing industry. In this paper, we examine two-stage lot scheduling problems with waiting time constraints as well as due dates where the objective is to minimize the variation of waiting times ($F3|\bar{d}_j|\sum|W_j - \bar{W}|$). The detailed structure of this two stage problem, described in the next section, is designated to sufficiently represent the practical challenges in the real-world wafer fabrication processes. To solve such a scheduling problem, we develop an MIP model for small problems and suggest an efficient approximate solution procedure for large problems by examining the optimal solutions and experimental results.

2 PROBLEM DESCRIPTIONS

2.1 Production Requirements

The real-world problems in the semiconductor industry involve hundreds of process stages and various lot types with distinct flow patterns. For such problems to be globally optimized, we are forced to decide the starting time of every lot at all stages. Thus it seems to be impossible to obtain such a globally optimized lot control due to both the problem size and the uncertainty in a fab environment. Therefore, the fab operators tend to locally optimize the schedule of lots by using dispatching rules on each step. Typical rules used in industry are based on priorities of the lots derived from their job characteristics such as due dates or waiting time constraints.

Although this local optimization on a single or several steps might seem to be quite simple, it is not the case in the real industry. As mentioned in the previous section, all lots have different due dates and process times. Also, some lots can have infinite waiting times, whereas the others have limited waiting times at the intermediate buffer. Furthermore, the fluctuation of waiting times should be reduced for all lots. Therefore, it is extremely complicated to schedule those lots together even in a single machine. For example, higher priorities on lots with tight waiting time constraints will increase the sojourn time of lots without a limited waiting time which results in large fluctuation of waiting times that is critical to quality deviations. In the next section, we define our problem setting which complies to the one that operators in a fab faces.

2.2 Problem Settings

We decompose the fab level problem to a smaller sized one with the purpose of analysis and modeling while maintaining the essential properties appearing in real world problems. The decomposed problem can be defined as a lot scheduling problem of two consecutive process stages. All lots should be processed in each step and have different due dates and waiting times. There are two different machines m_1 and m_2 at the first stage and one machine m_3 at the second stage where all machines are single lot processors. There are two lot types classified by the existence of waiting time constraints between the first and the second stage. The waiting time of a lot can be defined to be the elapsed time from completion at stage 1 to the entering of the process in stage 2. Depending on the lot type, each lot visits one of two machines at the first stage. Regardless of the lot type, every lot should pass through the identical single machine for the second stage. For example, the flow of lots having waiting time constraints are $m_2 \rightarrow m_3$ and the flow of lots without these constraints are $m_1 \rightarrow m_3$ in Figure 2.

We assume the waiting time constraint is distinct for each lot. All lots, regardless of the flow pattern, have a distinct due date which denotes the desirable limit for entering the process at the second stage. Also, each machine has its own buffer and we assume there is no limit on the size of storage capacity because each stocker or a buffer in real fab can basically contain dozens of lots. At the moment of scheduling, there can exist lots in the buffer of stage 2 whose process in the first stage is already completed. This is one of the critical factors that makes our problem difficult because, for those lots in stage 2, we can only control
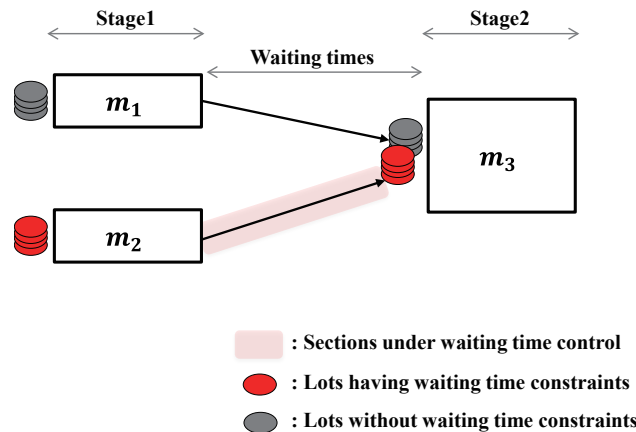
Figure 2: Two stage lot scheduling problem with waiting time constraints.

the timing of entering the process in the second stage. For our simplified problem, we assume there is no due date or time constraint for restricting the entering time of the process in stage 1.

3 OPTIMIZATION MODEL

3.1 Formal Description of the Problem

We describe our problem mentioned above formally. We will assume there are three machines $\{m_1, m_2, m_3\}$ and each of them is a single lot processor. There are N lots $\{J_1, J_2, ..., J_N\}$ to be scheduled and each of N lots has two process stages. We will denote the total set of N lots as $\mathbb{J}$. The first stage should be done by m_1 or m_2 depending on whether a lot has a waiting time constraint, and the second stage should be done by m_3. We assume there is a constraint on the duration of waiting occurring between the completion at m_2 and the starting at m_3. Let $\mathbb{J}_{m_k}$ be a subset of $\mathbb{J}$ including lots waiting for the entering of machine m_k. Also, define another subset $\mathbb{J}_{\text{stage}j}$ in which lots waiting in stage j are included. Therefore, the lots in $\mathbb{J}_{\text{stage}2}$ can be interpreted as the initial lots in the intermediate buffer whose process in the first stage is already completed. Denote the set of lots which have limited waiting times as $\mathbb{J}^L$ and a set of lots allowed to have infinite waiting times as $\mathbb{J}^I$. To obtain the simplicity in formulations, we define $R_{m_k} = \{(i, j) | \text{Both } J_i \text{ and } J_j \text{ need to be processed at machine } m_k\}$ for any machine k.

All lots in $\mathbb{J}$ are given a set of parameters denoting their process requirements. For each lot J_i, the process time at stage j is μ_{ji} where $J_i \in \mathbb{J}$ and $j = 1, 2$. Also, there are due dates and limited waiting times that constrain the entering time of the process at stage 2 where both are distinct for each lot. Thus, let d_i be a distinct due date for a lot $J_i \in \mathbb{J}$. The allowable waiting time for each lot J_i can be denoted as δ_i where the value is positive finite for all $J_i \in \mathbb{J}^L$ and infinite for all $J_i \in \mathbb{J}^I$.

In addition to the parameters used for characterizing the lots to be scheduled, we adopt some constant variables used to reflecting the production environment in the real industry. Let τ be a constant value representing the transport time between stage 1 and stage 2 where the value is identical for every lot. We assume the process in stage 2 is the bottleneck, so that m_3 tends to be busy at any moment. Therefore, we let ω be an initial workload which is the remaining process time of a lot already being processed by the resource m_3 at the beginning of the scheduling point. In other words, any of the lots waiting for the process in m_3 can enter the machine at least after ω amount of time.

3.2 Mixed Integer Programming Model

In this section, we formulate an MIP model to optimize the lot schedule by minimizing the waiting time variation and meeting the due dates and waiting time constraints. For each lot $J_i \in \mathbb{J}$, we introduce decision variables x_i for the starting time of the process in stage 1 and y_i for the starting time at stage 2. There is an

important property that variable x_i takes a negative value when J_i is a lot already in stage 2. In addition, we assume that the elapsed sojourn time of all lots in stage 2 are given as 0. Then, for lots that are already in stage 2 at the beginning, we can have $x_i + \mu_{1i} = 0$ so that the waiting time of these lots can be simply y_i. In addition as we assume the stage 2 is a bottleneck process, we can obtain the minimal makespan for our problem as $\omega + \sum_{J_j \in \mathbb{J}} \mu_{2j}$. Thus, by restricting m_3 not to have idle times, we will find a solution that is optimal to the waiting time variation while the minimum makespan is obtained. This assumption will be adopted in the MIP model by the following constraint.

$$y_i \leq \omega + \sum_{J_j \in \mathbb{J}, j \neq i} \mu_{2j} \quad \forall J_i \in \mathbb{J}$$

The decision variable $z_{ij}^k \in \{0,1\}$ will be used for the precedence relationship between lot J_i and J_j at machine m_k, where $(i,j) \in R_{m_k}$, $i \neq j$ and $k = 1,2,3$. The variable z_{ij}^k takes the binary value as below.

$$z_{ij}^k = \begin{cases} 1 & \text{if lot } J_i \text{ precedes } J_j \text{ at } m_k, \\ 0 & \text{otherwise.} \end{cases} \quad \text{where } k = 1,2,3$$

The variable w_i will be defined for the waiting time of lot J_i and let $\bar{w}$ be the average waiting time of all lots. Thus the variation of waiting times can be described by

$$var(w) = \frac{\sum_{J_i \in \mathbb{J}} |\bar{w} - w_i|}{|\mathbb{J}|}.$$

Thus we define $\sigma_i = |\bar{w} - w_i|$ and take $\sum_{J_i \in \mathbb{J}} \sigma_i$ as the objective function. The conventional big M notation is adopted to represent a sufficiently large value. The MIP model for our problem can be formulated as below.

3.2.1 Mixed Integer Programming Formulation

$$Minimize \quad \sum_{J_i \in \mathbb{J}} \sigma_i \tag{1}$$

Subject to

$$y_i \geq \omega \qquad\qquad \forall J_i \in \mathbb{J} \tag{2}$$

$$y_i \leq \omega + \sum_{J_j \in \mathbb{J}, j \neq i} \mu_{2j} \qquad\qquad \forall J_i \in \mathbb{J} \tag{3}$$

$$x_i + \mu_{1i} = 0 \qquad\qquad \forall J_i \in \mathbb{J}_{\text{stage2}} \tag{4}$$

$$y_i \leq d_i \qquad\qquad \forall J_i \in \mathbb{J} \tag{5}$$

$$y_i - (x_i + \mu_{1i}) \leq \delta_i \qquad\qquad \forall J_i \in \mathbb{J}^I \tag{6}$$

$$x_i + \mu_{1i} + \tau \leq y_i \qquad\qquad \forall J_i \notin \mathbb{J}_{\text{stage2}} \tag{7}$$

$$\bar{w} = \sum_{J_i \in \mathbb{J}} (y_i - x_i - \mu_{1i}) / |\mathbb{J}| \qquad\qquad \forall J_i \in \mathbb{J} \tag{8}$$

$$\sigma_i \geq \bar{w} - (y_i - x_i - \mu_{1i}) \qquad\qquad \forall J_i \in \mathbb{J} \tag{9}$$

$$\sigma_i \geq -\bar{w} + (y_i - x_i - \mu_{1i}) \qquad\qquad \forall J_i \in \mathbb{J} \tag{10}$$

$$x_i + \mu_{1i} \leq x_j + M * (1 - z_{ij}^k) \qquad\qquad \forall (i,j) \in R_{m_k}, \ k = 1, 2 \tag{11}$$

$$x_j + \mu_{1j} \leq x_i + M * z_{ij}^k \qquad\qquad \forall (i,j) \in R_{m_k}, \ k = 1, 2 \tag{12}$$

$$y_i + \mu_{2i} \leq y_j + M * (1 - z_{ij}^k) \qquad\qquad \forall (i,j) \in R_{m_k}, \ k = 3 \tag{13}$$

$$y_j + \mu_{2j} \leq y_i + M * z_{ij}^k \qquad\qquad \forall (i,j) \in R_{m_k}, \ k = 3 \tag{14}$$

$$x_i \geq 0 \qquad\qquad \forall J_i \notin \mathbb{J}_{\text{stage2}} \tag{15}$$

$$y_i \geq 0 \qquad\qquad \forall J_i \in \mathbb{J} \tag{16}$$

The objective (1) is to minimize the variation of waiting times, where each σ_i variable corresponds to the absolute value of the difference between the mean waiting time and the waiting time of lot J_i. The initial workload of resource m_3 is reflected in the constraint (2). Constraint (3) restricts machine m_3 not to become idle so that the makespan can be obtained as the right hand side of the inequality which is the minimum makespan available in the problem. The equations in (4) enforce the lots that are already in the intermediate buffer at stage 2 to have sojourn times as their y_i values. Constraints (5) and (6) ensure that all lots meet their due dates and waiting time constraints. The process requirements that all lots can enter their process at stage 2 only after their process in stage 1, and their transportation are completed, are guaranteed by (7). The equation sets (8)-(10) allow each σ_i to have a value of $|\bar{w} - w_i|$. Resource constraints (11)-(14) are used to avoid conflicts on each single lot processor m_k. The x_i values for lots already in stage 2 can be negative. The remaining variables for denoting entering times are imposed on non-negativity constraints by (15) and (16).

3.2.2 Example 1 : Scheduling problem with 9 lots

We solve a problem with 9 lots by the suggested MIP model. The parameter settings are illustrated in Table 1. We assumed the initial workload ω is given as 0.5 and there are three lots in stage 1 and six lots in stage 2. Each J_i having infinite δ_i values in the Table 1 are the lots without a waiting time constraint.

Table 1: Parameter settings and the optimal solution of example 1.

	Input				Output			
	μ_{1i}	μ_{2i}	d_i	δ_i	x_i	y_i	w_i	**Position**
J_1	0.3	0.5	7	1	-0.3	0.5	0.5	Stage2
J_2	0.2	0.5	1	∞	0	1	0.8	Stage1
J_3	0.15	1	1.5	∞	-0.15	1.5	1.5	Stage2
J_4	0.4	1	7	2.5	-0.4	2.5	2.5	Stage2
J_5	0.3	1	6	4	3.875	5.5	1.325	Stage1
J_6	0.25	0.5	5	2	3.425	5	1.325	Stage1
J_7	0.2	0.5	5	∞	1.975	3.5	1.325	Stage1
J_8	0.3	0.5	6	∞	2.375	4	1.325	Stage1
J_9	0.4	0.5	7	2	2.775	4.5	1.325	Stage1

The optimal objective is given as $\sum_{J_i \in \mathbb{J}} \sigma_i = 2.7$, and the x_i and y_i values are as in the output column in Table 1. The solution given by the MIP model is visualized by the Gantt-chart in Figure 3 where the four bars represent waiting times in stage 1, process times at stage 1, waiting times at stage 2 and the process times at stage 2 for each lot.
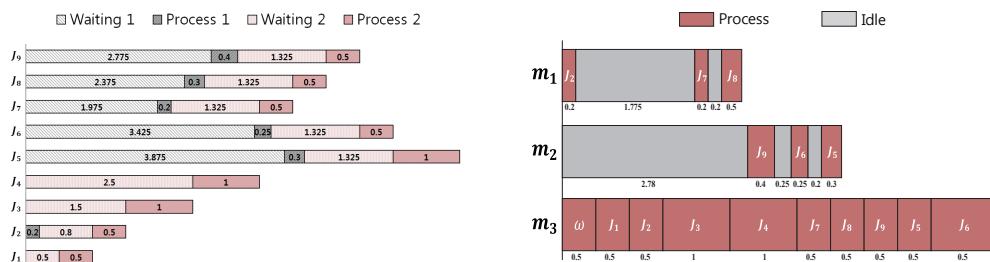

Figure 3: Gantt-chart for the optimal solution on example 1.

Computational experiments are accomplished by generating five cases where the due dates, process times, waiting time constraints are all distinct. For each case the μ_{1i} and μ_{2i} take the values in [0,1] and δ_i takes the value in [0,5]. The due date d_i takes the value in [ω, *Total Workload*] where the *Total Workload* represents the sum of all process times μ_{2i} in a given problem. The results in Table 2 show that the problem can be solved up to 25 jobs depending on the problem structure. Also, when the number of total lots to be scheduled are fixed, the computational time increases as the number of lots in stage 2 increases. The experiments are conducted by ILOG CPLEX 12.5 in Intel Core i3-2120 3.30GHz with 8 GB DDR-3 RAM environment. In the next section, by analyzing the optimal schedules, we suggest an approximate approach that can be adopted for large problems.

Table 2: Experimental results on the MIP model.

	Number of Lots				$\Sigma_{J_i \in \mathbb{J}} \sigma_i$	CPU Time (sec)
	m_1	m_2	m_3	Total # of Lots		
Case 1	3	3	3	9	2.7	0.96
	7	7	4	18	2	1.75
Case 2	7	5	6	18	4.33	1.64
	5	5	8	18	5.6	2.52
Case 3	5	6	9	20	9.25	2.98
	10	7	7	24	8.12	10.53
Case 4	10	6	8	24	8.95	12.06
	8	6	10	24	10.95	69.07
Case 5	9	6	10	25	10.94	78.39
	8	6	11	25	15.98	>2000*

4 APPROXIMATE SOLUTION PROCEDURE

As problems with more than 25 lots are unsolvable with the suggested MIP model, we propose an approximate procedure for obtaining the schedule of the lots. The procedure solves an MIP model on the reduced set of lots and the remaining lots are scheduled by additional steps based on the earliest due date (EDD) sequence and an elaborate timing control policy. Some insightful results and properties are used for the solution approach which were obtained by examining the optimal solutions from the MIP model of the entire job set.

4.1 Properties on Waiting Time Variation

First we decompose the lots in $\mathbb{J}$ into two subsets $\mathbb{J}_{free}$ and $\mathbb{J}_{fixed}$. We apply the MIP model only on the lots in $\mathbb{J}_{fixed}$ and then the lots in $\mathbb{J}_{free}$ are scheduled by an EDD sequence and timing rules together. To define the subset $\mathbb{J}_{fixed}$, let the maximum mean waiting time of lots in stage 2 at the beginning be

$$\bar{w}_{max} = \frac{1}{n} \sum_{i=1}^{n} (n-i)\mu_{2[i]}, \tag{17}$$

where $n = |\mathbb{J}_{stage2}|$ and $\mu_{2[i]}$ is the ith largest process time at stage 2. We adopted the longest process time first rule for finding the value of $\bar{w}_{max}$ in the equation (17). This equation is derived from the results on the minimum mean completion time and the relational properties between the mean completion time and the mean waiting time (Smith 1956, Bagchi 1989). Given this upper bound on the mean waiting time of the lots in stage 2, we newly define the following subsets.

Given the value of $\bar{w}_{max}$, we can define the subset $\mathbb{J}^{L'}$ consisting of the lots in $\mathbb{J}_{stage1}$ having smaller waiting time constraints than $\bar{w}_{max}$ as follows.

$$\mathbb{J}^{L'} := \{J_i | J_i \notin \mathbb{J}_{stage2}, J_i \in \mathbb{J}^L, \delta_i < \bar{w}_{max}\} \tag{18}$$

Now the subset $\mathbb{J}_{fixed}$ is defined as $\mathbb{J}_{fixed} = \mathbb{J}_{stage2} \cup \mathbb{J}^{L'}$ and the remaining lots comprise the subset $\mathbb{J}_{free}$. Intuitively, the lots in $\mathbb{J}_{fixed}$ are interpreted as lots whose waiting time values w_i are determined directly by the sequence of lots in stage 2. In other words, the waiting time of lots in $\mathbb{J}_{stage2}$ are determined by their y_i values and those in $\mathbb{J}^{L'}$ can have their waiting time values only within the range $0 \leq w_i \leq \delta_i$ if their δ_i values are less than or equal to $\bar{w}_{max}$. Thus, these lots comprise the subset $\mathbb{J}_{fixed}$. On the other hand, the waiting time of lots in $\mathbb{J}_{free}$ are freely controllable by deciding x_i even though their y_i values are determined if the x_i values of the lots does not make any resource conflicts at stage 1.

Now assume that we applied the MIP model to the lots in $\mathbb{J}_{fixed}$ and a schedule of lots in this subset is given. Also, assume there is no other constraint and our objective is to minimize the waiting time variation. Then, conceptually, the waiting time w_i for lots $J_i \in \mathbb{J}_{fixed}$, becomes a deterministic constant variable. In addition, the waiting time of lots in $\mathbb{J}_{free}$ can be treated as the decision variables and we can denote these variables as $W = \{W_1, ..., W_{|\mathbb{J}_{free}|}\}$. Now we deal with the waiting time variation of the entire lots to be scheduled as a function of W as in the following formula.

$$|\sigma|(W) = \frac{\sum_{J_i \in \mathbb{J}_{fixed}} |\bar{w} - w_i| + \sum_{J_i \in \mathbb{J}_{free}} |\bar{w} - W_i|}{|\mathbb{J}|}, \quad \text{where } \bar{w} = \frac{\sum_{J_i \in \mathbb{J}_{fixed}} w_i + \sum_{J_i \in \mathbb{J}_{free}} W_i}{|\mathbb{J}|} \tag{19}$$

Then, the following statement can be made.

Theorem 1 The waiting time variation $|\sigma|(W)$ is minimized when $W_i = \bar{w}_{fixed} \equiv \frac{\sum_{J_i \in \mathbb{J}_{fixed}} w_i}{|\mathbb{J}_{fixed}|}, \quad \forall J_i \in \mathbb{J}_{free}$.

The proof of this statement will not be discussed in this paper due to the space limitation. Thus, given a minimum waiting time variation for lots in $\mathbb{J}_{fixed}$ with the mean waiting time $\bar{w}_{fixed}$, it is desirable for the lots in $\mathbb{J}_{free}$ whose waiting times can be controllable to have their waiting times close to $\bar{w}_{fixed}$. If the

lots in $\mathbb{J}_{free}$ can have their waiting time values exactly the same as $\bar{w}_{fixed}$, the waiting time variation of the entire problem is minimized. In other words, the optimal waiting time variation is determined as the minimal waiting time variation of lots in $\mathbb{J}_{fixed}$.

As Theorem 1 assumes no due date constraint and resource constraints, we extend Theorem 1 to reflect these requirements. The main idea of Theorem 1 is extended and formalized as the following when the due dates and resource constraints are also regarded together. First, let $|\sigma|(W_{fixed})$ be the lower bound of the waiting time variation which is obtained by applying the MIP model on lots in $\mathbb{J}_{fixed}$ only. Let $\bar{w}_{fixed}$ be the average waiting time obtained for deriving the $|\sigma|(W_{fixed})$ value. Also, let $S(\mathbb{J})$ be a sequence of lots in $\mathbb{J}$ where the sequence denotes the entering order of process at m_3.

Theorem 2 Let $S(\mathbb{J})$ be an arbitrary sequence of lots in $\mathbb{J}$ where the sequence does not affect the waiting time of lots in $\mathbb{J}_{fixed}$ determined by finding the value of $|\sigma|(W_{fixed})$. For any given sequence $S(\mathbb{J})$, if the following conditions are satisfied, the $|\sigma|(W^*)$ which is the optimal waiting time variation of the entire problem is equivalent to $\frac{|\mathbb{J}_{fixed}|}{|\mathbb{J}_{fixed}|+|\mathbb{J}_{free}|}|\sigma|(W_{fixed})$.

$$\mu_{1[i]} - \mu_{1[j]} + \min(\delta_{[i]}, \bar{w}_{fixed}) - \min(\delta_{[j]}, \bar{w}_{fixed}) + \sum_{i \leq k < j} \mu_{2[k]} \geq 0 \quad \forall ([i],[j]) \in R_{m_k}, i < j, k = 1,2 \quad (20)$$

$$\min_{J_i \in \mathbb{J}_{free}} (\mu_{2i} + d_i) \geq \max_{J_i \in \mathbb{J}_{fixed}} (\mu_{2i} + d_i) \quad (21)$$

$$\min_{J_i \in \mathbb{J}}(y_i - d_i) = 0 \quad (22)$$

In the formula above, $J_{[i]}$ is the ith lot in a given arbitrary sequence $S(\mathbb{J})$. The condition (20) is satisfied when the lots in $\mathbb{J}_{free}$ can have their waiting time as $\bar{w}_{fixed}$ for a given sequence $S(\mathbb{J})$ without making any resource conflict at stage 1. If the initial workload of lots waiting in each machine is well balanced, with the assumption that the stage 2 is a bottleneck process, the condition (20) is easy to be satisfied. The condition (21) guarantees that the due date constraints can be satisfied although the lots in $\mathbb{J}_{fixed}$ precede all the lots in $\mathbb{J}_{free}$ at stage 2. The condition (21) is also easily satisfiable for problems where the lots already in stage 2 can be regarded as the lots that arrived earlier than those in stage 1. However, the condition (22) is highly conservative if the given arbitrary sequence $S(\mathbb{J})$ does not consider the due date constraints for the lots in $\mathbb{J}_{free}$. However, for an arbitrary sequence $S(\mathbb{J})$ where the sequence does not affect the waiting time of lots scheduled from obtaining the lower bound of the waiting time variation, the precedence relationship among the jobs in $\mathbb{J}_{free}$ only affects the due date constraints if the condition (20) is satisfied. In other words, the construction of $S(\mathbb{J})$ can be made without considering the waiting time variation and waiting time constraints. Then, for such a sequence $S(\mathbb{J})$, the conservativeness of the condition (21) can be weakened. To conclude, if any sequence $S(\mathbb{J})$ satisfies the conditions (20)-(22), we can obtain a full schedule of a problem by assigning a waiting time of each lot based on Theorem 1. A method to construct such a sequence $S(\mathbb{J})$ and an exact timing policy for applying Theorem 1 will be discussed in the next section. Finally, we suggest an exact solution approach that can be applied when the obtained lower bound is equivalent to the optimal objective of the original problem. The method provides the full schedule of the problem by solving MIP model only on the lots in $\mathbb{J}_{fixed}$.

4.2 Efficient Solution Approach

We suggest a stepwise procedure for obtaining a full schedule of the problem where the MIP model is only applied to $\mathbb{J}_{fixed}$. As the subset $\mathbb{J}_{fixed}$ is much smaller than the entire set of lots to be scheduled, even a large problem can be solved efficiently. The approach provides a schedule with all time constraints satisfied having a near optimal waiting time variation. The due date constraints, which are the least restrictive constraints, are not guaranteed to be satisfied. However, when the conditions (20) and (21) are satisfied for a given solution, the method provides an optimal objective value with all constraints satisfied. The

necessary condition for the optimality of a schedule given by the suggested approach is derived from the theorems given in Section 4.1.

- Step 1 : Using equations (17) and (18), define the set $\mathbb{J}_{fixed}$ for a given problem

- Step 2 : Solve the MIP model and obtain the values of x_i, y_i, w_i and $\bar{w}_{fixed}$ for the lots in $\mathbb{J}_{fixed}$ and let $\mathbb{J}'_{fixed}$ be defined as follows

$$\mathbb{J}'_{fixed} := \{J_i | y_i \leq y_k, \ \forall J_k \in \mathbb{J}_{fixed} \cap \mathbb{J}_{stage2}, \ J_i \in \mathbb{J}_{fixed}\} \tag{23}$$

- Step 3 : Reserve the schedule of lots in $\mathbb{J}'_{fixed}$ and let S' be the sequence of entering the process in stage 2 of lots not in $\mathbb{J}'_{fixed}$ where the lots are ordered by the earliest due date first policy

- Step 4 : All x_i and y_i value for lots not in $J_{[i]}$, where $J_{[i]}$ is the ith lot in the sequence S', are updated or newly obtained by the following equation

$$y_{[i]} = \omega + \sum_{J_i \in \mathbb{J}'_{fixed}} \mu_{2i} + \sum_{j<i} \mu_{2[j]} \tag{24}$$

$$x_{[i]} = \max(x_{[i-1]} + \mu_{1[i-1]}, \omega + \sum_{J_j \in \mathbb{J}'_{fixed}} \mu_{2j} + \sum_{j<i} \mu_{2[j]} - \mu_{1[i]} - \min(\delta_{[i]}, \bar{w}_{fixed})) \tag{25}$$

From the procedure above, the full sequence $S(\mathbb{J})$ and the entering time (x_i, y_i) for the entire problem can be obtained. If the given sequence $S(\mathbb{J})$ satisfy the condition (20) and the defined subsets $\mathbb{J}_{free}$ and $\mathbb{J}_{fixed}$ satisfy the condition (21), the optimality and feasibility of the schedule are guaranteed. While the proof of the first necessary condition can be followed by introducing the Jackson's lemma on the due date feasibility (Jackson 1955), the detailed proof will not be given in this paper. The second necessary condition follows from the feasibility on the resource constraints in stage 1. As the sequence on the lots in $\mathbb{J}_{free}$ is obtained based on the EDD policy, conditions (22) are satisfied if conditions (21) are satisfied. Even if the condition (21) and (22) are violated, the only violation occurs in due date constraints which is a less restrictive requirement compared to the reduction of waiting time variation and waiting time constraints. The upper bound on the number of lots violating due date constraints for such a schedule can be obtained as follows

$$N_{Tmax} = |\{J_i | J_i \in \mathbb{J}_{free}, d_i \leq \max_{J_k \in \mathbb{J}_{fixed}} d_k\}|.$$

The upper bound N_{Tmax} tends to take a small value for problems having well balanced lot distribution in each stage. Thus, when violations in due dates are allowed, the approach can be applied with efficient computational time with a smaller objective value than the one obtained from applying the MIP model on the entire set of lots. Table 3 shows the efficiency of the suggested solution approach based on set reduction procedures (RMIP). In case 1, the RMIP procedure which applies the MIP model on a reduced set of lots finds the optimal feasible solution with an efficient computation time. The due date constraints are not satisfied for the problem in case 2 by the RMIP model. However, even though the due date constraints are violated, the approximate procedure RMIP yields a better objective value. Thus, there exists a strong trade-off between the improvement in the objective value and the violation of due date constraints when the RMIP does not guarantee a solution with all due dates satisfied.

Works on obtaining a full schedule, which is always feasible to all constraints, for problems having more than 25 lots are still in progress. As the problem requirements consist of due dates, waiting time constraints, and waiting time variations, it is not trivial to obtain an efficient schedule by simple dispatching rules. For example, while the optimal sequence for the waiting time variance measure in a single machine

Yu, Kim, Jung, and Lee

Table 3: Efficiency of the approximate solution approach

| | Number of Lots | | | | $\|\mathbb{J}_{fixed}\|$ | $\Sigma_{J_i \in \mathbb{J}} \sigma_i$ | | Mean Tardiness | Feasibility of RMIP | CPU Time(sec) | |
	m_1	m_2	m_3	$\|\mathbb{J}\|$		MIP	RMIP*			MIP	RMIP
Case 1	8	6	11	25	13	8.88	8.88	0	0	953.6	215.6
Case 2**	6	6	11	25	16	15.99	11.83	1.075	1.075	>2000	118.1

RMIP* : MIP solution on reduced set of lots
Case2** : Case where the due date constraints are violated when solved by RMIP

problem tends to have V-shaped structure, this structure does not provide optimality in the two stage problems and also the feasibility on due dates and time constraints cannot be obtained without appropriate timing decisions (Eilon and Chowdhury 1977).

5 CONCLUSION

We discussed a two-stage lot scheduling problem with waiting time constraints and distinct due dates. We proposed an MIP model for minimizing the waiting time variations and satisfying the due date constraints and waiting time constraints. The MIP model finds the schedule where the bottleneck resource does not become idle so that the makespan measure and the throughput are guaranteed to be optimal. Computational experiments showed that the suggested MIP model solves problems having up to 25 lots. Thus we suggested an approximate solution procedure which solves the MIP model on the reduced set of lots. We also examined some insightful properties appearing in the problem such as the necessary condition on the optimality of the schedule obtained by the approximate solution method. The experimental results showed that the suggested method provides a solution efficiently where the quality of the schedule is also desirable for practical uses. Our ongoing work is to solve a problem with more than two stages and over three machines engaged. Also, an efficient heuristic algorithm which guarantees all constraints satisfied while the waiting time variation is obtained in the near optimal value should be further investigated.

ACKNOWLEDGMENTS

REFERENCES

Bagchi, U. 1989. "Simultaneous Minimization of Mean and Variation of Flow Time and Waiting Time in Single Machine Systems". *Operations Research* 37 (1): 118–125.
Baker, K. R., and G. D. Scudder. 1990. "Sequencing with Earliness and Tardiness Penalties : A Review". *Operations Research* 38:22–36.
Chen, B., C. N. Potts, and G. J. Woeginger. 1998. "A Review of Machine Scheduling: Complexity, Algorithms and Approximability". *Handbook of combinatorial optimization* 3:21–169.
Du, J., and J. Y.-T. Leung. 1990. "Minimizing Total Tardiness on One Machine is NP-HARD". *Mathmatics of Operations Research* 15:483–495.
Eilon, S., and I. Chowdhury. 1977. "Minimising Waiting Time Variance in the Single Machine Problem". *Management Science* 23 (6): 567–575.
Herroelen, W., B. D. Reyck, and E. Demeulemeester. 1998. "Resource-Constrained Project Scheduling: A Survey of Recent Developments". *Computers & Operations Research* 25 (4): 279 – 302.
Jackson, J. R. 1955. "Scheduling a Production Line to Minimize Maximum Tardiness". Technical report, DTIC Document.

Joo, B.-J., and Y.-D. Kim. 2009, April. "A Branch-and-Bound Algorithm for a Two-Machine Flowshop Scheduling Problem with Limited Waiting Time Constraints". *Journal of the Operational Research Society* 60 (4): 572–582.

Kim, J.-H., T.-E. Lee, H.-Y. Lee, and D.-B. Park. 2003. "Scheduling Analysis of Time-Constrained Dual-Armed Cluster Tools". *IEEE Transactions on Semiconductor Manufacturing* 16 (3): 521–534.

Klemmt, A., and L. Mönch. 2012. "Scheduling Jobs with Time Constraints between Consecutive Process Steps in Semiconductor Manufacturing". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. Uhrmacher, 194.

Kubiak, W. 1993. "Completion Time Variance Minimization on a Single Machine is Difficult". *Operations Research Letters* 14 (1): 49–59.

Lee, T.-E. 2008. "A Review of Scheduling Theory and Methods for Semiconductor Manufacturing Cluster Tools". In *Proceedings of the 2008 Winter Simulation Conference*, 2127–2135.

Lee, T.-E., and S.-H. Park. 2005, October. "An Extended Event Graph with Negative Places and Tokens for Time Window Constraints". *IEEE Transactions on Automation Science and Engineering* 2 (4): 319–332.

Rostami, S., and B. Hamidzadeh. 2002, August. "Optimal Scheduling Techniques for Cluster Tools with Process-Module and Transport-Module Residency Constraints". *IEEE Transactions on Semiconductor Manufacturing* 15 (3): 341–349.

Smith, W. E. 1956. "Various Optimizers for Single-Stage Production". *Naval Research Logistics Quarterly* 3 (1-2): 59–66.

Su, L.-H. 2003, March. "A Hybrid Two-Stage Flowshop with Limited Waiting Time Constraints". *Computers & Industrial Engineering* 44 (3): 409–424.

Wu, N., C. Chu, F. Chu, and M. Zhou. 2008, May. "A Petri Net Method for Schedulability and Scheduling Problems in Single-Arm Cluster Tools with Wafer Residency Time Constraints". *IEEE Transactions on Semiconductor Manufacturing* 21 (2): 224–237.

Yang, D.-l., and M.-S. Chern. 1995. "A Two-Machine Flowshop Sequencing Problem with Limited Waiting Time Constraints". *Computers & Industrial Engineering* 28 (1): 63–70.

Zhou, S., and X. Cai. 1996. "Variance Minimization : Relationship between Completion-Time Variance and Waiting-Time Variance". *The Journal of the Australian Mathematical Society* 38 (01): 126–139.

AUTHOR BIOGRAPHIES

Tae-Sun Yu received the B.S. degree in industrial and systems engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon. He is doing his doctoral studies at the Department of Industrial and Systems Engineering, KAIST. His research interests include modeling, scheduling, and control of discrete event dynamic systems in the semiconductor industry. His email address is taesunyu@kaist.ac.kr.

Hyun-Jung Kim received the Ph.D degree in industrial and systems engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, in 2013. Her research interests include modeling, scheduling, and control of discrete event dynamic systems in the semiconductor industry.

Chanhwi Jung received the B.S. degree in industrial engineering in 2005 from Konkuk University, Seoul. He is currently working as senior engineer of memory business at Samsung Electronics. His current research interests are system modeling, scheduling, and classification of data-based in the semiconductor industry.

Tae-Eog Lee received the Ph.D. degree in industrial and systems engineering from The Ohio State University, Columbus, OH, USA, in 1991. He is a Professor with the Department of Industrial and Systems Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, and the Head of the department. His research interests include cyclic scheduling theory, scheduling and control theory of timed discrete event dynamic systems, and their applications to automated manufacturing systems.