ON-TIME DELIVERY ACHIEVEMENT OF HIGH PRIORITY ORDERS IN ORDER-DRIVEN FABRICATIONS

Seo, J. C.[*]; Chung, Y. H.[**] & Park, S. C.[**,#]

[*] Samsung Electronics Co., Ltd, San #16 Banwol-dong, Hwasung, 445-701, Korea
[**] Department of Industrial Engineering, Ajou University,
San 5, Woncheon-dong, Yeongtong-gu, Suwon, Korea
E-Mail: jungchul.suh@samsung.com, yongho1230@gmail.com, scpark@ajou.ac.kr
([#] Corresponding author)

Abstract

This paper proposes a new dispatching rule to achieve the on-time delivery of high priority lots in order-driven fabrications (FABs). Most of conventional dispatching rules can be considered as the variants of classical rules like ODD, EDD, and CR. Although, many of conventional dispatching rules give good performance in terms of the on-time delivery, they do not consider the existence of high priority lots. We classify orders into two types for an order-driven FAB; regular orders and high priority orders. While regular orders are typically characterized by longer cycle times, looser target due dates but lower margins, the high priority orders have shorter cycle times, tighter target due dates and higher margins. If the deliveries of high priority orders are late, the manufacturer may have to pay a significant amount of penalty charges. The proposed dispatching rule employs the concept of reservations of high priority lots, and consists of two major steps; 1) finding a high priority lot for reservation, and 2) finding a tool for reservation. The first step tries to minimize the waiting time of high priority lots, and the second step tries to maximize the utilization of tools. Experimental results show that the proposed dispatching rule is superior over conventional rules with regard to on-time delivery of high priority lots.
(Received in September 2014, accepted in April 2015. This paper was with the authors 1 month for 1 revision.)

Key Words: Dispatching Rules, Factory Simulation, FAB, Job Shop Scheduling, On-Time Delivery

1. INTRODUCTION

The production of semiconductor chips requires a multiple-step sequence of photolithographic and chemical processing steps, during which electrical circuits are gradually created on a wafer. The entire manufacturing process of chips, from the beginning to the packaging of chips for shipment, takes a couple of months, and is performed in fabrications (FABs) which are highly capital-intensive production systems. Although, small companies can't afford such expensive FABs, they need to design their own chips because almost all electronic products require a number of semiconductor chips. Those small companies want a semiconductor manufacturer to produce their products according to their designs. Because of the reason, order-driven FABs have to produce a large number of product types and variants concurrently [1-5].

To be successful in competitive market environments, semiconductor manufacturers are struggling to satisfy customers' needs in terms of quality, quantity, and due dates. Among the needs, the achievement of better due dates is extremely important for order-driven FABs. The effective scheduling of order-driven FABs have been considered as one of the most complicated problems because of hundreds of steps for a product, re-entrant flows, and batch processing [6, 7]. A FAB can be seen as a complex job shop, and the job shop-scheduling problem is often considered as a sequencing problem to determine the processing order of operations on the machines [8].

In the literature the previous approaches on the job-shop scheduling can be classified into three groups; optimization algorithms such as mathematical programming, search-based meta-

heuristics such as genetic algorithms, and dispatching rules. Among the three groups, the dispatching rules are widely accepted in the industry because of the ease of implementation, quick in reacting to the changes encountered on the shop floor, low computation requirement, and flexibility to incorporate domain knowledge and expertise [9-11]. A dispatching rule dynamically determines a WIP (Work-In-Process) lot to be processed next once a machine becomes available. Currently, most of the dispatching rules are variants of classical rules like Operation due date (ODD), Earliest due date (EDD), and Critical Ratio (CR), and these rules is often used in an attempt to optimize the on-time delivery [12-19]. Although there have been various dispatching rules for on-time delivery, still FABs have great difficulties to achieve the on-time delivery.

For an order-driven FAB, there can be two different types of orders; regular orders and high priority orders. If customers need chips with urgency, they can issue high priority orders to semiconductor manufacturers with strict due dates which are tighter than those of regular orders. While regular orders are typically characterized by longer cycle times, looser target due dates but lower margins, the high priority orders have shorter cycle times, tighter target due dates and higher margins [20, 21]. If the deliveries of high priority orders are late, the manufacturer may have to pay a significant amount of penalty charges. Since high priority orders are more critical than regular orders, it is necessary to carefully control the due dates of high priority orders with minimum side effect to regular orders. Although, this issue has been well known in industry, dispatching rules considering the existence of high priority orders have rarely been brought into the focus [22].

Because the penalty charge of high priority orders is much higher than that of regular orders, it is necessary to ensure tool capacity of high priority lots. That serves as motivation for a dynamic tool allocation way to assign more capacity to high priority orders. The objective of this paper is to develop a dispatching rule to achieve better on-time delivery of high priority orders with minimum cost of due dates of regular orders. To prove the performance of the proposed dispatching rule, a simulation model is constructed based on a wafer FAB dataset, the MIMAC6 from Measurement and Improvement of Manufacturing Capacities. The simulation experiments are carried out with commercial software MOZART® developed by the VMS solutions [23]. The overall structure of the paper is as follows. Section 2 addresses the approach of this paper for the proposed dispatching rule, and Section 3 describes the details of the proposed dispatching rule. Section 4 provides the experimental design and analyses experimental results. Finally, concluding remarks are given in Section 5.

2. APPROACH TO DEAL WITH HIGH PRIORITY ORDERS

A typical FAB consists of hundreds of tool groups, and each tool group may contain multiple tools capable of assigned operations. Since a tool group contains homogeneous tools, tools belonging to the same tool group have the same capability. As shown in Fig. 1, each tool group has its own queue, a place for WIP lots to wait when they can't move on because all tools of the tool group are busy. A FAB produces various products concurrently, and each product has its own route according to its process plan. Whenever a tool becomes available, the tool group needs to determine a lot to be processed next among waiting lots in the queue of the corresponding tool group.

If we only focus on the on-time delivery of high priority orders without considering any side effect on the due dates of regular lots, we may think of a simple rule, which forces high priority lots to be processed first. For convenience, let's refer this simple rule as 'S-rule'. As shown in Fig. 2, the S-rule chooses a high priority lot to be processed next among waiting lots in the queue. If there are multiple high priority lots, it may apply the ODD rule to pick one from the multiple high priority lots. In the case of no high priority lots, the S-rule makes a

choice among regular lots based on the ODD rule. Although we can expect better on-time delivery of high priority lots, the S-rule may cause significant increase the tardiness of regular lots.
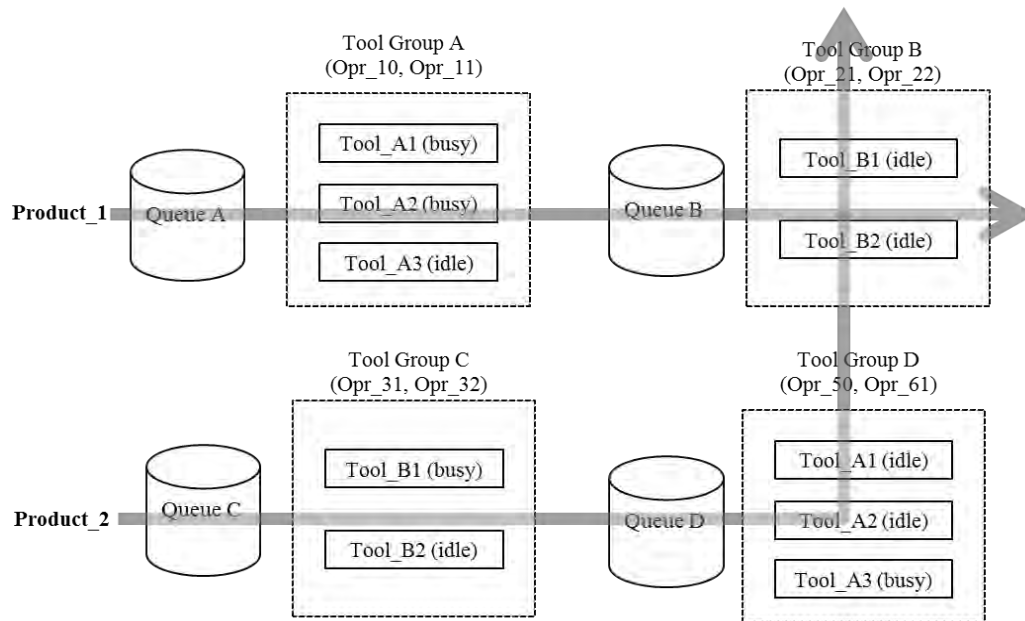

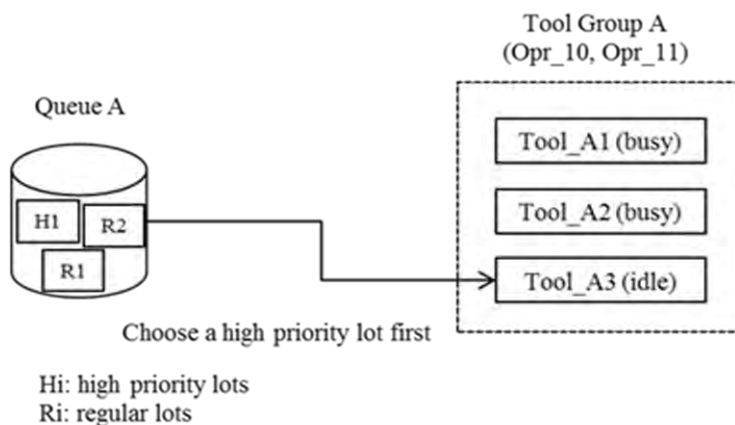Figure 1: Tool groups and homogeneous tools.

Figure 2: Simple dispatching rule: S-rule.

Our approach is to achieve the on-time delivery of high priority lots with the minimum tardiness of regular lots. To do so, we employ the concept of reservation, which means the provisional allocation of tool capacity.

The role of a dispatching rule is to determine a lot to be processed next when a tool becomes available, and most of conventional dispatching rules (ODD, EDD, CR, and S-rule) pick one among WIP lots which are waiting in the corresponding queue. Since conventional dispatching rules only consider lots in the queue, they can't support the reservation of high priority lots which are not in the queue yet. To overcome the limitations, the proposed dispatching rule takes into account high priority lots being processed in previous tools as well as the waiting lots in the queue (Fig. 3). As a result, it is possible to support the reservation of high priority lots, which means the provisional allocation of capacity to meet the on-time delivery of high priority lots. Similar with the S-rule, the proposed rule considers the due date high priority lots first, regular lots might be tardy. To minimize the tardiness of regular lots, the proposed dispatching rule takes into account tool utilization as well as on-time delivery of high priority order (HPO) lots.
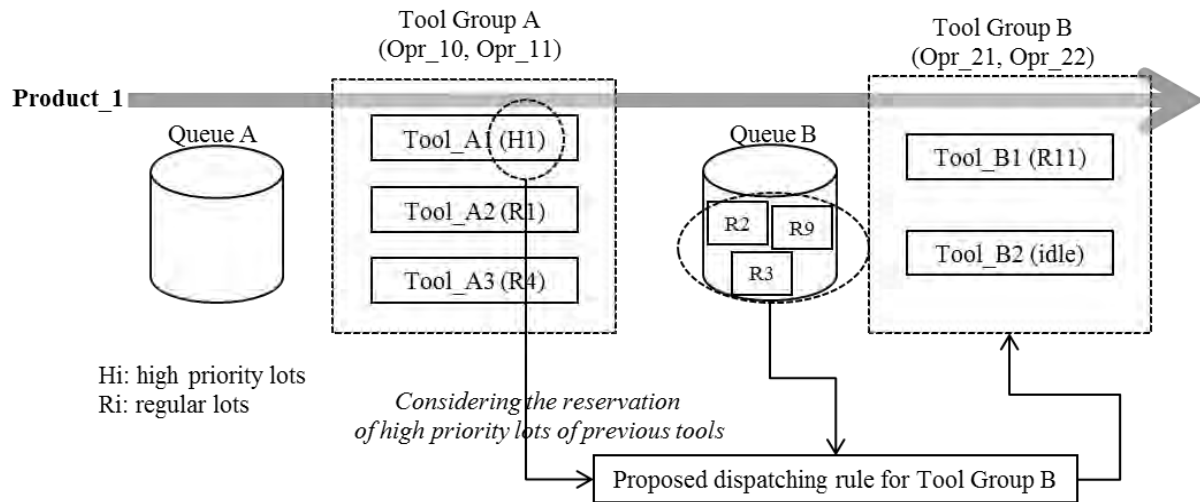
Figure 3: Proposed dispatching rule allowing the reservations of high priority lots.

3. DISPATCHING RULE WITH RESERVATIONS

This section describes the details of the proposed dispatching rule. As mentioned earlier, we consider not only WIP lots of the corresponding queue but also high priority lots of previous tool groups. A tool may reserve only one high priority lot at a time, and it cannot reserve any other high priority lots until it finishes the processing of the reserved high priority lot. Because of the reason, some high priority lots may wait in the queue without reservations together with regular lots. In this case, an available tool simply picks one by applying the ODD rule to the unreserved high priority lots. Otherwise, it is necessary to search proper a candidate for reservation among high priority lots of previous tool groups. For the formal explanation of the proposed reservation policy, we define several terms as follows.

- T_A: a tool which becomes available at time t_1 (current time).
- TG_K: a tool group including T_A.
- L_R: a lot for reservation.
- T_R: a tool for reservation.

   For the performance of the proposed dispatching rule, it is very important to develop a reasonable reservation policy. Let's assume that a tool group (TG_K) consists of n tools (T_1, T_2, .. T_n), and a tool (T_A, $1 \leq A \leq n$) becomes available at time t_1. The proposed dispatching rule consists of two major steps; 1) finding a high priority lot (L_R) for reservation, and 2) finding a tool (T_R) for reservation. The first step tries to minimize the waiting time of high priority lots, and the second step tries to maximize the utilization of tools. At this time, it is necessary to observe that T_A does not have to be the same with T_R. Since a tool group consists of homogeneous tools, we can freely choose a tool for reservation by considering the utilization.

   Whenever a tool (T_A) becomes free at time t_1, the first step of the dispatching rule searches a proper candidate for reservation among high priority lots of previous tool groups. If there are multiple high priority lots, it chooses the first arriving high priority lot (H-lot) by considering the remaining processing time and the transfer time. Let's assume that the arrival time of H-lot is t_H ($t_1 < t_H$). At this time, it is necessary to consider regular lots waiting in the queue of the tool group (TG_K). If there exist a regular lot, which can be finished before t_H, then we do not reserve H-lot to prevent the capacity loss. Otherwise, H_lot becomes L_R. The algorithm to determine a high priority lot for reservation can be described as follows.

- Finding a high priority lot for reservation (finding_LR)
   // Input: TG_K, T_A, t_1.

// Output: a lot for reservation or for an immediate processing.

Step 1) H-lot = Find the first arriving high priority lot to TG_K among high priority lots of the previous tool groups;

Step 2) If (H-lot ! = NULL)

$t_H = t_1 +$ Remaining processing time of H-lot + Transfer time to TG_K;

Else

$t_H = \infty$;

Step 3) R-lot = Apply ODD rule for WIP lots in the queue of TG_K;

Step 4) $t_R = t_1 +$ Remaining processing time of R-lot;

Step 5) If ($t_H < t_R$)

Return H-lot ; // for reservation ; (L_R = H-lot)

Else

Return R-lot ; // for an immediate processing of R-lot at T_A

Once a L_R is determined at the first step, it is necessary to find a tool (T_R) for reservation. As mentioned earlier, the tool group TG_K consists of n homogenous tools (T_1, T_2, .. T_n). Considering the existence homogeneous tools, we may think of a better assignment policy of reservations. In other words, we may try to find the best tool for the reservation in the tool group, which can minimize the capacity loss (idle time). Fig. 4 shows a tool group consisting of three homogeneous tools (T_1, T_2, and T_3). T_1 becomes free at t_1, and finds a reservation candidate (H-lot) according to the reservation algorithm. Since the arrival time of the H-lot is t_H, T_1 has to wait for the time period of $t_H - t_1$. If we assign the H-lot to T_2, we can save the idle time because T_2 becomes free at t_2 ($t_H - t_2 < t_H - t_1$). On the contrary, it is not desirable to assign the H-lot to T_3, because T_3 becomes free after the arrival of the H-lot.
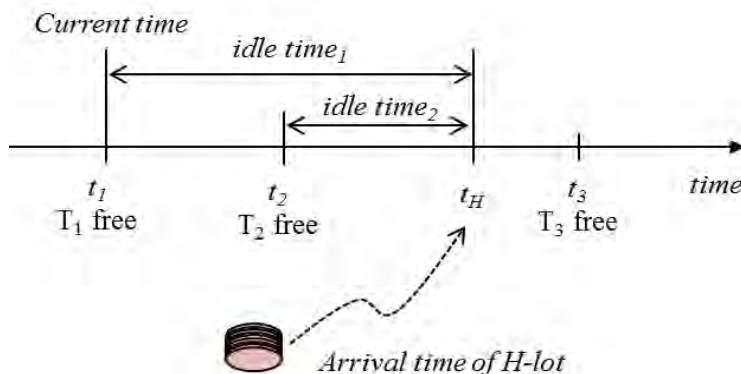

Figure 4: Simple dispatching rule: S-rule.

In this way, we can achieve better on-time delivery of HPO lots with the minimum capacity loss. The algorithm to determine a tool for reservation can be described as follows.

- Finding a proper tool for reservation (finding_TR)

// Input: TG_K, T_A, t_1, H-lot.

// Output: T_R.

Step 1) $t_H = t_1 +$ Remaining processing time of H-lot + Transfer time to TG_K;

Step 2) $T_R = T_A$;

Step 3) $idle\_time = t_H - t_1$;

Step 4) For ($i = 1$; $i <= TG_K.ToolsCount$; i++) {

$t_i =$ the time of a tool T_i becomes available;

if ($t_H - t_i < idle\_time$ && $t_H - t_i >= 0$) {

$idle\_time = t_H - t_i$;

$T_R = T_i$;

}

}

Step 5) Return T_R;

Based on the two major steps, the overall dispatching rule can be described as follows. It works whenever a tool (T_A) becomes free at time t_1.

- Dispatching rule with reservations
  // Input: TG_K, T_A, t_1.
  Step 1) L_R = finding_L_R (TG_K, T_A, t_1);
  Step 2) If (L_R is a regular lot) {
              Start an immediate processing of L_R at T_A;
              Return;
        }
  Step 3) T_R = finding_T_R (TG_K, T_A, t_1, L_R);
  Step 4) T_R reserves L_R ; // T_R has to wait until the finish of L_R

4. PERFORMANCE ANALYSIS

For the performance analysis of the proposed dispatching rule, we construct a simulation model by using a small wafer FAB dataset MIMAC6 from Measurement and Improvement of Manufacturing Capacities (MIMAC). It is necessary to refer to the MIMAC Final Report (1995) for the explanation details [24]. The configuration of the simulation model is summarized in Table I.

Table I: Configuration of simulation model used in this study.

Modelling	Value
Number of products (processes)	9
Number of tool groups	104
Number of tools	230
Wafers in a lot	24
Lots released per year	2706
Number of tools per tool group	2-7
Rework modelled	No
Yield loss (scrap) modelled	No
Raw processing time range (hours)	211-334
Number of processing steps range	234-355
Total number of processing steps	2541
Batching policy	Minimum batch size
Sequence dependent setup	Yes

To conduct the simulation, the MOZART® engine developed by VMS solutions was used. For comparison, the simulation experiments have been performed for three different dispatching rules; 1) the classic ODD rule, 2) the S-rule mentioned in Section 2, and 3) the proposed dispatching rule. In the case of ODD, we compute the operational due date in the following way: ODD = Due Date – Remaining Processing Time (RPT) × Flow Factor, where Flow Factor (FF) = target cycle time / RPT. As mentioned earlier, the S-rule simply chooses a high priority among waiting lots in the queue without the concept of reservations.
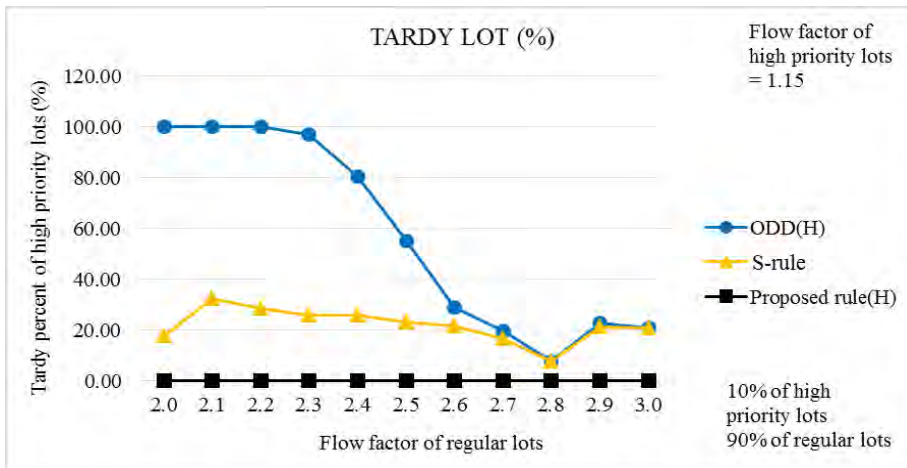
Figure 5: Percent of tardy high priority lots (10 % of high priority lots).
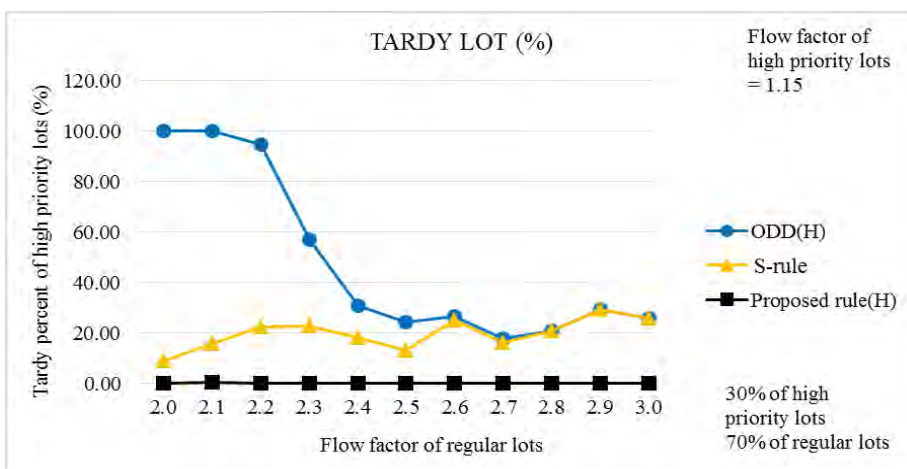

Figure 6: Percent of tardy high priority lots (30 % of high priority lots).

While regular orders are typically characterized by longer cycle times, looser target due dates but lower margins, the high priority orders have shorter cycle times, tighter target due dates and higher margins. Because of the reason, we give 1.15 as the flow factor of high priority lots, which means a very tight target due date. For the fixed flow factor of high priority lots, multiple experimentations have been performed for different flow factors of regular lots (2.0 – 3.0).

Figure 7: Percent of tardy regular lots (10 % of high priority lots).
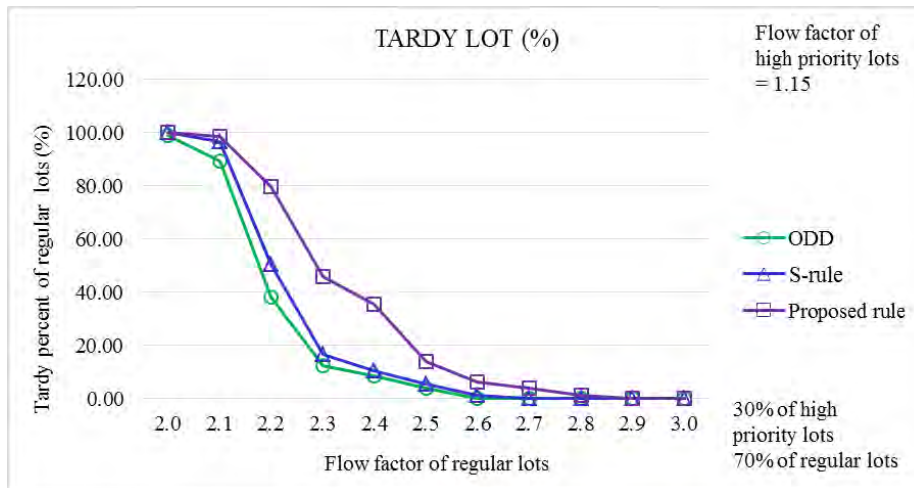
Figure 8: Percent of tardy regular lots (30 % of high priority lots).

To analyse the effect of the ratio of high priority lots, we perform the experimentations for two different configurations; 1) 10 % of high priority lots and 90 % of regular lots, and 2) 30 % of high priority lots and 70 % of regular lots. Figs. 5 and 6 show the percent of tardy high priority lots for 10 % of high priority lots and 30 % of high priority lots respectively. Since the ODD rule does not care about the high priority lots, it shows the highest percent of tardy high priority lots, which may cause significant amount of penalty charges. Although the S-rule gives better on-time delivery of high priority lots compared to the ODD rule, it still cause about 20 % of tardy high priority lots. In the cause of the proposed rule, it shows the perfect achievement of the on-time delivery of high priority lots for both configurations.

Figs. 7 and 8 show the percent of tardy regular lots for 10 % of high priority lots and 30 % of high priority lots respectively. Although the proposed rule gives worse performance for regular lots compared to two other rules, the difference is not significant. This is because the proposed rule achieves the on-time delivery of high priority lots at the expense of the tardy regular lots. If the deliveries of high priority orders are late, the manufacturer may have to pay a significant amount of penalty charges. Since high priority orders are more critical than regular orders, it is necessary to carefully control the due dates of high priority orders with minimum side effect to regular orders.

5. CONCLUSION

A wafer FAB is a complicated manufacturing system, and the FAB scheduling can be considered as a dispatching problem to determine a lot to be processed next. Conventional dispatching rules are trying to achieve the better on-time delivery of lots, and they are mostly the variants of classical rules like ODD, EDD, and CR. Although, many of conventional dispatching rules give good performance in terms of the on-time delivery, they do not consider the existence of high priority lots.

For an order-driven FAB, there can be two different types of orders; regular orders and high priority orders. While regular orders are typically characterized by longer cycle times, looser target due dates but lower margins, the high priority orders have shorter cycle times, tighter target due dates and higher margins. To ensure the on-time delivery of high priority lots, we propose a new dispatching rule with the concept of reservation, the provisional allocation of capacity. The proposed dispatching rule consists of two major steps; 1) finding a high priority lot for reservation, and 2) finding a tool for reservation. The first step tries to minimize the waiting time of high priority lots, and the second step tries to maximize the utilization of tools.

The simulation model has been constructed by using MIMAC6, and the experimentation has been performed by employing a commercial software system MOZART® developed by VMS solutions. The simulation results show that the proposed dispatching rule is superior over the ODD rule and the S-rule with regard to on-time delivery of high priority lots.

ACKNOWLEDGEMENTS

REFERENCES

[1] Gracanin, D.; Lalic, B.; Beker, I.; Lalic, D.; Buchmeister, B. (2013). Cost-time profile simulation for job shop scheduling decisions, *International Journal of Simulation Modelling*, Vol. 12, No. 4, 213-224, doi:10.2507/IJSIMM12(4)1.237

[2] Yu, D. H.; Huang, M. (2011). The batch machine scheduling of continuous caster with flexible jobs to minimize setup costs, *Concurrent Engineering: Research and Applications*, Vol. 19, No. 4, 325-334, doi:10.1177/1063293X11425192

[3] Kang, K. H.; Lee, Y. H. (2007). Make-to-order scheduling in foundry semiconductor fabrication, *International Journal of Production Research*, Vol. 45, No. 3, 615-630, doi:10.1080/00207540600792416

[4] Wu, M.-C.; Chiou, S.-J.; Chen, C.-F. (2008). Dispatching for make-to-order wafer fabs with machine-dedication and mask set-up characteristics, *International Journal of Production Research*, Vol. 46, No. 14, 3993-4009, doi:10.1080/00207540601085919

[5] Tran, T. A.; Park, J. Y. (2014). Development of integrated design methodology for various types of product-service systems, *Journal of Computational Design and Engineering*, Vol. 1, No. 1, 37-47, doi:10.7315/JCDE.2014.004

[6] Dang, T. T. (2013). Combination of dispatching rules and prediction for solving multi-objective scheduling problems, *International Journal of Production Research*, Vol. 51, No. 17, 5180-5194, doi:10.1080/00207543.2013.793857

[7] Park, S. C.; Ahn, E.; Chung, Y.; Yang, K. R.; Kim, B. H.; Seo, J. C. (2013). Fab simulation with recipe arrangement of tools, *Proceedings of the 2013 Winter Simulation Conference*, 3840-3849, doi:10.1109/WSC.2013.6721743

[8] Singh, A. (2013). Minimization of mean tardiness in a flexible job shop, *International Journal of Simulation Modelling*, Vol. 12, No. 3, 190-204, doi:10.2507/IJSIMM12(3)5.242

[9] Appleton-Day, K.; Shao, L. (1997). Real-time dispatch gets real-time results in AMD's Fab25, *Proceedings of the IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop*, 444-447

[10] Giegandt, A.; Nicholson, G. (1998). Better dispatch application – a success story, *Proceedings of the IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop*, 396-399

[11] Kaban, A. K.; Othman, Z.; Rohmah, D. S. (2012). Comparison of dispatching rules in job-shop scheduling problem using simulation: A case study, *International Journal of Simulation Modelling*, Vol. 11, No. 3, 129-140, doi:10.2507/IJSIMM11(3)2.201

[12] Bahaji, N.; Kuhl, M. E. (2008). A simulation study of new multi-objective composite dispatching rules, CONWIP, and push lot release in semiconductor fabrication, *International Journal of Production Research*, Vol. 46, No. 14, 3801-3824, doi:10.1080/00207540600711879

[13] Chiang, T.-C.; Fu, L.-C. (2004). Solving the FMS scheduling problem by critical ratio-based heuristics and the genetic algorithm, *Proceedings of the 2004 IEEE International Conference on Robotics and Automation*, Vol. 3, 3131-3136, doi:10.1109/ROBOT.2004.1307544

[14] Dileepan, P.; Ahmadi, M. (2010). Scheduling rules for a small dynamic job-shop: A simulation approach, *International Journal of Simulation Modelling*, Vol. 9, No. 4, 173-183, doi:10.2507/IJSIMM09(4)1.165

[15] Jayamohan, M. S.; Rajendran, C. (2000). New dispatching rules for shop scheduling: A step forward, *International Journal of Production Research*, Vol. 38, No. 3, 563-586, doi:10.1080/002075400189301

[16] Lodree Jr, E.; Jang, W.; Klein, C. M. (2004). A new rule for minimizing the number of tardy jobs in dynamic flow shops, *European Journal of Operational Research*, Vol. 159, No. 1, 258-263, doi:10.1016/S0377-2217(03)00404-1

[17] Rose, O. (2002). Some issues of the critical ratio dispatch rule in semiconductor manufacturing, *Proceedings of the 2002 Winter Simulation Conference*, Vol. 2, 1401-1405, doi:10.1109/wsc.2002.1166410

[18] Zhou, Z.; Rose, O. (2012). WIP balance and due date control in a wafer fab with low and high volume products, *Proceedings of the 2012 Winter Simulation Conference*, 2019-2026, doi:10.1109/wsc.2012.6464989

[19] Jovanovic, J. R.; Milanovic, D. D.; Djukic, R. D. (2014). Manufacturing cycle time analysis and scheduling to optimize its duration, *Strojniski vestnik – Journal of Mechanical Engineering*, Vol. 60, No. 7-8, 512-524, doi:10.5545/sv-jme.2013.1523

[20] Ehteshami, B.; Petrakian, R. G.; Shabe, P. M. (1992). Trade-offs in cycle time management: hot lots, *IEEE Transactions on Semiconductor Manufacturing*, Vol. 5, No. 2, 101-106, doi:10.1109/66.136270

[21] Trybula, W. J. (1993). Hot jobs, bane or boon, *Proceedings of the 15th International Electronic Manufacturing Technology Symposium*, 317-322, doi:10.1109/IEMT.1993.398186

[22] Crist, K.; Uzsoy, R. (2011). Prioritising production and engineering lots in wafer fabrication facilities: A simulation study, *International Journal of Production Research*, Vol. 49, No. 11, 3105-3125, doi:10.1080/00207541003801259

[23] Ko, K.; Kim, B. H.; Yoo, S. K. (2013). Simulation based planning & scheduling system: MozArt®, *Proceedings of the 2013 Winter Simulation Conference*, 4103-4104

[24] Fowler, J.; Robinson, J. (1995). *Measurement and improvement of manufacturing capacities (MIMAC): Final report*, Technical Report 95062861A-TR, SEMATECH, Austin