

Analysis of Lot Sizing and Planning Bucket Strategies for Semiconductor Backend Operation

Malcolm Yoke Hean Low
Mike Long Foong Liow
Peter Lendermann
Singapore Institute of Manufacturing Technology
71 Nanyang Drive, Singapore 638075
{yhlow,lfliow,peterl}@simtech.a-star.edu.sg

Jurgen Potoradi, Long Kim Tng,
Wai Khang Lim, Daniel Quadt
Infineon Technology Asia Pacific Pte Ltd
168 Kallang Way, Singapore 349253
{juergen.potoradi, long-kim.tng, wai-khang.lim,
daniel.quadt}@infineon.com

Abstract

Capacity planning in existing semiconductor assembly and test operation is usually done in weekly buckets, and incoming customer orders are checked “capable-to-promise” against these weekly buckets. Consequently, delivery of a customer order on a specific date can be ensured only if they are planned for in the previous planning week. The overall leadtime from customer order processing to final delivery can amount to more than six weeks. Significant leadtime reduction can be made through moving from weekly to daily planning buckets. At the same time, in the hyper-complex, make-to-order environment of the assembly and test manufacturing process, where order quantities are not necessarily a multiple of the optimal lot sizes due to customer order constraints, the issue of lot sizing has become even more critical than before. In this paper, we investigate the issue of lot sizing and the impact of planning bucket on the order fulfillment rate of the backend operation. A simulation model based on actual factory data is constructed and experiments were carried out to determine the best lot size to use for the factory as well as to evaluate the impact of switching from a weekly to a daily planning bucket.

1. Introduction

In semiconductor backend assembly and test operation, products can be made and supplied to customers from storage location such as wafer stock, die bank, and distribution centre (DC), depending on volume and order frequency. Capacity planning in assembly and test operation is usually carried out in weekly buckets, and incoming customer orders are checked “capable-to-promise” against these weekly buckets. Consequently, delivery of a customer’s order on a specific date can be ensured only if they are planned for in the previous planning week. The overall leadtime from customer order processing to final delivery can amount to more than six weeks. This includes the processing time for customer orders, the order fulfillment leadtime (OFLT), the time taken for shipment to DC, and from the DC to customer.

To sharpen the competitive edge, further reduction of this overall leadtime is essential. An important contribution to such leadtime reduction can be made through moving from weekly to daily planning buckets. At the same time, in the hyper-complex, make-to-order environment of the assembly and test manufacturing process, where order quantities are not necessarily a multiple of the optimal lot sizes due to customer order constraints, the issue of lot sizing has become even more critical than before. Additionally, high variability on the shopfloor also has significant impact on the schedule adherence. In this setting, the establishment of an appropriate balance between the variability of the production lines and the flexibility/sophistication of the scheduling process is critical.

In this paper, we investigate the issue of lot sizing and the impact of planning bucket on the order fulfillment rate of the backend operation. A simulation model based on actual factory data is developed and experiments were carried out to determine the best lot size to use for the factory as well as to evaluate the impact of switching from a weekly to a daily planning bucket.

2. Related Work

The semiconductor industry is highly capital intensive and the products it produces are also of high value [7]. While significant productivity improvement (in terms of equipment utilization, inventory and lead-times) has been achieved in the semiconductor front-end process, the same improvement has not been readily replicated in the back-end assembly and test operation. The semiconductor backend assembly and test process remains the bottleneck in the overall industry goal of bringing higher quality products to market quicker and at a lower cost [2]. In order to operate effectively in an increasingly competitive environment, semiconductor assembly and test manufacturers are constantly looking for ways to improve capital efficiency and order lead-time by increasing throughput and reducing cycle-time. Increasingly, these companies are moving toward time-based competition. A clear understanding on the impact of lead-time on expected profit, and the inter-relationship between manufacturing cycle-time, lead-time reliability (earliness and tardiness), throughput, inventory levels and resource utilization is of significant interest to these manufacturers [17]. Improvement on order lead-time allows a company to establish competitive advantage over its competitors [16]. From the competitor perspective, the manufacturer is able to respond to any changes in market condition in a timely fashion. From the customer perspective, the order can be delivered to the customer in a speedier manner.

Lots scheduling is an important aspect of backend assembly operation as it ensures orders are processed in a timely manner. In most of the semiconductor backend assembly shopfloor, simple heuristics in the form of priority dispatching rules such as First-In-First-Out (FIFO) and Shortest-Processing-Time (SPT) are being used to schedule lots onto machines. More detailed shop floor management will involve the use of finite capacity scheduling (FCS) which allocates resources for manufacturing operation over the time horizon. For example, a rule-based heuristics FCS system for semiconductor backend assembly is described in [18] where large-scale scheduling problems are solved effectively in a complex industrial environment.

Simulation model analysis of existing operation can be used to discover and validate opportunities for cycle-time reduction without incurring additional cost to the factory. In [5], the authors presented simulation methodology and findings pertaining to the analysis of the Assembly, Burn-in, and Test operations for a backend assembly and test facility. A list of recommendations was proposed to reduce cycle-time in the factory. Using a validated model of the backend factory, the authors showed that the cumulative impact of these recommendations is a 41% reduction in the average cycle-time and many of the recommendations identified can be implemented at no additional cost to the factory.

Lot-splitting is a technique used in a make-to-order environment to break large orders into smaller transfer lots and offers the ability to move parts more quickly through the production to improve make-span performance [14]. An order can be split equally or unequally, with the number of splits ranging from one to the number of units in the order. Time bucket is another technique used by planning managers which can constrain the re-planning frequency of the planning system and affects the length of internal work-order lead-times. A study in [12] shows that the length of time bucket has an important effect on the performance of equal split strategy.

Product mix is an important production decision in semiconductor manufacturing. Suitable time bucket of planning, be it one day, one week or one month, must be used for workload and product mix calculation [3]. The high production mix typically found in an assembly and test operation often leads to wide cycle-time distribution among production orders. In [13], the author carried out a simulation based cause and effect analysis of cycle-time distribution in semiconductor backend using a validated model of an actual semiconductor backend facility. The study concluded that lot release scheduling to the first operation has the greatest impact on cycle-time distribution and throughput in semiconductor backend manufacturing. The author also noted that theoretical cycle-time distribution and lot size distribution have complex effects on actual cycle-time distribution on the production shopfloor.

Lot sizing has a significant impact on the production system cycle-time [15]. Determining an optimal production lot size for the backend factory is not a straightforward exercise since the optimal lot size in the assembly is generally different from the one in the test. In [11], a study was carried out for the analysis of lot sizing for two models: one for the combined assembly/end-of-line areas and one for the burn-in/test areas. The study showed that the operating curves produced for different lot sizes for the two models can be used to determine the optimum lot size that maximizes throughput while maintaining the established cycle-time target for both the assembly area and burn-in/test area.

Lot sizing can also directly affect inventory and scheduling. Heuristic can be employed for simultaneous lot sizing and scheduling to minimize cycle-time as well as set-up and inventory holding costs [1]. A detailed survey of lot sizing and scheduling is available in [6]. In [8], a hierarchical lot sizing and scheduling approach for the semiconductor assembly process is described. The lot sizing phase is bottleneck-oriented and integrates back-ordering, setup carry-over and parallel machines. The scheduling phase first generates a schedule based on the product family, and then uses this schedule to arrange individual jobs of a customer order within the time slots for each product family.

3. Simulation Model

In order to study different lot sizing strategies and the effects of switching from a weekly to a daily planning bucket, a virtual representation of an existing backend manufacturing operations which portrays the shopfloor including selected variability (random) effects is developed using the WITNESS commercial off-the-shelf simulation package version 2003 [10]. Figure 1 shows the line configuration of the model developed.

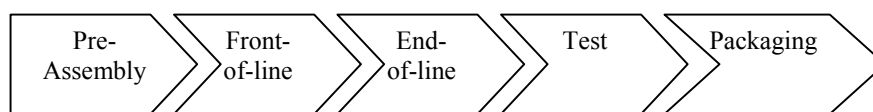


Figure 1: Line Configuration

The simulation model is triggered by release of material to the first production step, wafer mount, according to a simplified release sequence. From there the material flows through the rest of the Assembly process according to dispatching rules (e.g. first-come-first-serve or dedicated line) that are applied on the shopfloor to handle selection of lots for processing on the individual machines. For the Test operations, a customized dispatching rule (that considers accessories availability and setup avoidance) to select lots to be processed on machines is used. A merging process takes place between the Assembly and the Test operations to emulate the merging process of multiple assembly lots into a test lot. Lots that emerge from the merging station enter an automated storage and retrieval systems (AS/RS) for temporary storage, modelled by a fixed delay in the simulation model.

Input data (product demand, processing and setup times etc) is fed into the WITNESS model from an external Excel file. In this study, two product families (product family types A and B) which consist of a total of 74 product types are included. Two types of release sequences are used in this study: 1) Uniform distributed release – the 1st week order demand plus backlog obtained from historical data is equally distributed over 7 days and released into the simulation system with type A packages first and follow by type B packages; 2) Manual line planner release – this release schedule was generated by line planner from the shopfloor. The 1st week order demand plus backlog is segregated into “small” demand order and “large” demand order. The “small” demand order are consolidated and released into the system first, whereas the “large” demand order are divided equally over the rest of the 7 days production period.

The simulation runs to be analysed start from an empty factory and go through one year of simulation with two months of warm-up period, therefore information regarding the initial shopfloor status (work-in-progress (WIP) data and machines status) is not required. The weekly demand quantities are broken down into integer multiples of the optimal lotsize for assembly plus the remaining quantity (for example: demand for 22,000 devices and optimal lotsize of 1,500 will result in 14 lots of 1,500 and 1 lot of 1,000) to be released into the factory uniformly across 1 week. The Assembly and Test processes will have different lot sizes to maximise the up-time of testers. Lots of the same product type can be merged to form larger lot sizes for Test operations, subject to the condition that lots are released from the same planning bucket (i.e. within the same production week for weekly planning bucket, and within the same day for daily planning bucket).

4. Experimental Results

4.1 Analysis of Weekly vs Daily Planning Bucket

Under a weekly planning bucket with manufacturing lead-time m , orders received in week k (day k_1 to day k_7) will be released from day in week $k+1$ (from day $k+1_1$ to day $k+1_7$, assuming all raw materials are ready by day $k+1_1$ in week $k+1$). The earliest due date that can be quoted for these orders will be on day $(k+1)_7+m$. For daily planning bucket, an order received in day k_1 can be released as early as day k_2 and can be quoted an earliest due date of k_2+m . In this case, the potential saving of 1 to 7 days on order lead-time will only hold true if the manufacturing lead-time m remains the same. To study the impact on manufacturing lead-time when moving from a weekly to a daily planning bucket, a series of experiments are conducted to first determine the lead-time required to obtain a pre-determined order fulfillment rate of 99% under a weekly planning bucket. The daily order quantity released for each product type is manually assigned by production line planner. Experimental results show that this order fulfillment rate can be achieved with a 10-day manufacturing lead-time.

We conducted another set of experiments to determine what percentage of average weekly order quantity the factory can handle when a daily planning bucket strategy is employed. For the experiments on daily planning bucket, the same weekly demand quantity is used with daily product arrival sequence following a Pareto order (i.e. order with higher demand is released earlier in a week and order with smaller demand later in a week). The total daily release quantity is capped based on the assigned quantity used in the experiments for the weekly planning bucket. The total number of production lots per week generated by the weekly and daily planning bucket is similar.

The experiments on daily planning bucket are repeated with manufacturing lead-times of 8.5, 9.25 and 10 days respectively. For each experiment, the order fulfillment rate is obtained for different percentages of weekly demand. Our first set of experiments show that under a weekly planning bucket with 100% weekly demand, a 99% order fulfillment rate can be achieved with a 10 days manufacturing lead-time. Note that under a weekly planning bucket with a 10 days manufacturing lead-time, all lots release in the week (from day 1 to day 7) have a common due-date at day 17. In the case of the daily planning bucket with 10 days manufacturing lead-time, lots that are released in different days, say x and y , have due-dates $x+10$ and $y+10$ respectively.

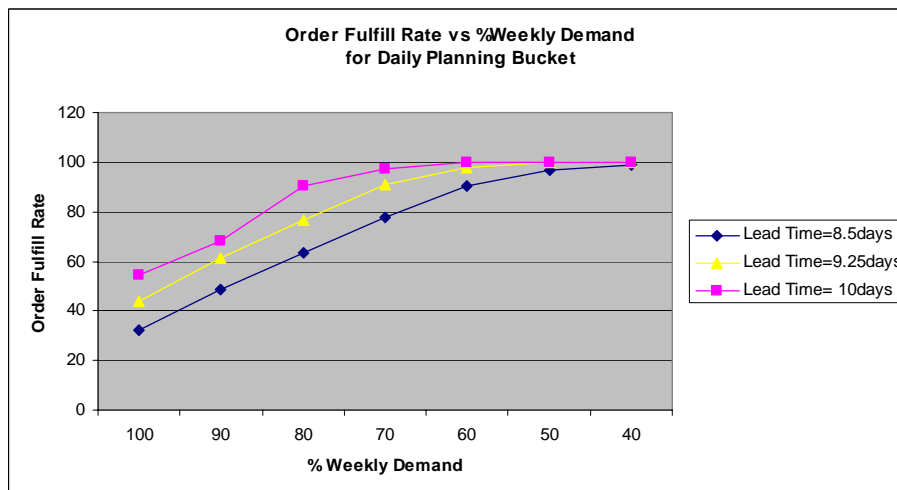


Figure 4: Order Fulfill Rate vs. Weekly Demand for Daily Planning Bucket

The experimental result from Figure 4 shows that under a daily planning bucket with 10 days manufacturing lead-time, 99% order fulfillment can only be sustained with 60% weekly demand or less. To achieve 99% order fulfillment for daily planning bucket with 10 days order lead-time for 100% weekly demand, more capacity has to be added to the production line and more sophisticated planning approach that takes into account machine setup requirements is needed.

Another set of experiments are conducted to determine the additional manufacturing lead-time required to achieve 99% order fulfillment rate at 100% weekly demand when moving from a weekly to daily planning bucket. Figure 5 shows the order fulfillment rate for different manufacturing lead-times under the daily planning bucket. In order to maintain a 99% order fulfillment rate without increasing the production capacity or additional planning effort, the manufacturing lead-time would have to be increased to 24 days (compared to a 10 days manufacturing lead-time under weekly planning bucket).

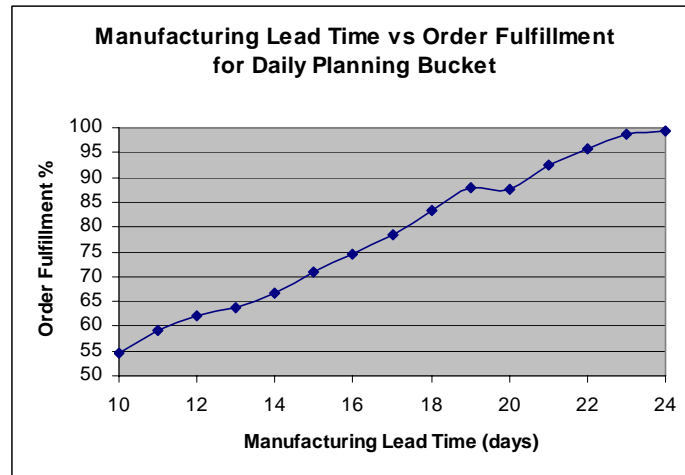


Figure 5: Manufacturing Lead-time vs. Order Fulfillment for Daily Planning Bucket

4.2 Analysis of Lot Sizing

Production lots that complete the assembly process are merged according to a Test-to-Assembly lotsize ratio to form a test lot. In order to examine the effects of lot sizing policy on factory performance indicators such as throughput and cycle-time, experiments with different Test-to-Assembly lotsize ratios are conducted. To simplify the analysis, the 74 product types are classified into three groups. Group 1 consists of those products from product family A that has small assembly lot size (≤ 1440). Group 2 consists of those products from product family A that has larger assembly lot size (≥ 4000). Group 3 consists of those products from product family B only (≥ 4800).

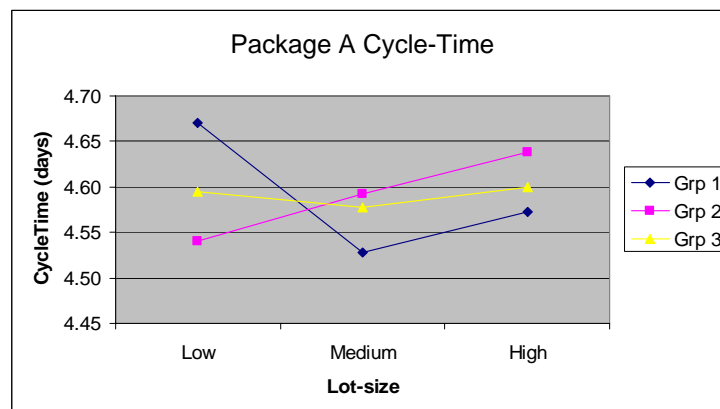


Figure 6: Package A Family Cycle-Time

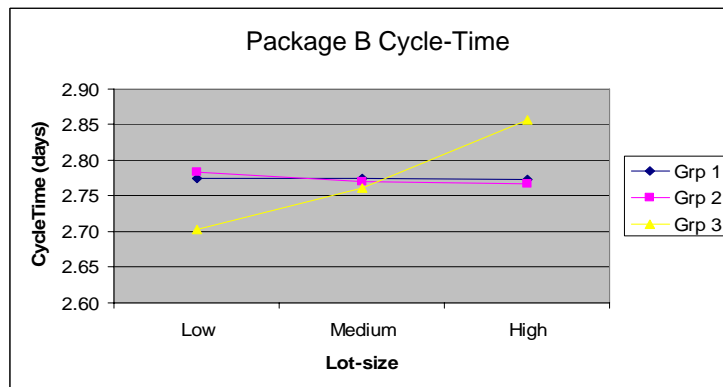


Figure 7: Package B Family Cycle-Time

Figures 6 and 7 show the results of a full factorial analysis on how the Test-to-Assembly lotsize ratio (for each of the three groups) affects the product family cycle-time. Note that each data point is an average of 9 results obtained by varying the Test-to-Assembly lot size of the other two groups. For Group 1, we use a Test-to-Assembly lotsize ratio of 2.0, 4.0, and 6.0 for the three lot size levels (low, medium and high). For Groups 2 and 3, we use a ratio of 1.0, 1.5, and 2.0. In Figure 7, Group 1 and Group 2 have no effect on the package B family cycle-time since both groups consist of only packages from the package A family. Similarly, Figure 6 shows that Group 3 has little effect on the package A's family cycle-time.

The experimental results indicated that a Test-to-Assembly lotsize ratio of 4.0 for Group 1, and 1.0 for Groups 2 and 3 gave the best product family cycle-time. To further verify that this combination of Test-to-Assembly lotsize ratio maximizes the capacity of the factory, another set of experiments was carried out by varying the capacity and Test lotsize with the aim of maintaining a fixed overall cycle-time of 8.7 days for Group 1.

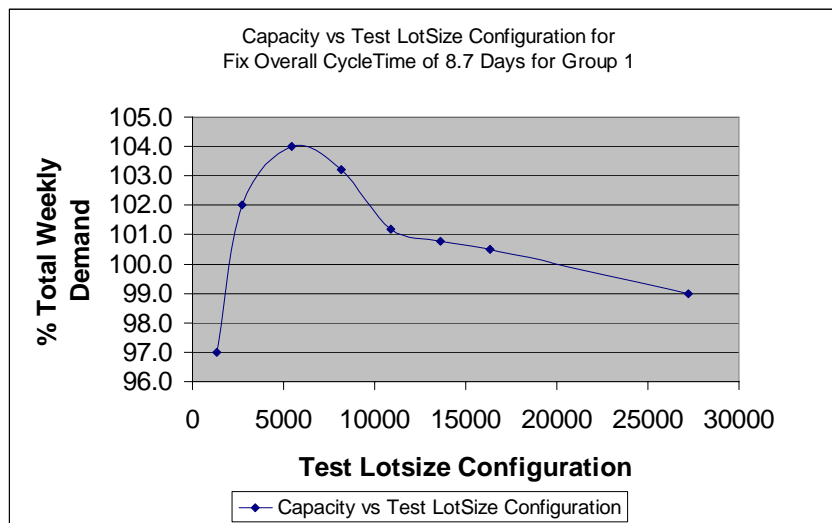


Figure 8: Capacity vs. Test Lotsize Configuration for Fixed Overall Cycle-Time of 8.7 Days for Group 1

Figure 8 shows the percentage of total weekly demand that can sustain an overall cycle-time of 8.7 days for products in group 1. Given the range of lotsize for Group 1 is 1280-1440, the experimental results show that a Test-to-Assembly lotsize ratio of 4 for Group 1 (~5000) yields the highest throughput. With a lower or higher Test-to-Assembly lotsize ratio, the factory can only sustain a lower production capacity.

5. Conclusion and Future Work

In this paper, we have described our work in developing a simulation testbed for a semiconductor backend factory and using it to analyze the impact of different lot sizing and planning bucket policies on the manufacturing shopfloor. Our experimental results show that the factory capacity has to drop to as low as 60% of the existing level in order to maintain a pre-defined order fulfillment rate of 99% when moving from a weekly to a daily planning bucket. We also carried out a lot sizing study which shows that the optimal Test-to-Assembly lot size ratio of 4 provides the maximum capacity for the factory.

For our future study, we will focus our investigation to the following areas: 1) How can policies that are used to handle events caused by variability on the shopfloor be analyzed in the simulation testbed?; 2) How can a scheduling system be integrated into the simulation in order to handle and analyze demand changes and their consequences for release of material into the production system, also taking into account the WIP status on the shopfloor?; 3) To what extent can a better enforcement of schedule adherence improve the operational performance of the manufacturing system?

Faced with rapid changing demands, many semiconductor backend manufacturers have opted to either produce stock in advance or use a subcontractor to provide additional capacity when they need it. As decision making on outsourcing can take place at strategic, tactical and operational levels within a company [4], simulation can be used as an effective tool to aid in the decision making process. We are also currently working on a prototype symbiotic simulation system that integrates simulation system and the underlying physical semiconductor backend system in a mutually beneficial way [9]. The symbiotic simulation system will be used to study different strategies and policies for outsourcing production to subcontractors based on tactical and operational requirements of the manufacturing shopfloor.

References

1. Anwar, M.F. and Nagi, R. Integrated Lot-sizing and Scheduling for Just-In-Time Production of Complex Assemblies with Finite Set-ups. *International Journal of Production Research*, 35(5), pp. 1447-1470, 1997.
2. Bibbee, S., "Fixing the Weakest Link: Immediate Strategies for Test Process Transformation". *Microwave Product Digest*, Octagon Communications, August Issue. 2004
3. Chou, Y.-C. and Hong, I.-H. A Methodology for Product Mix Planning in Semiconductor Foundry Manufacturing. *IEEE Transaction on Semiconductor Manufacturing*, 13(3), pp. 278-285, 2000.
4. Dekkers, R. Strategic Capacity Management: Meeting Technological Demands and Performance Criteria. *Journal of Materials Processing Technology*, 139(2003), pp. 385-393, 2003.
5. Domaschke, J., Brown, S., Robinson, J., and Leibl, F. Effective Implementation of Cycle Time Reduction Strategies for Semiconductor Backend Manufacturing. In *Proceedings of the 1998 Winter Simulation Conference*, pp. 985-992, 1998.
6. Drexl, A. and Kimms, A. Lot Sizing and Scheduling - Survey and Extensions. *European Journal of Operational Research*, 99, pp. 221-225, 1997.
7. Jain, S., Lim, C.C., Gan, B.P., and Low, Y.H. Criticality of detailed modeling in semiconductor supply chain simulation. In *Proceedings of the 1999 Winter Simulation Conference*, pp. 888-896, 1999.
8. Kuhn, H. and Quadt, D. Lot sizing and scheduling in semiconductor assembly – a hierarchical planning approach. In *Proceedings of the International Conference on Modeling and Analysis of Semiconductor Manufacturing (MASM 2002)*, pp. 211–216, 2003.
9. Low, M.Y.H., Lye, K.W., Lendermann, P., Turner, S.J., Chim, R.T.W., and Leo, S.H. An Agent-based Approach for Managing Symbiotic Simulation of Semiconductor Assembly and Test Operation. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multi Agent Systems (to appear)*, pp. 2005.
10. Markt, P.L. and Mayer, M.H. Witness Simulation Software A Flexible Suite of Simulation Tools. In *Proceedings of the 1997 Winter Simulation Conference*, pp. 711-717, 1997.
11. Potoradi, J. and Lee, W.K. Determining Optimal Lot size for a Semiconductor Back-end Factory. In *Proceedings of the 1999 Winter Simulation Conference*, pp. 720-726, 1999.
12. Riezebos, J. Time Bucket Length and Lot-splitting Approach. *International Journal of Production Research*, 42(12), pp. 2325-2338, 2004.
13. Sivakumar, A.I. Simulation Based Cause and Effect Analysis of Cycle Time Distribution in Semiconductor Backend. In *Proceedings of the 2000 Winter Simulation Conference*, pp. 1464-1471, 2000.

14. Smunt, T., Buss, A., and D. Kropp. Lot-Splitting in Stochastic Flow Shop and Job Shop Environments. *Decision Sciences*, 27(2), pp. 215-237, 1996.
15. Spearman, M.L. Lot Sizing to Minimize Cycle Time in a Sequential Operation, Factory Physics, Inc. 2004.
16. Tersine, R.J. and Hummingbird, E.A. Lead-time Reduction: The Search for Competitive Advantage. *International Journal of Operations & Production Management*, 15(2), pp. 8-18, 1995.
17. Weng, Z.K. Strategies for Integrating Lead Time and Customer-Order Decisions. *IIE Transactions*, 31(1999), pp. 161-171, 1999.
18. Xiao-Feng, Y., et al. A Rule-based Heuristic Finite Capacity Scheduling System for Semiconductor Backend Assembly. *International Journal of Computer Integrated Manufacturing*, 17(8), pp. 733-749, 2004.