Proceedings of the 2002 Winter Simulation Conference
E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds.

USING SIMULATION-BASED SCHEDULING TO MAXIMIZE DEMAND FULFILLMENT IN A SEMICONDUCTOR ASSEMBLY FACILITY

Juergen Potoradi
Ong Siong Boon

Infineon Technologies
FTZ, Batu Berendam,
75350 Melaka, MALAYSIA

Scott J. Mason

University of Arkansas
4207 Bell Engineering Center
Fayetteville, AR 72701, U.S.A.

John W. Fowler
Michele E. Pfund

Arizona State University
P. O. Box 875906
Tempe, AZ 85287, U.S.A.

ABSTRACT

This paper describes how a large number of products are scheduled to run in parallel on a pool of wire-bond machines to meet weekly demand. We seek to maximize demand fulfillment subject to system constraints. The schedule is generated by a simulation engine and used to control the machines at execution time and also to plan for the start of material. By using online data for equipment status and WIP availability, the schedule adapts to “unforeseen” changes on the shop floor after a simulation run. The frequently updated schedule redirects the line towards maximum demand fulfillment based on the latest status of the line.

1 SEMICONDUCTOR ASSEMBLY

Semiconductor manufacturing is among the most complex manufacturing processes as described by Sze (1985). Semiconductor devices are highly miniaturized, integrated electronic circuits consisting of thousands of components. Every semiconductor manufacturing process starts with raw *wafers*, thin discs made of silicon or gallium arsenide. Depending on the device and wafer diameter, up to a few hundreds of identical chips can be made on each wafer, building up the electronic circuits layer by layer in a *Wafer Fab* (see Figure 1). Next, the wafers are sent to *Probe*, where electrical tests identify the individual die that are not likely to be good when packaged. The bad die are either physically marked or an electronic map is made of them so that they will not be put in a package. The probed wafers are sent to an *Assembly* facility where the “good” die are put into the appropriate package. Finally, the packaged die are sent to a *Test* facility where they are tested in order to ensure that only good products are sent to customers.

Wafer Fab and Probe are often called the “Frontend” and Assembly and Test are often called the “Backend”. While Frontend operations are typically done in highly industrialized nations, Backend operations are often done in countries where labor rates are cheaper. Considering the scale of integration, the type of chip, the type of package, customer specs, the whole manufacturing process may require up to 500 processing steps and up to three months to produce.

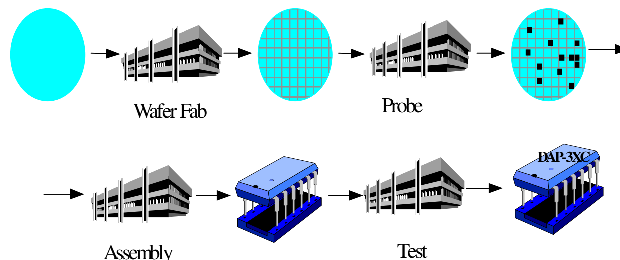

Figure 1: Steps of Semiconductor Manufacturing

Crucial factors of competitiveness in semiconductor manufacturing are the ability to rapidly incorporate advanced technologies in electronic products, ongoing improvement of manufacturing processes, and last (but not least) the capability of meeting due dates for an optimal customer satisfaction. In a situation where prices as well as the state of technology have settled at a certain level, the

capability of meeting due dates along with the reduction of cycle time probably has become the most decisive factor to stand the fierce competition in the global market place. Consequently, operations managers are under increasing pressure to ensure short and predictable cycle times.

Semiconductor assembly is sometimes called packaging. There are usually more types of parts being made in an assembly factory than in a wafer fab, but each part type requires 10-20 sequential processing steps instead of 400-500 reentrant processing steps. One difficulty in modeling assembly operations is the fact that a lot is often divided into sub-lots with each sub-lot being sent to the next machine when it completes an operation. Thus, one lot may be being processed across several machines at the same time. Another difficulty is that there is often a very significant amount of setup required to changeover from one product type to another. Finally, batching machines are also often present in assembly factories. Brown *et al.* (1999) document recent assembly modeling work done at Infineon Technologies. In this paper, we investigate the use of a simulation model to schedule weekly production at an Infineon assembly plant.

2 MATERIAL FLOW

The assembled integrated circuits (ICs) produced in Infineon's Melaka factory are for various applications in the automotive, wireless and communication industry. They are customized for the specific application and customer. Some of the applications are anti-lock braking systems (ABS), airbag systems, mobile phones and communication networks.

The high product diversity significantly impacts the first two assembly steps, i.e., the die-attach and wire-bond processes (see Figure 2). These are the processes with individual settings with respect to the device (the actual semiconductor die) that determines the IC functionality. To change these machine settings takes an average of 1 hour. The remainder of the process is to protect the device (molding) and to provide the interface (pins) to the application. The standardized dimensions of the molded ICs together with the number and location of the pins is known as a package. The process steps from molding onwards depend solely on the package (see Figure 2).

The interesting areas for scheduling application in the Melaka assembly factory are the die-attach and wire-bond areas. This is because the areas have 50 times higher product diversity and also because of the much higher number of machines in these areas. The higher number of machines is based on the significantly lower throughput of these machines. A wire-bonder generally has only 2.5%-5% the throughput capability of the downstream machines. A die-attach machine generally only has 25%-50% of the throughput capability of the downstream machines. The scope of the work reported herein is die-attach to wire bonding. To control the material flow from molding onwards we installed tools for due date driven dispatching.

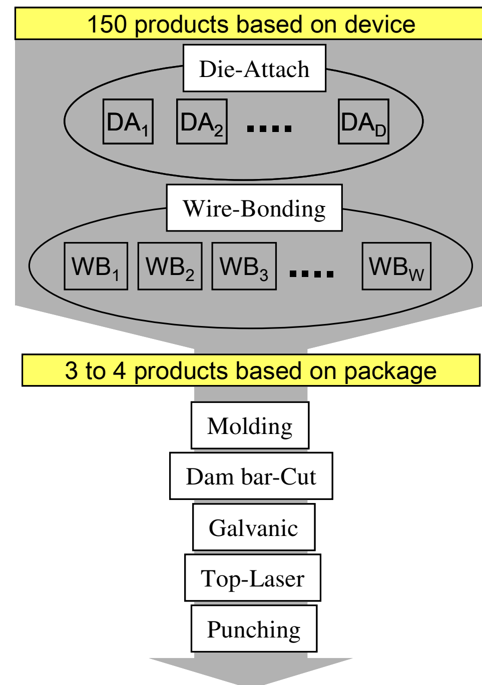

Figure 2: Assembly Flow Diagram

3 APPLICATION

To get the full benefit out of a controlled material flow in an assembly facility, lot release plays a major part. In environments with linear material flow all material proceeds from lot release via non-bottlenecks to the bottleneck. This is different than reentrant flow environments (e.g., wafer fabs), where just a small portion (e.g., 5% for 20 cycles) of the material arrives directly from the lot release to the bottleneck without passing the bottleneck itself in a prior cycle.

Based on this, our approach is to control the lot release and all non-bottlenecks in such a way to feed the bottleneck. The wire-bonders are considered the bottleneck. This is the area where the number of setups (product changes) needs to be controlled to ensure sufficient capacity.

We start with a plan (schedule) for the wire bonders and derive from there a sequence for the die-attach and lot release to suit this plan. The wire-bond plan is generated using a discrete event simulation engine that is described in the following section. Based on the time horizon there are two complementary applications of the wire-bond scheduler targeting the same objective of maximizing demand fulfillment:

1. Daily schedule generation for lot release planning.
2. Hourly schedule generation for setup control.

The lot release is planned two to three days ahead, to allow time for upstream processes and material preparation. For this daily planning is sufficient. This also includes daily model adjustments by an expert.

For setup control at the execution level, the schedule must always capture the latest status of the line. Due to unpredictable events the schedule becomes obsolete within hours or even minutes. Ideally, the schedule is regenerated after each major unforeseen event.

4 SCHEDULE GENERATION

We describe the process of schedule generation in three steps, from data preparation to model generation and execution and finally to Gantt-chart generation and evaluation.

4.1 Input Data

Equipment status and WIP data is extracted online from shop floor monitoring systems. The latest demands are captured from the planning systems. This is the most volatile input to the model that might change on an hourly basis. Therefore we considered it most important to make this data fast and easy accessible.

Product, process speed and general equipment information is maintained directly in the commercial simulation software, Factory Explorer® from Wright Williams and Kelly (Chance, 1996). The software provides an Excel®-based user interface, that makes data entry and maintenance comfortable. The model consists of 82 wire-bonders and 130 products. Die-attach is not modeled explicitly, but merely as a delay step.

4.2 Model Execution

With the current automation level the model is maintained and executed on a daily basis. Infineon plans in a next step, to fully automate the data entry from the shop-floor and planning systems. That way the model-update becomes much faster and does not require an expert, thus can be done on a more frequent basis.

Each simulation run generates a schedule for a time window of one week. This time horizon is sufficient to plan for the lot starts and also to confirm the delivery for the running work week.

The model is initialized with actual lots available in the line as well as outstanding demand due in the current week plus the following two work weeks. The outstanding demand scheduled for the first days serves as start plan for die-attach.

4.3 Gantt Chart

The event-file from a simulation run is then converted into a more readable Gantt chart. Figure 3 shows part of the overall Gantt chart with a three day time horizon on the X-axis and 26 wire-bonders on the Y-axis.

The tool used to generate the Gantt chart was developed to allow a visual evaluation of the schedule. The

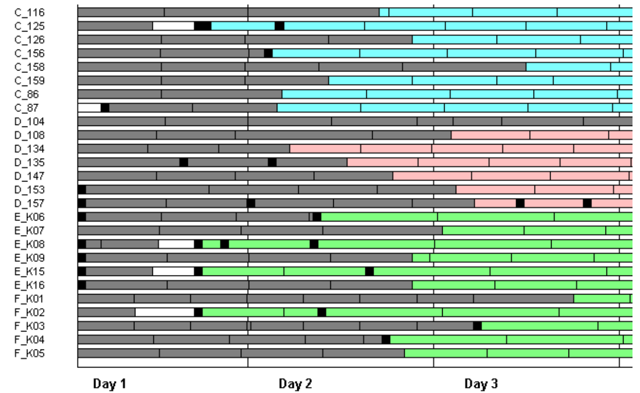

Figure 3: Gantt Chart, Showing Available WIP in Gray

horizontal bars consist of segments in different colors representing the activities scheduled for the respective wire-bonder. Black segments stand for setups, white segments for idling periods. The other segments are individual lots. Lots in gray color are determined by a certain selection criteria. In figure 3 the gray lots are available as WIP in the line. Lots of a certain product or with a certain due date are other possible selections.

5 EXPERIMENTAL RESULTS

Once the wire bonder schedule has been generated automatically (“the base schedule”), we use a two-step approach to improve the base schedule in terms of the average number of setups per day and job tardiness. By incorporating this two steps into the final model we enhance the local dispatch decisions and hopefully increases the average output per day.

5.1 Base Schedule

The initial base schedule contains product, machine, demand and process flow information with setup constraints. A preliminary review of the base schedule reveals more than 20 individual wire bonders running the same product at the same time, albeit for a short period (see gray highlighted area of the Gantt chart in figure 4). This is indicative of an unnecessarily large number of setups being performed while the workload for a given product is spread evenly across a large number of machines.

5.2 Step 1

We must ensure the number of setups required by the base schedule is feasible. Capacity calculations from Factory Explorer® estimate the number of machines required to bond the required demand in the amount of time available. The results of these calculations are used to restrict the number of machines that can simultaneously share the same setup ID. Figure 5 reveals a reduction from 50 setups in the base

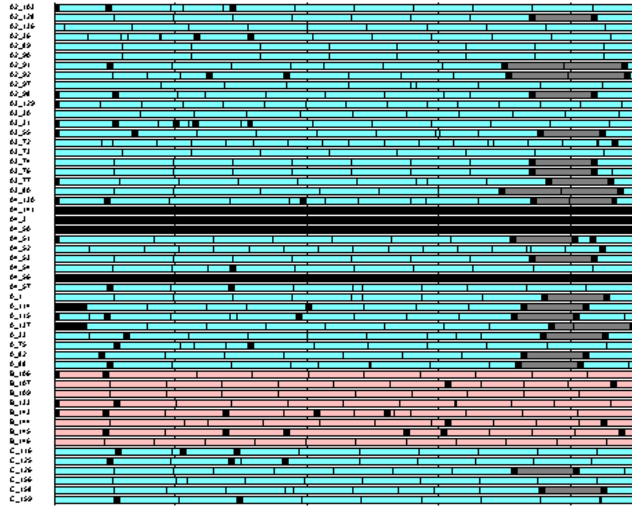

Figure 4: Gantt Chart for Base Schedule

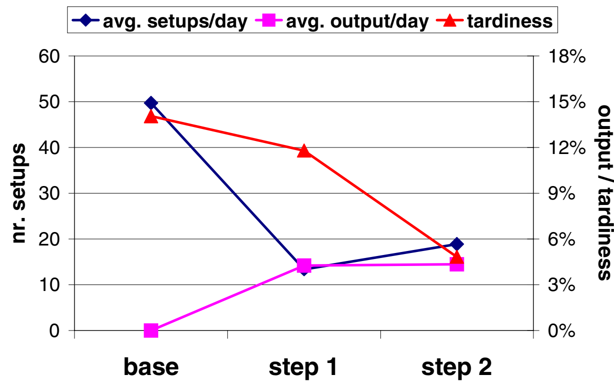

Figure 5: Experimental Results for Two Step Process

model to 13 after step 1 (refer to the left-hand vertical axis). Further, this reduction in setups also produces a 4% increase in output, thereby resulting in superior demand fulfillment (refer to the right-hand vertical axis). The output is shown as percent increase relative to the base schedule.

Job tardiness % is defined as the fraction of the WIP due in workweeks one and two that is not wire bonded after day 4 (Wednesday). Finally, the percent of tardy jobs is slightly reduced from 14% to 11% after Step 1 (refer to the right-hand vertical axis).

5.3 Step 2

To further improve upon Step 1 by reducing job tardiness, we prioritize the lots based on their due date. Using a simple prioritization schedule, we assign those lots due in week 1 a higher priority than those due in weeks 2 and 3. Similarly, those lots due in week 2 are given a higher priority than week 3 lots.

Figure 5 indicates our priority scheme causes more setups (an average of 19 per day). However, no significant impact on output is evident. Further, the number of daily setups prescribed by Step 2 is feasible in Infineon's factory, given there are two setup technicians available at any time in the line. Because a single setup requires one hour, Infineon can theoretically perform 48 setups per day. This in turn drives a low utilization of the setup technicians. Therefore, we do not expect to see jobs experiencing long times waiting for the setup operators (Hopp and Spearman, 1996).

Finally, Figure 5 does show a significant reduction in job tardiness percent after Step 2, a reduction to 4%. Experimental results indicate the fine pitch (FP) machines are the primary reason for this improvement.

6 CONCLUSIONS AND FUTURE RESEARCH

In this paper, we have described a recently developed system to schedule weekly production in the assembly plant of a major semiconductor manufacturer. The system uses a discrete event simulation engine to schedule the bottleneck equipment group, the wire-bonders. We include prioritized lots based on due date and include additional resources for setups to ensure that the schedule generated is feasible. This methodology is providing significant improvements in the factory's tardiness performance. In the future, we will investigate the efficacy of replacing the simulation-based schedule generator with an optimization scheme.

ACKNOWLEDGMENTS

The team gratefully acknowledges the Director of Infineon Technologies Advanced Logic in Melaka and the General Manager of Infineon Technologies AP for their support of this research. Scott Mason, John Fowler and Michele Pfund are partially supported by SRC grant 2001-NJ-880, which is jointly funded by SRC and International SEMATECH.

REFERENCES

Brown, S., J. Domaschke, and F. Leibl. 1999. No Cost Applications For Assembly Cycle Time Reduction, *International Conference on Semiconductor Manufacturing Operational Modeling and Simulation*, pp. 159-163.

Chance, F. 1996. Factory Explorer® users' guide.

Hopp, W. J. and M.L. Spearman. 1996. *Factory physics: foundation of manufacturing management*. Chicago: Irwin.

Sze, S. M. 1985. *Semiconductor Devices: Physics and Technology*. New York: John Wiley & Sons.

AUTHOR BIOGRAPHIES

JUERGEN POTORADI is a Production Manager with Infineon Technologies (formerly Siemens Semiconductor Division). He is currently responsible for production operations, short term planning and logistics in a backend

factory in Malaysia. Before, he was a Factory Modeling and Simulation analyst and project leader for implementing simulation techniques and methodologies in Singapore and Malaysia. Mr. Potoradi received his undergraduate and graduate degrees in Computer Science from the University of Wuerzburg in Germany. He has extensive experience in simulation analysis of semiconductor wafer fab and back-end production operations. His email address is <juergen.potoradi@infineon.com>.

ONG SIONG BOON obtained his Bachelor in Computer Science from the University Sains Malaysia. He joined Infineon Technologies as Systems Engineer 2 years ago. He has been involved intensively in a variety of projects supporting production and logistic in the IT manufacturing area. This includes development of real-time production monitoring and planning application as well as online reporting tools for cycle time and equipment performance. His research interests include real-time computing, and application of artificial intelligence in planning and shop-floor control problems. His email address is <siong-boon.ong@infineon.com>

SCOTT J. MASON is an Assistant Professor in the Department of Industrial Engineering at the University of Arkansas and a registered Professional Engineer in the state of Arkansas. Prior to his current position, he spent eight years working on factory modeling, simulation, and capacity analysis projects at SEMATECH, Advanced Micro Devices, Intel, Wright Williams & Kelly, National Semiconductor, Micrel Semiconductor, and Seagate Technologies. Dr. Mason received his B.S.M.E. and M.S.E. from The University of Texas at Austin, and his Ph.D. from Arizona State University. His interests include modeling and analysis of semiconductor manufacturing systems, applied operations research, and factory scheduling and production control. He is a member of IIE and INFORMS and serves as a Technical Advisor to Integral Wave Technologies. His email address is <mason@uark.edu>.

JOHN W. FOWLER is an Associate Professor in the Industrial Engineering Department at Arizona State University. Prior to his current position, he was a Senior Member of Technical Staff in the Modeling, CAD, and Statistical Methods Division of SEMATECH. He received his Ph.D. in Industrial Engineering from Texas A&M University and spent the last 1.5 years of his doctoral studies as an intern at Advanced Micro Devices. His research interests include modeling, analysis, and control of semiconductor manufacturing systems. Dr. Fowler is the co-director of the Modeling and Analysis of Semiconductor Manufacturing Laboratory at ASU. The lab has had research contracts with NSF, SRC, SEMATECH, Infineon Technologies, Intel, Motorola, ST Microelectronics, and Tefen, Ltd. He is also an Area Editor for ***SIMULATION:*** *Transactions of the Society for Modeling and Simulation International* and Associate Editor of *IEEE Transactions on Electronics Packaging Manufacturing*. He is a member of ASEE, IIE, IEEE, INFORMS, POMS, and SCS. His email address is <john.fowler@asu.edu>.

MICHELE E. PFUND recently received her Ph.D. in Industrial Engineering at Arizona State University, specializing in the area of operations research. She received her M.S. in Industrial Engineering from Purdue University and her B.S. in Chemical Engineering from Case Western Reserve University. She serves as the editor for the INFORMS student journal OR/MS Tomorrow. Her e-mail address is <michele.pfund@asu.edu>.