Capacity Adjustment in Semiconductor Fabrication During Peak Periods


ABSTRACT: Semiconductor demand is dominated by a three-year cycle compounded by unpredictable slumps and peaks in customer orders. During the peaks, demand may exceed available foundry capacity, and during a slump period, the drop in demand often results in plant utilization of 50% capacity. The constraint in the overall supply chain lies in the capacity of the semiconductor fabrication plant to fabricate chips on the wafer. Cycle times are usually 30 to 60 days with 300 to 900 steps, of which 30% are reentrants to the same equipment depending on recipe complexity. The largest semiconductor suppliers continue to expand factory capacity to cater to periods of high demand. However, smaller suppliers need other methods to maneuver through the demand cycle. This research applies a capacity adjustment approach for smaller high-mix foundries to optimize throughput and meet short-term spikes in demands using existing resources. A what-if simulation model that adjusted the shared equipment capacity was successful in increasing fabrication output over a two-month period by selecting wafer lots with few remaining shared equipment steps. The output for the first month was increased 21% versus the standard approach, which was followed by an output decrease of 50%. This approach has been successfully implemented in a wafer fabrication foundry.

KEYWORDS: Semiconductor Fabrication, foundry, cycle time, capacity; re-entrance.

1. INTRODUCTION

Analysis conducted by Semico Research in 2007 shows that the aggregate sales of game player devices, primarily those from Sony and IBM, reached 1 million within three months[1], a staggering feat when compared with black-and-white television sets which took almost 20 years to reach a similar figure. Since then, the demand for electronic goods has intensified, evidenced by October 2011 preorder sales for the iPhone 4S reaching more than 1 million on the first day[1]. To meet these demand spikes, electronics manufacturers are forced to rethink production methods to produce more products within shorter cycle times.

Global semiconductor revenues cycle with economic conditions [2], [3]. Figure 1 shows peak demand in 2010, followed by a major decline in 2011, and a further decline in 2012. Semiconductor trends reports, the NASDAQ Semiconductor Outlook reports[4], the semiconductor utilization reports [5], and the SC-IQ [6], [7] reports, are consistent in showing an annual cyclic trend in semiconductor revenue. Whether the cyclic trend starts at beginning of the year, middle of the year, or end of the year depends on the companies’ target market segments. The rapid changes in demand stress semiconductor fabrication plant utilization. A company usually takes action during the down cycle, such as workforce layoffs to cut costs and improve efficiency[8][9] or maintains a safety stock to meet uncertain demand and forecast errors[10]. Manufacturers need a solution with the potential to meet a demand surge during a peak period followed by a demand drop. As business cycles are expected to continue [2], a solution to improve efficiency during a cycle is essential.

Manufacturers’ supply chains are constrained by semiconductor fabrication plant capacity, which have a cycle time per chip of 30 to 60 days [11], [12] depending on the complexities involved. The relationship between plant cycle time and operational utilization is exponential [13]–[18]. Numerous publications have discussed cycle time strategies and propose utilizing bottleneck equipment with the threshold set between 80% to 90%. WIP inventory controls ensure a linear WIP profile that avoids unnecessary stress on equipment, cycle time expansion resulting from unpredictable and long queue times, and missed delivery dates [19]. Existing literature also suggests
continuous improvements in equipment throughput, efficiency, and continued investment in equipment to expand capacity [11], [13], [20]. But few publications explore the potential benefits of utilizing the market trends to improve plant capacity. Such an approach would help reduce capacity constraints over the three-year semiconductor demand cycle [21], in which, during the up period, demand exceeds the available capacity and, during the down period, utilization is less than 50% of plant capacity.

Figure 1: Semiconductor annual growth, from 2007 to 2014, follows the peaks and troughs of global business cycles. [2]

This study investigated a novel method of increasing capacity in a semiconductor fabrication foundry by improving the shared equipment capacity. This improvement targeted semiconductor fabrication foundries using complementary metal-oxide semiconductor core technology on 200 mm wafers that process 20 to 50 different products at any one time. Although this study focuses on facilities that fabricate 200 mm wafers, the outcomes and findings can be applied to other wafer fabrication facilities that produce wafers of different sizes, because they use a 200 mm capacity equivalent as the benchmark when stating capacity [20]. Globally, there are 71 fabrication plants manufacturing 200 mm wafers [21], but the method also benefits 28 foundries with 300 mm wafer fabrication facilities [22] The electronics industry as a whole benefits from this study because it offers an alternative method of increasing capacity during periods of high demand, which cannot be achieved by adding extra equipment because equipment shipment, installation, and process qualification takes longer than the peak period time frame [11], [23].

2. DEVELOPMENT OF SIMULATION MODEL FOR MANUFACTURING OPERATION

While other research focuses on the development of manufacturing prediction models [12], [15], this analysis focuses on finding a solution that increases semiconductor chip output during short, temporary periods of high demand. The investigation is underpinned by a simulated what-if analysis, modeled using the software AutoSched AP with programming platform of Advanced Productivity Family from Applied Material, a semiconductor equipment and software company. This software provides the integration platform with real time and historical manufacturing database AutoSched AP. This integration platform allows the user to develop and configure a simulation model easily specially to extract information related to WIP profile and WIP scheduling module. The simulation model than is developed to configure factory configuration which include, equipment, WIP, process flows, cycle time and others manufacturing information.

Table 1: Generic Output from the Simulation Model

<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Lot</th>
<th>Quantity</th>
<th>Step</th>
<th>Equipment</th>
</tr>
</thead>
<tbody>
<tr>
<td>3/14/17</td>
<td>12:22</td>
<td>S-xxx</td>
<td>25</td>
<td>OQAXX-XX-02</td>
<td>OQAXXX001</td>
</tr>
<tr>
<td>3/14/17</td>
<td>13:11</td>
<td>S-xxx</td>
<td>25</td>
<td>OQAXX-XX-01</td>
<td>OQAXXX003</td>
</tr>
<tr>
<td>3/14/17</td>
<td>14:48</td>
<td>S-xxx</td>
<td>25</td>
<td>ETSXXX-XX-02</td>
<td>ETSXXX001</td>
</tr>
<tr>
<td>3/15/17</td>
<td>05:13</td>
<td>S-xxx</td>
<td>25</td>
<td>DIFXXX-XX-16</td>
<td>DIFXXX006</td>
</tr>
<tr>
<td>3/15/17</td>
<td>08:42</td>
<td>S-xxx</td>
<td>25</td>
<td>CLINXXX-XX-28</td>
<td>CLINXXX004</td>
</tr>
<tr>
<td>3/15/17</td>
<td>12:42</td>
<td>S-xxx</td>
<td>25</td>
<td>ETHXXX-XX-66</td>
<td>ETHXXX112</td>
</tr>
<tr>
<td>3/15/17</td>
<td>16:17</td>
<td>S-xxx</td>
<td>25</td>
<td>PHOXXX-XX-37</td>
<td>PHOXXX022</td>
</tr>
<tr>
<td>3/16/17</td>
<td>02:26</td>
<td>S-xxx</td>
<td>25</td>
<td>TFDXXX-XX-64</td>
<td>TFDXXX112</td>
</tr>
<tr>
<td>3/16/17</td>
<td>09:37</td>
<td>S-xxx</td>
<td>25</td>
<td>TFDXXX-XX-48</td>
<td>TFDXXX007</td>
</tr>
<tr>
<td>3/16/17</td>
<td>11:51</td>
<td>S-xxx</td>
<td>25</td>
<td>ETHXXX-XX-62</td>
<td>ETHXXX111</td>
</tr>
<tr>
<td>3/16/17</td>
<td>21:42</td>
<td>S-xxx</td>
<td>25</td>
<td>TFMXXX-XX-14</td>
<td>TFMXXX001</td>
</tr>
</tbody>
</table>

The types of information used in the model include work in progress, equipment utilization, setup time, preventive maintenance, and process efficiency. The data were taken from two subsequent months to reflect the historical semiconductor revenue trends shown in Figure 1.

The expected output of the simulation model is show in Table 1. The data was then summarized into daily output (Move) for equipment and the expected (forecast) completion of each product or lot, and plotted against actual data in the semiconductor fabrication factory to validate move accuracy. Accuracy validation did not consider accuracy of product forecast due to product priority changes during the respective periods that caused a high number of outliers in the analysis. The simulation model produced consistently accurate predictions of move per day results versus actual moves, shown in Figure 2. Daily output from equipment reflected the accuracy of the simulation model compared to actual [15].

The output from equipment per day (Move) was then measured. Move per day is defined as the output of wafers from respective equipment per day after completing the process. In the same day, a product can be processed in 6 to 12 different processing steps using mostly different equipment.

Accuracy was defined as how close the forecast is compared to actual practice, and is critical to understand factory capability to provide in order to use the model for improvement.
analysis. Thus, the equation to validate the accuracy is as follows:

$$\text{Accuracy} = \frac{(F - A)}{A} \times 100\%$$  

(1)

where:  
A = Accuracy of the Move  
F = Forecasted Move  
A = Actual Move

When the model was executed for the first time using information gathered through analysis and a survey, 73% accuracy was achieved. Redefining the equipment into categories, such as single processing, process chamber configuration, recipe availability, closed the accuracy gap to 96%.

The simulation model was validated using Paired t-test to compare the values of forecast versus actual. The analysis, performed in MS Excel, hypothesized:

1. Baseline Hypothesis ($H_0$): Forecast Move is same as actual move.
2. Alternate Hypothesis ($H_1$): Forecast is different than actual move.

The model validation results, representing actual factory operation at 99% confidence level (CI). The results from p value shows that it is more than 0.05, therefore there is no enough evidence to reject baseline Hypothesis at 99 CI.

Improvement made to achieve the simulation model to make it acceptable in hypothesis testing include getting accurate processing time based on process start and process end, correctly configure on the batching equipment into interval batching for wafer clean equipment with bath process type and typical batching, efficiency setting reflect to the interval time and apply right dispatching rules that similar used in the shop floor.

![Table 2: Paired t-test on forecast versus actual output at 99% confidence level](image)

<table>
<thead>
<tr>
<th>Statistical parameter</th>
<th>Estimated</th>
<th>Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>750.36</td>
<td>782</td>
</tr>
<tr>
<td>Variance</td>
<td>0.585</td>
<td>1024</td>
</tr>
<tr>
<td>Observations</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>0.1136</td>
<td></td>
</tr>
<tr>
<td>Pooled Variance</td>
<td>912</td>
<td></td>
</tr>
<tr>
<td>Hypothesized Mean Difference</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Df</td>
<td>32</td>
<td></td>
</tr>
</tbody>
</table>

3. PSEUDO CAPACITY FOR OUTPUT INCREASES

The complexity of semiconductor wafer fabrication lies in dealing with 300 to 1,000 processing steps, and the repetition of similar processes. In this case study, the technology was 0.16 µm for 200 mm diameter wafers with an average of 32 masking layers. The process is repeated 32 times for photolithography tools, and the equipment used in these repetitive processing steps are placed either in the front-end or back-end. The average number of processing steps that share the same equipment in the front-end for this particular case is 6.3, and that share the back-end equipment is 5.7. The sources of Overall equipment effectiveness (OEE) formulation are based on FabTimes and SEMI which widely used for semiconductor equipment performance benchmarking [3], [16], [24], [25]. The capacity analysis is based on the following formulation. The components of Overall Equipment Effectiveness (OEE) are illustrated below, the definition for Run, Idle and Down are based on semi E10 standard.[15].

$$\text{Daily Capacity} \ SP = \text{WPH} \times \text{OEE} \times \text{Yield} \times 24\text{hrs}$$  

(2)

$$\text{OEE} = \text{Avail} \times \text{Perf} \times Q$$  

(3)

Capacity SP = Capacity for single step reentrance  
F = Forecasted Move  
A = Actual Move  
Q = Quality  
Avail = Equipment Availability  
Perf. = Performance Efficiency

Where, daily Capacity Wafer Fab with number of re-entrance illustrated in equation (4):

$$\text{Daily Capacity} = \frac{\text{Daily Capacity SP}}{\text{Number of Passes}}$$  

(4)

As overall capacity for a group of similar equipment is measured as capacity (day) multiplied by the number of pieces of equipment in the same group, reducing the number of passes improves capacity. Figure 3 shows that decreasing the number of passes from seven to six increases...
capacity by 16.7%. As the number steps sharing equipment is reduced capacity increased. The relation is closed exponential and the gains are huge. Therefore, by applying the same concept, the factory output can immediately increase for the current month to meet sudden demand increases but also meet the expectation of lower output needed for the following period due to lower demand. The results from this approach can be further enhanced if number of work in progress (WIP) is higher at the end of processing steps due to constraints equipment at the backend. Immediate capacity gain can be used to gain additional output for the respective month or quarter by reducing the number of step passes per equipment during the same time period. The additional output can be achieved by slowing down the work in progress (WIP) for higher remaining passes and focusing only on products with the respective remaining passes shared with the bottleneck equipment. This automatically resets the capacity number to a higher value and allows for additional output from the factory as described in Figure 3.

![Capacity (day) versus Number of shared steps per equipment](image)

**Figure 3:** Potential capacity increase (versus capacity standard) by reducing the number of shared-equipment steps. By decreasing the number of steps from seven to three, the output capacity of the first month increased 2.3 times.

This strategy works well in situations where the demand in the next period is likely to slow down because the capacity gains in the first month are made by selecting products that have a small number of passes remaining, leaving products with more processing steps and longer cycle times for the following month. To accurately predict cycle time for the next period, the simulation model should be run multiple times to fully understand the WIP profile of the systems, along with the remaining mask and potential capacity gains.

### 4 DISCUSSION OF RESULTS

The simulation model was successful in increasing back-end fabrication capacity by adjusting the shared equipment capacity, in accordance with the proposed theory (see Figure 3). As the opportunity to change factory output occurs at quarter change, the analysis selected month September (Month S) and month October (Month O) to represent months in two quarters (Month S = end of quarter, Month O = start of quarter). By selecting wafer lots with few remaining shared equipment steps, the monthly output for Month S was increased 21% to 27,337, versus the standard approach, which would only produce 22,511 wafers as shown in Figure 4.

There are 2 lines in the chart. The first line which is drawn lower than the other line representing baseline capacity based on average number of passes of step the product has in the factory. This is the standard (std.) approach for capacity forecasting. The second line is representing capacity with lower number of passes for product and the line is drawn increased higher than first line at the end of Month S. Number of passes is lower due to the strategy to select pool of product that able to meet the maximum output of the current month which Month S without considering product to for Month O that usually need to be delivered consistently. Based on the optimization approach reducing number of passes the new configured capacity for the wafer fab to meet cyclic demand if the demand is available for Month S, the output can be generated by additional 21%.

![Standard method versus capacity adjustment for Month S (end of quarter)](image)

**Figure 4:** Standard method versus capacity adjustment for Month S (end of quarter). This graph shows that artificially integrating capacity with WipP successfully increased the overall output of the factory from its original limit by 21%.

Figure 5 shows the similar output graph as in Figure 4, but for Month O. The extended simulation results to show trend of the forecasted output for Month O. The output for Month O is lower by 50% which to meet lower demand as in the cyclic trend. Figure 6 shows the comparison summary.
Figure 5: Standard method versus capacity adjustment for Month O (start of quarter). The impact of artificial capacity adjustment on Month O reduced output by more than 50%. In situations where a sharp drop in demand is expected, this is acceptable, and may fit into actual cycle trend of global semiconductor demand.

Other research that proposes continuous improvement does not support the rapid capacity changes needed during cyclic swings in demand [13], [26], [27].

Figure 6: Output capacity of standard method versus capacity adjustment method steps for two months output and cycle time. The capacity adjustment method boosted production by 21% and improved cycle time by 16% (1.8 DPML) to meet periods of high demand (Month S). This was followed by a drop in production output of more than 50% (Month O) as cycle time increased to 2.54 DPML.

The capacity adjustment method concept increases capacity for complex processes, with steps ranging from 300 to 1,000. The output increment in this analysis is 21% for Month S, but could be potentially higher or lower depending on the WIP profiles, equipment status and availability. The cycle time impact on the following month is significant, be as critical as 27% for wafer outputs of the following month due to the high waiting time for WIP that must be processed after re-prioritization. These capacity changes fit the established trend for higher demand in quarter three and lower demand in quarter four, and fit the SICAS capacity utilization graphs [5].

Furthermore, the capacity improvements are greater than other recent multi-objective dynamic scheduling techniques.

In addition, this research recognizes the WIP profile (WipP), which determines the location of the WIP and the number of steps remaining. This identifies the target number of shared steps to increase capacity to more than standard capacity, in fixed wafers per hour, yield, and overall equipment efficiency. To address this, a new related equation was established extended from Equation (4):

\[
P_{\text{Psuedo Cap}} = \text{Current Output} + \sum_{n=1}^{k} \left( \text{WipP} \times \text{Daily Capacity Remaining Passes} < k \right)
\]

where:
- \( n \) is the target number of WIP remaining of output for respective month with remaining passes
- \( k \) is the maximum number of output based on single passes capacity

Fewer shared steps between equipment reduced the complexity and increases the output of the factory. In the situation where the number of shared steps or complexity is high for respective factory like high mix and low volume, its will creates more opportunity to increase the true capacity of the factory. [10], [28] has also proposed a potential shop-floor control concept to maximize output in Month S and reduce output in Month O.

5. CONCLUSION

This analysis focused on finding a method for companies to maximize revenues during unpredictable demand cycles, such as when customers increase orders to capture market demand surges, or extend the due dates of existing purchase orders. It is not permanent solution to increase factory output, but a method to artificially increase wafer factory capacity for a particular month or quarter, followed by a reduction in output. Continuous improvement through other methods is needed to generate long term capacity improvements.

The results, generated from a what-if simulation model, were successful in increasing fabrication output by adjusting the shared equipment capacity. This was achieved by processing wafer lots that required fewer shared equipment passes. Based on this model’s data, production output for the first month (Month S) increased 21% to 27,337, versus the standard approach, which would only produce 22,511 wafers. This was followed by significantly reduced output for the second month—from 27910 wafers for the standard method to 12331 wafers for this adjusted method. There was also a corresponding impact on cycle time with a 16% (1.8 DPML) increase.
CAPACITY ADJUSTMENT IN SEMICONDUCTOR FABRICATION DURING PEAK PERIODS

DPML) improvement to meet periods of high demand, followed by an 41% increase in cycle time to 2.54 DPML (versus the standard of 2.0 DPML). Based on wafer selling price estimates from the TSMC 2012 financial report[20], [21], [29]–[31], this approach could increase a company's revenue by USD 5.8 million compared with holding the equivalent cost of wafers in inventory during the cyclic trend.

As the semiconductor industry demand follows business cycles [34], with unpredictable spikes and slumps adding to the problem, wafer manufacturers need to find solutions that maximize revenues during unstable times. This technique, successfully [35] implemented in a wafer fabrication foundry, provides wafer manufacturing companies an immediate solution to increase revenues during peaks in demand [36].

ACKNOWLEDGEMENT

The research presented is cosponsored by two parties, Collaborative Research in Science, Engineering & Technology (CREST) grant (no: P16C1-12) and Silterra Malaysia Sdn. Bhd.

REFERENCES


Development of Capacity Indices for Semiconductor Facilities,” in Proc. 2011


