INTEGRATING TARGETED CYCLE-TIME REDUCTION INTO THE CAPITAL PLANNING PROCESS

Navdeep S. Grewal
Alvin C. Bruska
Timbur M. Wulf

Wafer Systems Department
Seagate Technology,
One Disc Drive
Bloomington, MN 55435, U.S.A.

Jennifer K. Robinson

C2MS Productivity Solutions
7171 Buffalo Speedway, #639
Houston, TX 77025, U.S.A.

ABSTRACT

This paper describes the development and application of an integrated static capacity and dynamic simulation analysis methodology for purchasing equipment capacity. The goal of the study is to address targeted cycle time objectives in a start up Recording Head Wafer manufacturing facility at Seagate Technology, Minneapolis, MN. The short product cycle time, coupled with the competitive nature of the disc drive industry, has made cycle time reduction one of the most important objectives of production capacity planning. This paper describes an equipment procurement strategy in which static capacity analysis is used to identify an initial equipment set with a low slack capacity variable on each tool group. Simulation analysis is then used to identify the critical tool groups that contribute to cycle time delays. The Seagate Industrial Engineering team used the simulation analysis tool Factory Explorer® from Wright, Williams & Kelly to perform the cycle time reduction analysis. This targeted approach is compared to the traditional static capacity planning approach of globally applying reserve capacity buffers of 20% or more to achieve the same cycle time reduction goal. Overall, the targeted approach has proven to be efficient in terms of minimizing capital equipment expenditures and also effective on the factory floor.

1   INTRODUCTION

In the competitive semiconductor industry, manufacturers closely monitor their manufacturing performance measures. The foremost performance measure for any semiconductor company is the manufacturing facility's (fab) cycle time. The process studied here is the manufacturing of wafers to make disc drive heads. The reentrant wafer process has more than 400 complex steps across 100 advanced tools with random uptime and processing times. Continuous process improvement and the introduction of new technology have led to shorter product life cycles, while simultaneously making the wafer manufacturing process more complex. Shorter product life cycle times have also made it necessary to reduce the wafer cycle times while maintaining the same level of production capacity. Many benefits may be attributed to reduced cycle times, including shorter learning curves, reduced scrap, and general process improvement (Nemoto *et. al.*, 1996. Potti and Mason, 1997). This paper outlines a capacity planning methodology formulated to include cycle time objectives in the capital purchasing procedure, using both simulation and static capacity analysis.

Simple spreadsheets are useful for analyzing capacity quickly. However, they cannot accurately assess cycle time repercussions. Another modeling approach is that of analytical queuing network models (AQNM). These models can provide quick estimates of steady state results regarding total system output and average resource utilization. They can be invaluable in making fast turn around decisions and in screening alternative scenarios. Another benefit to AQNM models is that they require a relatively small number of data inputs. However, some drawbacks exist. Unlike simulation models, which provide transient state results, AQNM models usually analyze the system under "steady state" conditions. They also generally require limiting assumptions about the system characteristics like rework, reentrant flow (multiple visits to the same tool group), and non-exponential random failures. Such dynamic and detailed analysis requires the use of discrete event simulation. Therefore, this paper does not discuss AQNM models further. Interested readers are referred to research papers by Suri and Diehl (1988) and Suri *et. al.*, (1993) for more information.

In most static capacity models, excess capacity of 10% to 30% is maintained across all equipment groups to fulfill cycle time objectives. This "brute force procedure" of installing excess buffer capacity at all the tool groups is in practice a very costly method of ensuring low cycle times. A more cost-effective method is to first plan a tool set with a smaller buffer of slack capacity across all equipment

groups, and then purchase high cycle time contribution tools to reduce overall fab cycle time. This method does not guarantee a mathematically "optimal" cycle time (best cycle time for lowest cost). For Seagate, however, it has rendered an acceptable cycle time at a much lower cost than the less efficient approach of maintaining a large global slack capacity variable across all the tools.

Seagate made their latest expansion of the wafer manufacturing facility at Minneapolis, MN by commissioning a new fab. One of the key objectives assigned to the Industrial Engineering capacity planning group was to develop an organized approach for benchmarking the new fab cycle time and purchasing equipment capacity to meet the cycle time objectives. Seagate hired Wright, Williams & Kelly (WWK) to assist the Industrial Engineering team in this effort. The analysis tool was WWK's Factory Explorer® modeling package. Factory Explorer® (FX®) is an integrated software package, capable of cost modeling, capacity analysis, and detailed factory simulation. FX® uses an Excel® spreadsheet as the front end for loading data and setting model parameters. This integration with Excel® allows users to exploit Excel®'s data manipulation features when storing data. Furthermore, it reduces the model preparation time significantly and simplifies the modeling task, compared with other user interfaces.

Creating a FX® model at Seagate requires loading data into several Excel® worksheets. One worksheet contains the product level information, including, for each product, the product name, start rate, default priority, lot size, release pattern, and process flow name. Another worksheet contains tool group information, including, for each tool group, number of workstations, downtime parameters, dispatch rule, tool capital cost, and minimum and maximum load size. Additional worksheets contain step-level data for each process flow defined in the model. For each step, tool group, processing time, and rework and scrap parameters are defined.

2 METHODOLOGY

The simulation project was divided into three modeling and analysis phases. Phase I involved collecting model input data, identifying critical system performance measures, and preparing the base model. In Phase II, the model was analyzed in detail by reviewing the output reports. An iterative method was then used in Phase III to assess the equipment capacity and develop an equipment purchase plan that would achieve cycle time goals for various phases of the production ramp.

2.1 Phase I

Phase I of the project included data preparation activities such as gathering equipment process times from the time standard database and obtaining engineering process time estimates for new tools. A single process flow (single product) was modeled. Historical equipment downtime data was collected from the maintenance group's equipment resource tracking system. To reduce the complexity of the project, material handling time between the stations was excluded from the analysis. Similarly, operators were not modeled. Inline process yields, rework data, and scrap data were downloaded from the shop floor control system. Setups were minimized through application of a setup avoidance dispatch rule. A smaller wafer lot size was assumed than the typical large size lots used by semiconductor manufacturers. Smaller lot sizes make equipment utilization very sensitive to batch load size. The default dispatch rule was first in first out (FIFO). The equipment loading rule invoked combined lots with the same recipes for processing. The minimum or maximum batch load size were set per rules followed on the production floor at each equipment group. For the random failures, mean time between failures and mean time to repair were modeled using an exponential distribution. Maintenance events were modeled with a constant distribution. Factory shutdowns were not modeled, nor were back-up tools or alternate process paths. Key modeling assumptions are summarized in Table 1.

Table 1: Key Modeling Assumptions

Maximum tool utilization set at 85% (for base model)
No material handling time modeled
Infinite labor assumed
Single product modeled

A base model was prepared with start up equipment bought for the new fab. Base runs were executed at a low start rate to verify the FX® simulation model. The base model output reports were analyzed to assess key system performance measures such as throughput, cycle time and equipment utilization. The fab loading was then set to the minimum anticipated production volume level. The three main performance measures for verification were: equipment count; equipment utilization by category (e.g. off-line % and busy %); and product cycle time.

When analyzing capacity, FX® first computes the available capacity for each tool group by downgrading the total scheduled time by unscheduled downtime, maintenance events, setups and repairs. FX® then predicts an estimated capacity loading % value for each tool group

that represents the percentage of available capacity at the tool group that is currently being used for production. (Chance, 1997). The suggested tool quantity in each group is then determined such that the capacity loading % value for the tool group stays below a user-defined global maximum loading (85% in this case). Percentage capacity loading in this study is considered to be same as equipment utilization. The user can either tell FX® to use the suggested equipment count, or can just use the calculated capacity loading number for comparison purposes. If the suggested quantities are used, the FX® capacity analysis module calculates the resources required to support the wafer schedule, while maintaining a maximum capacity loading of 85% on any tool group. This is very helpful in identifying the resource requirements without needing simulation, and also in avoiding unstable simulation runs. The model can also be loaded with actual tool counts and the pre-simulation capacity analysis used to refine the model.

The base case model was used to generate the tool list for the new fab. The capacity analysis was run and output data such as the bottleneck resource chart were reviewed to identify the capacity constraints. A sample bottleneck chart is shown in Figure 1, and displays the results of the capacity analysis, with top tools ranked by overall capacity loading. This chart illustrates the capacity usage of each tool group, broken down into free time, processing time, and various components of down time. The capacity analysis in general helps in short-listing alternative scenarios for subsequent detailed simulation analysis. This report was also used for model verification by comparing the equipment downtime (Off-line %) estimates against the observed values. The model was further validated and verified by comparing the FX® tool counts against the Seagate's spreadsheet static capacity model estimates and by comparing the model's behavior with actual shop floor data. The global capacity loading factor for all the tools was set at 85%.

The base model capacity analysis was also used to estimate the theoretical cycle time or "Raw process cycle time". Raw process cycle time is defined as the total time it takes to process a wafer lot, independent of queuing times, machine downtimes, rework, yield and other non product value-added times (Chance, 1997). After analyzing the FX® capacity analysis output reports, the base model simulation runs were executed. Key system performance measures tracked were tool utilization, system throughput, mean cycle time and queue delay time at each workstation. The average wafer cycle time predicted by the FX® simulation analysis was compared for verification against the factory cycle time and throughput data.
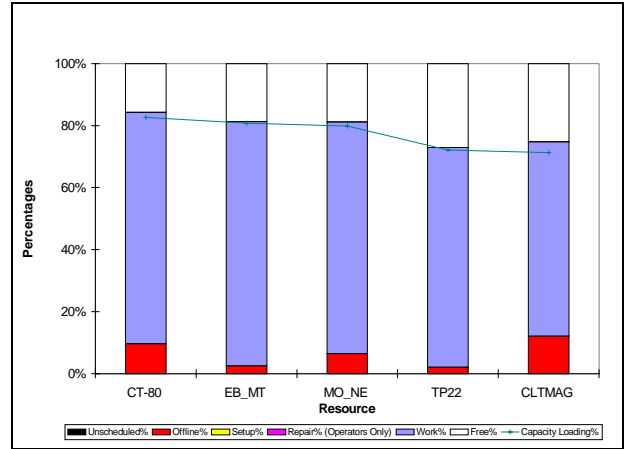

Figure 1: Example of Bottleneck Resource Chart

The top cycle time equipment contributors were identified by reviewing the FX® report that graphically represents the average time spent in queue and in service at each station. A sample is shown in Figure 2. This report lists the key tool groups that contribute to the product cycle time, ranked in order of cycle time delays. For various reasons, these tools are not always the bottleneck tools in terms of capacity. For example, although tool group BAS-02 is not heavily loaded it still contributes approximately one full day to the total cycle time. Simulation analysis helps to detect such cycle time contributors. This information sometimes leads to low-cost cycle time improvement opportunities. Instead of purchasing additional equipment, batch loading policies and dispatch rules can perhaps be modified to lower cycle time. This type of cycle time and queue size analysis is beyond the realm of pure static capacity analysis. Typically, in a static capacity analysis, the aforementioned tool would have never been suspected of causing cycle time delays because of its excess capacity.
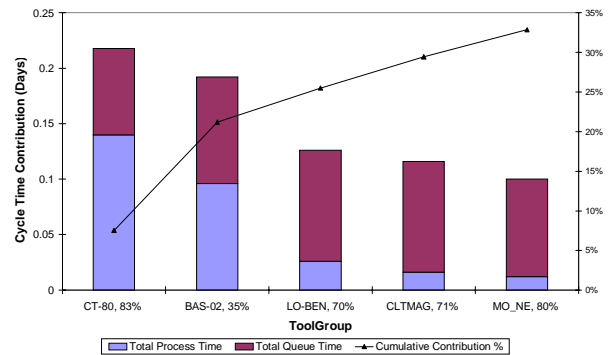

Figure 2: Example of Cycle Time Contribution Chart

2.2 Phase II

After validation and verification of the base model were completed, the simulation software was used to develop a capital equipment plan for a moderate factory production target. The cycle time target was set between two and three X, where X is the theoretical cycle time of the process. A series of simulation runs were performed for various global equipment capacity loading values. For each value, FX® generated the required minimum equipment set, and then ran the simulation to estimate the corresponding cycle time. Figure 3 shows the total equipment cost and average cycle time for each capacity loading value explored.
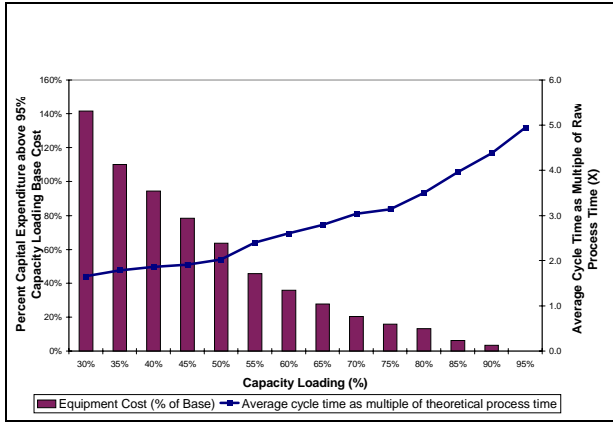

Figure 3: Cycle Time vs. Percent Capital Expenditure above the Base Cost for the Moderate Production Volume Level for Various Capacity Loading Values

Equipment sets with higher capacity loading values (e.g., 90%) have lower total equipment cost but longer cycle times compared to equipment sets planned with lower capacity loading values such as 70%. The main reason for this reduction in cycle time is the addition of more bottleneck servers at lower capacity loading levels. Also, the lower capacity loading equipment sets have fewer one-of-a-kind tools (tool groups containing only a single server). Although the total equipment cost is lower for the higher capacity loading models, the cycle times are significantly longer, especially for factories loaded above 85%. This data was used to illustrate the system behavior and to generate costs for equipment sets with large amounts of slack capacity applied to all tool groups. The latter were used for comparison with equipment sets derived via the informed capacity planning method studied in this paper. Cycle times were also observed to be lower for higher production volume level factories. This effect is shown in Figure 4, which plots the average cycle time against three production volume levels. For each volume level, an equipment set with 85% capacity loading was generated and simulation was run to estimate cycle time.
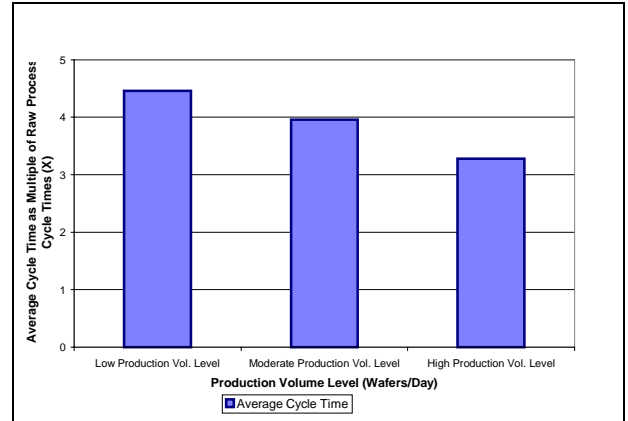

Figure 4: Average Cycle Time Chart for Various Production Volume Levels with 85% Capacity Loading.

The next step was targeted cycle time reduction, starting with a high capacity loading factory. WWK, based on their experience in a research project at a leading integrated circuit manufacturer (Fowler *et. al.,* 1997), proposed to use a heuristic optimization method to reach an acceptable solution. This heuristic required multiple analysis passes. For each pass, multiple candidate models were developed and investigated for cycle time reduction (Chance, 1996). The best candidate model became the base model for the subsequent simulation analysis. The specific steps in the analysis are shown below (from Chance, 1996):

Inputs: Production volume level, Capacity loading percentage, Budget Limit $ X Million.

Analysis Procedure:

1. Run Factory Explorer® capacity analysis to create a base model with minimum cost tool set.
2. Run Factory Explorer® simulation to estimate base cycle time and total queue delay time contribution by tool group.
3. For each of the top five tool groups in the base model (ranked by contribution to queue delay):
    a) Starting with current base model, add one tool to the selected tool group to form a candidate model.
    b) Run FX® simulation to estimate the cycle time for the candidate model.
    c) If the new cycle time is statistically significantly lower than the base cycle time, compute the ratio of cycle time reduction to tool fixed cost.
4. For the candidate model with the best (largest) reduction per dollar ratio, record the tool added and replace the base model with the candidate model.
5. Go to Step 3 or terminate (a) if the budget limit is reached or (b) if no candidate model results in a statistically significant reduction in cycle time.

Seagate investigated several different production volume levels, to plan for various points in the production ramp of the factory. Several different initial capacity loading values were also explored, to see how these differed in terms of the final recommended tool set. For each of the plans analyzed, the budget was never exhausted and the heuristic reached a stage when additional tools did not result in any significant reduction of cycle time. All simulations were run for two years, with a warm up period of six months to clear model statistics and minimize initialization bias.

2.3 Phase III

The chart in Figure 5 depicts the results of the cycle time optimization heuristic performed for the moderate production volume level. The analysis was also conducted for low and high production volume levels. For all levels, cycle time and total tool expenditure were measured for initial capacity loading values of both 70% and 85%.
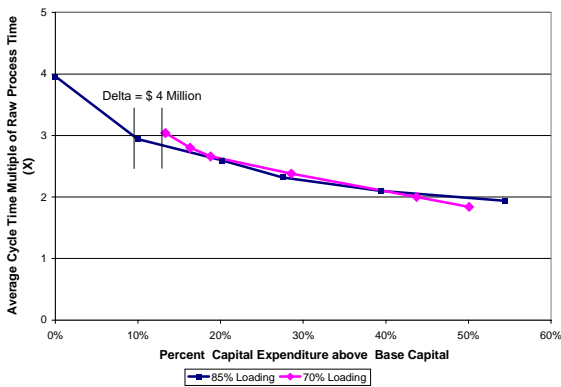

Figure 5: Cycle Time vs. Percent Capital Expenditure above the Base Capital for Moderate Production Volume Level with 70% and 85% Initial Loading

Key findings from this analysis are summarized below:

- The equipment set generated by starting with a global 85% capacity loading resulted in reasonable cycle times for all the production volume levels analyzed. This cycle time could be improved by purchasing additional high cycle time contribution tools, as shown in Figure 5.
- To achieve the average cycle time objective of 3X days, total equipment cost analysis was performed by using the two approaches, the aforementioned heuristic approach and the "brute force method" i.e., maintaining large slack variable across all the tools. When the analysis was run with a 70% capacity loading, the first iteration (with no extra tools) resulted

in a total tool cost of $ 75.6 Million with an average cycle time of 3X days. When starting with an 85% capacity loading and purchasing cycle time contributing tools a similar cycle time was achieved at a much lower cost of $ 71.5 Million – a net saving of nearly $4 Million, as shown in Figure 5.

- Regardless of which initial capacity loading number is used, this study shows that using the analysis procedure described in the previous section leads to a much more cost effective tool set than does a "brute force" approach of creating large slack capacity across all tool groups. To achieve an average cycle time of 2X days using the simulation procedure described above costs $ 9 Million less than it would cost to reach 2X days by the "brute force" method. This is because cycle time does not drop to 2X until the suggested loading is as low as 45% when globally applied (Figure 3). Overall, the most cost effective informed strategy is planning a minimal equipment tool set with a high capacity loading factor and then lowering the cycle time by selectively purchasing additional capacity.
- The graph in Figure 5 also shows the cycle time reduction achieved by adding more tools. The smooth slope gradient represents the achievement of a cycle time limit beyond which the addition of more tools would not statistically reduce the cycle time. A substantial amount of capital would have to be spent to attain significant cycle time reduction beyond 2X days.

3   CONCLUSION AND RECOMMENDATIONS

This project provided Seagate management with the information needed to purchase cost-effective equipment sets that could achieve cycle time objectives at various production volume levels. The analysis was also helpful in establishing the theoretical wafer cycle time benchmark and in predicting the average wafer cycle time. More experiments could be done to show cost savings by planning capacity at a higher capacity loading factor (e.g. 90% or 95%) and then purchasing additional equipment to reduce the cycle time. The scope of the project could also be expanded to perform more detailed analysis by including additional factors such as labor constraints and hot lots. This combined capacity and simulation analysis technique, targeted at high cycle time contribution tools, saved Seagate a significant amount of money by recommending the purchase of fewer tools than would have been needed by applying more slack capacity across all tools.

REFERENCES

Chance, Frank. 1996. "Developing Optimal Capital Purchase Plans," Seagate and Wright, Williams & Kelly internal project report.

Chance, Frank. 1997. Factory Explorer® Version 2.3 User Manual.

Fowler, John W., Steven Brown, Hermann Gold, and Alexander Schoemig. 1997. "Measurable Improvements in Cycle-Time-Constrained Capacity," *Proceedings of the 1997 IEEE/UCS/SEMI International Symposium on Semiconductor Manufacturing (ISSM)*.

Nemoto, Kazunori, Elif Akcali and Reha Uzsoy. 1996. "Quantifying the Benefits of Cycle Time Reduction in Semiconductor Wafer Fabrication," *Proceedings of the 1996 IEEE/CPMT International Electronics Manufacturing Technology Symposium*, 130-136.

Potti, Kishore and Scott J. Mason. 1997. "Using Simulation to Improve Semiconductor Manufacturing," Semiconductor International, July, 289-292.

Suri, Rajan and Greg W. Diehl. 1988. "Manuplan: A Precursor To Simulation For Complex Manufacturing Systems," Proceedings of the 1988 Winter Simulation Conference, 411-420.

Suri, Rajan, J. L. Sanders and M. Kamath. 1993. "Performance Evaluation of Production Networks," *Handbooks in Operations Research and Management Science, Vol. 4: Logistics of Production and Inventory*, edited by S. C. Graves, A. H. G. Rinnooy Kan, and P. H. Zipkin, Elsevier, Amsterdam, The Netherlands, 199-286.

AUTHOR BIOGRAPHIES

NAVDEEP S. GREWAL is an Industrial Engineer of Wafer Systems Department at Seagate Technology. He received a B.S. (1991) degree in Industrial Production Engineering from Gulbarga University (India) and an M.S. (1994) from University of Louisville. He is a member of Alpha Pi Mu and I.I.E.

ALVIN C. BRUSKA is an Industrial Engineering Tech of the Wafer Systems Department at Seagate Technology. Before making a complete circle back to IE, he was a Unit Manager and a Tech Manager at Seagate. Prior to Seagate he was an Industrial Engineer, a Methods Engineer, a supervisor, a Superintendent, a Business Unit Manager and a Plant Manager at Naval Systems Division of FMC Corporation. He holds a B.S. (1968) degree in Mechanical Engineering (Industrial Engineering option) from the University of Minnesota.

TIMBUR M. WULF is a Manager of the Wafer Systems Department at Seagate Technology. He received his B.S. in Electrical Engineering from Michigan Technological University and his M.BA. from the College of St. Thomas. He is a member of SEMI.

JENNIFER K. ROBINSON is a senior analyst for C2MS Productivity Solutions. Prior to C2MS, she worked as an independent consultant in the semiconductor industry. Her clients in this industry have included SEMATECH, Digital Equipment Corporation, Seagate Technology, and Siemens AG. Jennifer holds a B.S. (1989) degree in civil engineering from Duke University, an M.S. (1992) degree in Operations Research from the University of Texas at Austin and Ph.D. degree in Industrial Engineering from the University of Massachusetts at Amherst (1998). Her research interests center on factory productivity measurement and improvement.