Proceedings of the 2008 Winter Simulation Conference
S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler eds.

FRAMEWORK FOR EXECUTION LEVEL CAPACITY ALLOCATION DECISIONS FOR ASSEMBLY – TEST FACILITIES USING INTEGRATED OPTIMIZATION - SIMULATION MODELS

Shrikant Jarugumilli
Mengying Fu

Dept. of Industrial Engineering
502 Goldwater Center
Arizona State University
Tempe, AZ 85287, USA

Naiping Keng
Chad DeJong

Intel Corporation
5000 West Chandler Boulevard

Chandler, AZ 85226, USA

Ronald Askin
John Fowler

Dept. of Industrial Engineering
502 Goldwater Center
Arizona State University
Tempe, AZ 85287, USA

ABSTRACT

We present a framework for capacity allocation decisions for Assembly-Test (A-T) facilities that is comprised of an optimization model and a simulation model. The optimization and simulation models are used iteratively until a feasible and profitable capacity plan is generated. The models communicate using an automated feedback loop and at each iteration the model parameters are adjusted. We describe the role of the optimization model, the simulation model and the feedback loop. Once the capacity plan is generated, it is passed down to the shop-floor for implementation. Hence, decision makers can develop accurate and more profitable execution level capacity plans using the integrated model which utilizes both optimization and simulation models. In this paper, we focus on the optimization model for capacity planning for the entire A-T facility at the individual equipment (resource) level for a two-week planning period and briefly discuss the simulation and the adjustment model.

1 INTRODUCTION AND PROBLEM DESCRIPTION

Semiconductor manufacturing involves a series of complicated processes which span several weeks. The industry is characterized by short product life cycles and high equipment costs. Most of the companies in this industry use the same type of equipment and have similar processes. Hence, companies have to constantly innovate on the silicon technology and utilize their existing resource capacity efficiently in order to survive the fierce competition. Allocating capacity to the already present equipment in an optimal or near optimal manner is challenging due to the complicated process flow, high product mix, and complex bill of material structure which is specific to the industry.

The semiconductor manufacturing process can be broadly classified into the following stages: Fabrication, Sort, Assembly and Test. On the wafer fabrication side, the integrated circuits are fabricated on silicon wafers by growing films of material with different electric characteristics and patterning using photolithography and etching processes. At Sort, the integrated circuits (dies) on each wafer are tested for their performance characteristics e.g., speed and power consumption. At the Assembly-Test facilities the individual dies are packaged and tested before they are shipped to the customer.

The focus of this work is to present an initial optimization model for the capacity allocation decisions for the Assembly-Test plants. A-T plants consist of a series of operations which share overlapping resources, i.e. multiple operations might be performed on a single resource. There are often several machines which can be utilized to perform a given operation.

We also propose a conceptual framework for short-term capacity allocation decisions which consists of an optimization model, a simulation model and an adjustment model. The optimization model is used to generate a capacity plan for the factory in two-hour time periods for a couple of shifts and a shift-wise plan for subsequent shifts for a two week planning horizon. Even modeling only the bottleneck stages in the facilities, the optimization problem is very large in terms of number of variables (both binary and real) and constraints. Also, the modeling complexities include other factory specific rules and conditions such as: limited equipment, multiple levels of product mapping (or BOM Structure), product flow complexity (many to many relationships), high product and volume mix, shifting bottlenecks, setups (conversions), qualification requirements, among others. The solution of the optimization model generates a short-term capacity plan which is passed on to the simulation model using a adjustment model.

The simulation tool is used to model the entire factory including the non-bottleneck operations to get high fidelity estimate of the performance measures for the plan generated by the optimization model. The adjustment model

978-1-4244-2708-6/08/$25.00 ©2008 IEEE

passes the information between the models and also resets the parameters or constraints as required.

While implementing execution level (i.e. short-term) capacity plans it is important that the software tool performance (in terms of the run time) is low so that there is not a big difference in the shop-floor status during the run of the tool itself.

While allocating capacity, the planners need to ensure the availability of the raw material. This problem gets a bit complicated because of the high product mix and the complicated bill of material structure. Hence, the planning needs to be carried out on multiple time horizons. By making sure that these time horizons are small we can ensure real-time control and implementation of shop floor operations.

Capacity requirements planning is based on long-term demand forecasts of the various final products. It is interesting that these long term planning assumptions typically change and are not valid at the execution level on the shop floor. This causes a gap in what is being produced versus what needs to be actually produced. In order to narrow this gap, planners need to have real-time control over the shop floor operations to make real-time decisions, which is a big challenge considering the huge product mix which ships out of its factories, the process and the product complexities. Often it is seen that these constraints may lead to underload or overload situations for particular time periods resulting in the loss of capacity or creation of a new bottleneck. It is the goal of this research to solve this problem of assigning the right capacity to all equipment on the shop-floor in order to satisfy the demand and have shorter cycle times for the final products.

Based on this problem definition, the short-term capacity allocation problem is considered. Our approach includes modeling of major machines at the various stages in two hour time intervals for a two week planning period under operating rules which are specific to Intel's A-T factories.

2 LITERATURE REVIEW

Over the years, spreadsheet based models (Occhino 2000) for industrial capacity planning have been very popular since they are easy to use and provide reasonable results in a short amount of time. Recently, other techniques including simulation (Chou and Everton 1996, etc.), and mathematical programming (Karabuk and Wu 2002, Uribe et al. 2003, etc.) have been used as advanced planning tools in industry.

In their work, Dillenberger et al. (1993) present a production allocation model for machines and time periods in a single-level, multi-capacitated production environment with initial setups and setups which are partially dependent on production sequence. In this particular work, they consider product-dependent setups for a large scale problem instances. They extend this work in Dillenberger et al. (1994), where they solve the capacity requirements at the execution level. In their model they take into account considerable shop floor information including: machine availability, minor and major setups, storable and non-storable resources and they account for costs associated with overlapping setups. Though this model comes very close to our problem, this model uses an excessively large number of binary variables and only models a single stage of operations.

Quadt and Khun (2005) describe a conceptual framework for lot-sizing and scheduling of flexible flow shops. The problem described in their research is a large time bucket lot sizing model that attempts to effectively utilize the bottleneck (i.e. a single stage model). They also capture the concept of setup state carryover but do not model the set up time carry over. Other complexities such as product substitution and limited enabler constraints are missing. In this paper, only the framework is presented and the actual solution and results are not presented.

3 INTEGRATED MODEL FRAMEWORK

The capacity allocation tool we are designing will consist of an optimization model, a simulation model and a feedback mechanism as shown in figure 1. In this section we will describe the scope of each of these models.

3.1 Optimization Model

The optimization model mainly handles the deterministic parameters representing the factory operations, e.g. processing times, setup times, etc. The various bottleneck operations are modeled in the optimization model and the non-bottleneck operations are considered as additional time delays between the modeled operations.

The input parameters for the optimization model include: demand for each product, the product flow and the process flow, complete bill of material, setup times, processing times, throughput times for non-bottleneck operations, transfer times, number of machines at bottleneck stages, yields, potential bottlenecks, machine availability, WIP limits, qualification matrix, product priority, material availability, factory calendar among others. The model generates the capacity allocation plan for each machine for a specific operation at each stage in two hour time periods for the first two shifts and in shiftly time periods for the next thirteen days. The model also accounts for sequence dependent setups which last for a couple of minutes to a few hours and reports the volume of the unmet demand. The objective function of the optimization model is to minimize the cycle time, the work in process and the missed demand and is subject to inventory balance, capacity and setup constraints. The optimization model was developed using CPLEX 10.2.

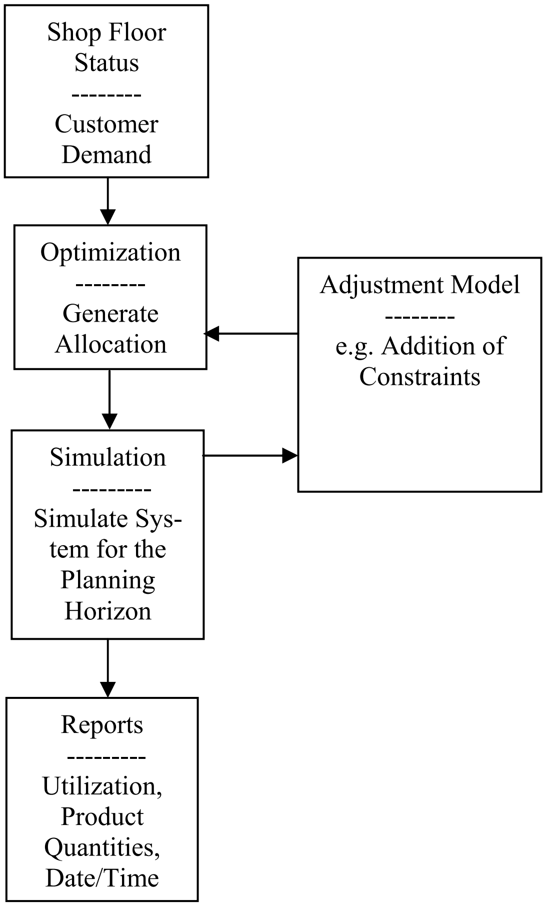

Figure 1: Integrated Framework Model

Decision Variables
Continuous Variables:

B_{pt}**:** Back order volume of the product family ‘p’ at the end of period ‘t’ at the last stage ‘N’

X^n_{ptm}**:** Production volume of product family ‘p’ during period ‘t’ on machine ‘m’ at stage ‘n’

I^n_{pt}**:** Inventory volume of product family ‘p’ at the end of period ‘t’ at (after) stage ‘n’

W^n_{ptm}**:** Setup time for product family ‘p’ on machine ‘m’ in stage ‘n’ in period ‘t’

R^n_{ptm}**:** Quantity of product ‘p’ used to meet demand for product ‘q’ in period ‘t’ at stage ‘N’

L^n_{ptm}**:** Cumulative setup time for product family ‘p’ on machine ‘m’ in stage ‘n’ at the end of period ‘t’

Binary Variables (Setup Tracking Variables):

Y^n_{ptm} 1: if setup operation for product family ‘p’ on ma chine ‘m’ stage ‘n’ is finished in period ‘t’
0: otherwise

Z^n_{ptm} 1: if product family ‘p’ can be produced at end of period ‘t’ and can be produced in ‘t+1’ without incurring a setup
0: otherwise

U^n_{ptm} 1: if setup operation for product family ‘p’ is going on at the end of the period ‘t’ and will continue at the beginning of period ‘t+1’ on machine ‘m’ at stage ‘n’ at the end of period ‘t’
0: otherwise

Parameters:
M: number of parallel machines at each stage
N: number of stages
P: number of product families
T: number of periods
κ: a big number

B_{p0}: Initial back order volume of product family ‘p’ at (after) stage ‘N’

I^n_{p0}: Initial inventory of product family ‘p’ at (after) stage ‘n’

h^n_p: Inventory holding cost of product family ‘p’ at stage ‘n’

o_p : Back order cost of product family ‘p’ at stage ‘N’

C: Capacity of machine per time period (in machine hours available per period)

D_{pt}: Demand volume of product family ‘p’ in period ‘t’ at the last stage ‘N’

s^n_{pm} : set up time for set up product family ‘p’ on machine ‘m’ in stage ‘n’

t^n_{pm} : unit processing time of product family ‘p’ on machine ‘m’ in stage ‘n’

c^n_{pqt} : unit substitution cost for using product ‘p’ to satisfy ‘q’ demand in time ‘t’ at stage ‘N’ at time ‘t’. Note: we assume all cost parameters for product ‘p’ are strictly greater than ‘q’

ζ^n_{pq} 1 : if product family ‘p’ can be used to satisfy demand for product family ‘q’ at stage N
0 : Otherwise

g_{ql} : the number of die type ‘q’ which gets into package assembly for product ‘l’

Objective Function:

$$\min(\sum_{p,t,n} h_p^n I_{pt}^n + \sum_{p,t} o_p B_{pt} + \sum_{\substack{p,q,t\\ p>q}} c_{pqt}^n R_{pqt}^n) \qquad (1)$$

Subject to:

$$I_{pt-1}^n + \sum_m X_{ptm}^n - \sum_m X_{ptm}^{n+1} - \sum_q \zeta_{pq}^n R_{pqt}^n + \sum_q \zeta_{qp}^n R_{qpt}^n = I_{pt}^n$$

for all p, t, m, n<N (2)

$$I_{pt-1}^n - B_{pt-1} + \sum_m X_{ptm}^n - D_{pt} - \sum_q \zeta_{pq}^n R_{pqt}^n + \sum_q \zeta_{qp}^n R_{qpt}^n = I_{pt}^n - B_{pt}$$

for all p, t, m (3)

$$\sum_m X_{ptm}^n \le \sum_{k=1}^{k*} I_{pt-1m}^{n-k} \qquad (4)$$

$$X_{ptm}^n \le \kappa (Z_{pt-1m}^n + Y_{ptm}^n) \text{ for all p, t, m, n} \qquad (5)$$

$$\sum_p t_{pm}^n X_{ptm}^n + \sum_p W_{ptm}^n \le C_t \text{ for all t, m, n} \qquad (6)$$

$$s_{pm}^n * Y_{ptm}^n \le L_{pt-1m}^n + W_{ptm}^n \qquad (7)$$

$$\sum_p Z_{ptm}^n + \sum_p U_{ptm}^n \le 1 \text{ for all t, m, n} \qquad (8)$$

$$W_{ptm}^n \le C_t (Y_{ptm}^n + U_{ptm}^n) \text{ for all p, t, m, n} \qquad (9)$$

$$Z_{ptm}^n \le 1 - \sum_{\substack{q=1\\ q\ne p}}^{p} U_{qtm}^n \text{ for all p, t, m, n} \qquad (10)$$

$$Z_{ptm}^n \le Z_{pt-1m}^n + Y_{ptm}^n - Y_{qtm}^n \text{ for all p, t, m, n, q;}$$

$$q \ne p \qquad (11)$$

$$L_{ptm}^n \le L_{pt-1m}^n + W_{ptm}^n \text{ for all p, t, m, n} \qquad (12)$$

$$L_{ptm}^n \le U_{ptm}^n * s_{pm}^n \text{ for all p, t, m, n} \qquad (13)$$

$$\sum_{p,t,m,n} W_{ptm}^n \le \mathit{fixednumber} \text{ for all p, t, m, n} \qquad (14)$$

$$\sum_{p,t,m,n} Y_{ptm}^n \le \mathit{fixednumber} \text{ for all p, t, m, n} \qquad (15)$$

$$X_{ptm}^n + X_{pt+1m}^n \ge \mathit{const} \text{ for all p, t, m, n} \qquad (16)$$

$$I_{pt}^n \ge \text{ lower limit for all p, t, n} \qquad (17)$$

$$R_{pqt}^n \le \text{ limit on substitution for all p, t, q} \qquad (18)$$

$$U_{ptm}^n = 0 \text{ for t =6,12 etc… end of shift}$$

for all p, m, n (19)

The objective function (1), minimizes the sum of the inventory holding costs across all the operations and products, the back order costs at the last stage, and the product substitution costs, which are the costs incurred due to conversion to a lesser value product. The last two terms were added to the objective function to distinguish between various alternate optimal solutions by incorporating a penalty function which decides the machine-job pair at the lowest cost. This is subject to various inventory balance, capacity and setup constraints which are described below.

The constraint sets (2) and (3) are the inventory balance constraints which ensure the "conservation of material" principle. Constraint set (3) is specifically for the last stage while constraint set (2) is for all other operations excluding the last stage. Constraint set (4) ensures that the conservation of material is ensured when the product can be processed at multiple stages within a specific time period, i.e. the processing time for several operations is less than the length of the time period itself, k* represents the number of stages the product can flow thorough within a given time period. The capacity constraint set (5) ensure that a lot manufactured in a given period only if the machine is setup for the product; this could be possible as a result of setup state carry over from the previous period or a completion of the setup operation in a given period. The value of 'κ' is defined as the maximum production that can be done on a resource in a given time period. Constraint set (6) ensures that the amount of time available in a time period is either utilized in making a product or is utilized for making setups or is accounted as idle time. The value of 'C_t' for the first shift will be 2 hours, followed by the subsequent shifts which will be 12 hours. Constraint set (7) ensures that if a setup is completed in a given time period, then the total cumulative time to complete the setup across all the previous periods is less than or equal to the setup time for the specific product and machine combination. Constraint set (8) ensures that at the end of the time period the machine is either carrying a setup state to the next time period or is undergoing a setup for a particular product. Constraint set (9) ensures that the setup time cannot be initialized for a product on a particular machine at a particular stage in a given time period unless we complete the setup or the setup operation is in progress during the time period. Constraint sets (10) and (11) ensure that the value of the setup carryover variable is reset depending on if a setup operation is complete or is in progress. Constraint set (12) adds the setup time incurred in a given time period to the cumulative setup time from the previous periods. Constraint set (13) resets the value of the cumulative setup time if the setup is not active at the end of a given time period. Constraint sets (14) and (15) limits on setup time and number of setups. Constraint set (16) ensure that between any two setup operations, a minimum number of lots are produced. Constraint set (17) ensures that minimum inventory of products is maintained at all stages at all times. Constraint set (18) sets the limit on the number of products of type 'p' which can be used for satisfying the demand of product 'q'. Constraint set (19) ensures that the setup operations have to be fully completed before the end of a shift.

3.2 Simulation Model

The role of the simulation model is to lay out the schedule as per the optimization model's output in a deterministic setting. The input to the simulation model includes all the parameters used for the optimization along with the output generated by the optimization model and the processing times for the non-bottleneck operations. The simulation model is mainly used to get estimates of various performance measures such as: cycle time, tool utilization, throughput time for non-bottleneck operations based on the current solution given by the optimization model. The simulation model is a more detailed model which consists of the various non-bottleneck operations and other factory parameters which were not captured in the optimization model. The simulation model was built in AutoSched AP. (AutoSched AP User Manual v 8.0, 2004)

3.3 Adjustment Model

The adjustment model is used to transfer and adjust data between the optimization model and the simulation model. Initially, the adjustment model sends the schedule generated by the optimization model to the simulation model. Once the simulation runs to completion, various statistics are collected and checked for feasibility and practicality for implementation. Based on the simulation results some parameters and constraints might need to be modified. This information is sent back to the optimization model.

These iterations continue till the desired factory metrics are achieved or the factory metrics converge in a couple of successive iterations. Also, the user can predefine the number of iterations between the models before the results are accepted.

4 CONCLUSIONS AND RESUTLS

The main objective of the paper was to present the framework for capacity planning using an integrated optimization and simulation models. In this paper, we have presented an optimization model which considers the setup times varying from a couple of minutes to a few hours and models the capacity lost due to setups accurately. Also, we have presented a conceptual framework for development of a capacity planning tool comprising of an optimization model, a simulation model and a feedback loop. We believe accurate and more profitable execution level capacity plans can be generated using an integrated model which utilizes both simulation and optimization models. We intend to present our initial results at the conference in December.

ACKNOWLEDGMENTS

We would like to thank the Intel Research Council for funding this project.

REFERENCES

Brooks Automation, Inc. 2004. AutoSched AP User's Guide v 8.0 Chelmsford, MA.

Chou, W., and J. Everton. 1996. Capacity planning for development wafer fab expansion. In *Proceedings of the 1996 IEEE/SEMI Advanced Semiconductor Manufacturing Conference,* Cambridge, MA, 17-22.

Dillenberger, C., L. F. Escudero, A. Wollensak, and W. Zhang. 1993. *On solving a large-scale resource allocation problem in production planning*. In: Fandel, G., Gulledge, T., and Jones A. (Eds), Operations Research in Production Planning and Control, Springer, Berlin, Germany, 105-119.

Dillenberger, C., L. F. Escudero, A. Wollensak, and W. Zhang. 1994. On practical resource allocation for production planning and scheduling with period overlapping setups. *European Journal of Operational Research* 75:275-286.

Karabuk, S., and S. D. Wu. 2002. Decentralizing semiconductor capacity planning via internal market coordination. *Inst. Ind. Eng. Trans*. 34:743–759.

Occhino, T. J. 2000. Capacity planning model: the important inputs, formulas, and benefits. In *Proceedings IEEE/SEMI Advanced Semiconductor Manufacturing Conference*, 455-458.

Quadt, D., and H. Kuhn. 2005. Conceptual framework for lot-sizing and scheduling of flexible flow lines. *International Journal of Production Research* 43:2291–2308.

Uribe, A. M., J. K. Cochran, and D. L. Shunk. 2003. Two-stage simulation optimization for agile manufacturing capacity planning. *International Journal of Production Research* 41:1181–1197.

AUTHOR BIOGRAPHIES

SHRIKANT JARUGUMILLI is a graduate student currently working as a graduate research associate in the Industrial Engineering Department at Arizona State University. He did a summer internship with Intel Corporation in Summer 2007. He has a MS degree in Engineering Management from University of Missouri, Rolla and a B.E. (Industrial Engineering and Management) from Visvesvaraya Technological University, India. His research interests include modeling and analysis of manufacturing systems. He is also the Vice-President of the ASU Chapter of INFORMS. His e-mail address is <sjarugumilli@gmail.com>.

MENGYING FU is currently pursuing the Doctoral degree in Department of Industrial Engineering, Arizona State University, Arizona State, US. Her academic interests include production scheduling and integer optimization. She received the B.S. in Industrial Engineering from Tsinghua University, Beijing, China. Email: <Mengying.Fu@asu.edu>.

NAIPING KENG is a Principal Engineer at Intel Corporation, where he has designed and implemented various planning and scheduling tools for fab, sort, and assembly and test for the past 19 years. He is the chief architect of several production planning and execution tools for assembly and test (A/T) factories. His current projects and interests include machine setup optimization, shiftly full factory capacity allocation and alignment, automatic generation of capacity and material feasible production goals, and business process design. Dr. Keng received a Ph.D. in Computer Science from Southern Methodist University in 1989.

CHAD DEJONG is a Systems Engineer with the Operations Decision Support Technologies group at Intel Corporation. He is primarily responsible for the design, layout, and modeling of automated material handling systems. He earned a B.S. in Industrial and Operations Engineering from the University of Michigan. He also earned an M.S. in Industrial and Systems Engineering from Georgia Institute of Technology, and completed the Management of Technology certificate program. His previous professional background includes hospital and health care systems, and automotive component manufacturing. Current research and professional interests in the semiconductor industry are in whole factory capacity and operations modeling, supply chain simulation, model communication, model design and execution time reduction, and the statistical validation of modeling tools. His email address is <chad.d.dejong@intel.com>.

RONALD ASKIN, Ph.D., is Department Chair and Professor of Industrial Engineering at Arizona State University. He received his PhD from Georgia Institute of Technology and has 30 years of experience in the development, teaching and application of methods for production systems analysis. He is a Fellow of IIE and has published extensively. His list of awards includes a National Science Foundation Presidential Young Investigator Award, the Shingo Prize for Excellence in Manufacturing Research, IIE Joint Publishers Book of the Year Award, and the *IIE Transactions* Development and Applications Award.

JOHN FOWLER is a Professor of Industrial Engineering at Arizona State University and was the Center Director for the Factory Operations Research Center that was jointly funded by International SEMATECH and the Semiconductor Research Corporation. His research interests include modeling, analysis, and control of semiconductor manufacturing systems. Dr. Fowler is a member of IIE, INFORMS, and SCS. He is an Area Editor for SIMULATION: Transactions of the Society for Modeling and Simulation International and an Associate Editor of IEEE Transactions on Semiconductor Manufacturing. He is an IIE Fellow and is on the Winter Simulation Conference Board of Directors.