ISSN: 0974-5823 Vol. 7 No. 8 August, 2022

International Journal of Mechanical Engineering

Model to Evaluate Production Performance under Capacity Fluctuation of High Throughput Equipment in Wafer Fabrication

Ying-Mei Tu [1]

[1] Department of Industrial Management, Chung Hua University, Taiwan.

Abstract –

In order to get more profit from advanced technology, 300mm small and medium sized wafer fabs introduced high-throughput equipment to enhance the wafer processing technology due to fab space constraint. There will have a huge capacity fluctuation risk of high-throughput equipment due to the characteristic of small quantity and dedicated process constraint. Therefore, how to maintain production performance, such as higher throughput and shorter product cycle time, is a considerable challenge for managers.

The major task of this work is to develop a performance evaluation system under capacity fluctuation condition. Managers can easily figure out the reduced outputs and the increased production cycle times of product when the production capacity fluctuation happened. A fab simulation model will be established to realize the effect of various factors that cause capacity fluctuations upon production performance. In addition, a mathematical model will be constructed as well by using the queuing theory to evaluate the impacts upon waiting time by the capacity fluctuation factors. By the same token, Little's Law will be used to calculate the impacts upon outputs simultaneously.

Index Terms : *Wafer fabrication, High throughput output equipment, Capacity fluctuation, Queueing theory.*

INTRODUCTION

As everyone knows that due to rapid upgrading of wafer processing technology and more profits of advanced technology, fabs attend to manufacture products by using advanced technology. Therefore, the current 300mm small and medium sized wafer fabs are suffering a common dilemma, which is the production scale of this kind of factory is not only sizable but also mixed with different technology generations. Thus, high-throughput equipment was introduced due to fab space restriction. The characteristic of high-throughput equipment includes small quantity and dedicated process constraint. It means there will have a huge capacity fluctuation occurred of these machines. Under such circumstances, how to maintain production performance, such as throughput and product cycle time, is a considerable challenge for managers. Consequently, the planning and management of high-throughput equipment under small-scale production must be discussed and studied.

Most of the previous research on production capacity focused on capacity assessment, capacity optimization, bottleneck prediction, capacity planning and management, but little attention was paid to the prediction and control of capacity fluctuations (Chen, *et al*. 2013). In capacity planning, the historical average availability and process time data of the machine are almost used to estimate whether the machine capacity meets the demand (Geng, *et al*. 2009). Generally, the target can be achieved when the machine availability is sable but it is difficult to achieve the predetermined goal if the machine availability is more fluctuant. Due to the large quantity of same type machines, the equipment availability of large scale fab will be more stable, even if some machines are not in production conditions which include machine down, prevent maintenance or taken for engineering purpose. Nonetheless, the fluctuation of capacity availability of high-throughput equipment under

small-scale production will seriously affect the overall production activities. In order to solve the availability fluctuation issue, managers will prepare sufficient WIP to avoid capacity loss of downstream machines in the past. However, for advanced processes, almost every station has a time constraint, so it is completely infeasible to reserve high WIP(Working In Process). Therefore, managers must understand the relationship between the stability of production capacity and production performance, and set the production capacity stability of equipment under acceptable production performance, so as to ensure smooth production and maintain better production performance.

LITERATURE REVIEW

In terms of capacity planning, it generally refers to medium and long-term capacity planning. In the past, many scholars had proposed various capacity planning models for the semiconductor manufacturing industry, and most of them are based on queueing network, mathematical planning or mean value analysis to estimate the numbers of machines required for each machine group (Iwata et al., 2003; Walid and Gharbi, 2002; Chou and You, 2001; Romauch & Hartl 2016; Ortner 2008; Wood et al., 1994). Nowadays, the purpose of capacity planning is not only to meet future demands, but also to reduce production time, minimize production costs, and improve customer satisfaction. However, these studies do not take future demand uncertainty and equipment investment risks into account. In the high demand uncertainty and investment risk industry such as semiconductor manufacturing, this kind of planning results are bound to be insufficient. Therefore, some studies have considered the uncertainty of future demand in capacity planning (Chien et al., 2018; Swaminathan, 2000; Hood et al., 2003; Chou et al., 2007). Chien *et al.* (2018) proposed an uncertain multi-objective decision strategy framework based on uncertainty theory for capacity expansion and migration problem to minimize the potential loss of capacity oversupply or shortage under uncertain environment. Hood *et al.* (2003) used the method of integer programming and modeled the future demand change trend into several "scenarios" with different probability of occurrence, so as to plan the best product mix under the condition of uncertain future demand. Unfortunately, this study did not take technological changes into consideration. More importantly, it did not pay much attention to the calculation and formulation of the probability of occurrence of "scenarios", and the planning results calculated by using scenarios were nothing more than using the probability of occurrence of various scenarios. Another technique that is often used in highly uncertain environments is to explore them through system simulations (Hung and Leachman, 1996; Mula et al., 2006), which has the advantage of incorporating uncertainties in the environment. The factors of uncertainty are set in the simulation system in the form of random variables, and then the optimal decision-making under the influence of these uncertain factors is obtained. However, the semiconductor industry is associated with extremely complex processes and many uncertain factors, planning using system simulation will be a time-consuming, labor-intensive and costly task (Uzsoy et al., 1992). Based on these researches, it can be found that the issues of capacity fluctuation are rarely considered, and the supply of production capacity is presented in a static and average way. As mentioned above, since most of production process could share machines with the same function in the past, the numbers of machines in the workstation was quite large, and the production capacity was more stable. Thus, the issues of capacity fluctuation need not pay more attention in capacity planning accordingly. Nevertheless, the current situation has changed, and many workstations have obvious production capacity fluctuations, which significantly increase the difficulty to achieve the original production goals. Apparently, the problem of production capacity fluctuations must be considered in production capacity planning without a doubt.

ESTIMATION MODELS

In general, throughput and product cycle time are the major production performance indicators in wafer fabrication. To understand the impact of high-throughput equipment on production performance, the equipment was first divided into three different categories: bottleneck machines, non-bottleneck machine in front of bottleneck machine, and non-bottleneck machine behind bottleneck machine. Since wafer fabrication is a re-entrant process, the tool defined in this study is located before or after the bottleneck tool, which means the next or previous station is the bottleneck tool. There are two major factors influence the fluctuations in production capacity, the number of machines and the stability of individual machines. The more numbers of machines, the better stability of workstation productivity, that means it will have less impact on the overall

workstation productivity when one machines is down or repaired. Based on the simulation results, it reveals that MTTR (Mean Time To Repair) has a significant impact on the capacity stability of individual machine. And, the longer MTTR will cause the worse the production capacity stability. Therefore, the productivity stability of the workstation will be affected by the numbers of machines and the MTTR of individual machine. Furthermore, they will affect the overall production performance as well. In order to well manage the impacts from capacity fluctuation of high-throughput workstations, a throughput and cycle time evaluation model is established to provide managers with the impact information of production capacity fluctuations. Through this information, managers can understand the influences on throughput and product cycle time when the production capacity fluctuates. Moreover, manager can find out how much capacity should be increased or which factors should be adjusted if the original performance should be maintained.

From the perspective of wafer manufacturing industry, there are three workstations are treated as high-throughput machines include Dry Strip, Ion Implanter and Photolithograph, which are all sequential processing stations. Usually, the influence on subsequent workstations will be weakened gradually to the downstream stations. Accordingly, this research will focus on the high-throughput workstation and the following two workstations into the evaluation system. Basically, the product cycle time evaluation system will start with the high-throughput workstation as the first station and build two stations after that. Furthermore, the bottleneck station will be added into the evaluation segment in case the high-throughput workstation is behind the bottleneck station, there will be total four stations in such a situation. Based on this concept, the evaluation model of the production segment will include the high-throughput workstation and two downstream stations or plus the bottleneck workstation before the high-throughput workstation. This evaluation system will be divided into two parts: the impact on product cycle time and throughput separately.

EVALUATION OF THE IMPACT ON CYCLE TIME

In the evaluation system, each workstation is a queueing system connected with the upstream and downstream. The upstream output will affect the production of the workstation, and the throughput of the workstation will affect the downstream operation as well. In general, the cycle time of a product includes processing time and waiting time. And, the processing time of a product is more stable than waiting time. As the waiting time is also affected by environmental factors, the difficulty in predicting the production cycle time is mainly caused by the variation of the waiting time. Thus, the GI/G/m model is applied to calculate the impact on the waiting time of the production segment due to the capacity fluctuation of high-throughput workstation. In queueing network, the service rate of the previous queueing system will be directly equal to the arrival rate of the next queueing system. The capacity fluctuation of the workstation will affect the waiting time of the downstream station and this effect will continue to be deferred and gradually weakened. Hence, the production segment constructed in the evaluation model only consists three to four stations including the high-throughput workstation. In addition, for the situation of downtime, PM and engineering purposes that induces the workstation cannot be used for production, the modification of the queueing system proposed by Tu & Chen 2009 will be applied to add the impact of capacity fluctuation on the waiting time. Based on the influence of waiting time by process, the processing of wafer fabrication can be roughly divided into three types: sequential processing, batch processing and un-batch sequential processing. The model is calculated as follows:

(A) Notation

The following data were required for the waiting time model:

λ_{fj}	mean arrival rate of product *f* at workstation *j*
λ_{of}	demand rate of product *f*
V_{fj}	the reentrant numbers of product *f* at workstation *j*
b	batch size
τ_j	mean service time of workstation *j*
$\tau_j^{'}$	the adjusted mean service time of workstation *j*

τ_{fk}	the service time of product *f* at step *k*
s_{fk}	the workstation visited by product *f* at step *k*
t_f	total number of operation steps of product *f*
C^2_{sfk}	the SCV of service time of product *f* at step *k*
$C^2_{s_j}$	the SCV of service time of workstation *j*
$C^{2'}_{s_j}$	the adjusted SCV of service time of workstation *j*
$MTTR_{jl}$	mean Time To Repair of machine *l* at workstation *j*
$MTBF_{jl}$	mean Time Between Failure of machine *l* at workstation *j*
$\hat{\rho}_b$	utilization of fictitious batch machine
$\hat{\lambda}_b$	mean arrival rate of fictitious batch machine
$\hat{\tau}'_b$	adjusted mean service time of fictitious batch machine
$\hat{C}^{2'}_{s_b}$	the adjusted SCV of service time of fictitious batch machine
m_{j-1}	total numbers of machines at the batch workstation j-1 which is the upstream of workstation j
ρ_{j-1}	utilization of the batch workstation j-1 which is the upstream of workstation j
τ'_{j-1}	adjusted mean service time of the batch workstation j-1 which is the upstream of workstation j
$C^2_{a_j}$	the SCV of inter-arrival time of workstation *j*
$EW_{j(Q)}$	the expected waiting time of product by batch type
$EW_{j(B)}$	the expected waiting time of any product p to form a batch
P	The numbers of product type
EW_j	expected waiting time of workstation *j*
EW_{jQ}	expected waiting time of batch entity in front of workstation *j*
EW_{jU}	expected waiting time of un-batch entity in front of workstation *j*
ΔCT	the influence of product cycle time by capacity fluctuation
NEW_i	the waiting time of process segment after capacity fluctuation
OEW_i	the waiting time of process segment before capacity fluctuation
N	numbers of reentry
n	numbers of workstation
$\Delta output$	the influence of output by capacity fluctuation
WIP	working in Process within the segment of capacity fluctuation
PT_i	average process time of workstation i within the segment of capacity fluctuation

(B) Calculation of GI/G/m Parameters

There are some parameters should be calculated in GI/G/m queueing model including arrival rate (λ), average service time (τ), squared coefficient of variation (SCV) of inter-arrival time (C_a^2) and SCV of service time

(C_s^2). Equation (1) is the arrival rate aggregating arrival rates of individual product into the mean arrival rate (λ_{fj}) of the workstation. Equation (2) and (3) are the service time and squared coefficient of variation (SCV) of service for individual products which are aggregated into the mean service time (τ_j) and SCV of service time ($C_{S_j}^2$)of the workstation separately. Based on the previous study [18], the machine failure behaviors have been considered in GI/G/m model and the mean service time and SCV of service time were modified in Equation (4) and (5). Furthermore, there are some parameters should be modified for the un-batch process which is the sequential process in the downstream of batch process. The parameters of fictitious batch machine in un-batch process can be referred to Tu [19]. The parameters calculations are as follows.

$$\lambda_{fj} = \frac{V_{fj}\lambda_{of}}{b} \tag{1}$$

$$\tau_j = \frac{\sum_f \sum_{\{k|s_{fk}=j\}}^{t_f} \lambda_{of}\tau_{fk}}{\lambda_j} \tag{2}$$

$$C_{s_j}^2 = \frac{\sum_f \sum_{\{k|s_{fk}=j\}}^{t_f} \lambda_{of}\tau_{fk}^2(C_{sfk}^2+1)}{\lambda_j\tau_j^{'2}} - 1 \tag{3}$$

$$\tau_j^{'} = \frac{\lambda_j\tau_j + \sum_{l=1}^{m_j}\frac{MTTR_{jl}}{MTBF_{jl}+MTTR_{jl}}}{\lambda_j + \sum_{l=1}^{m_j}\frac{1}{MTBF_{jl}+MTTR_{jl}}} \tag{4}$$

$$C_{s_j}^{2'} = \frac{\lambda_j\tau_j^2(C_{S_j}^2+1) + \sum_{l=1}^{m_j}\frac{MTTR_j^2}{MTBF_j+MTTR_j}(C_{d_{jl}}^2+1)}{(\lambda_j + \sum_{l=1}^{m_j}\frac{1}{MTBF_{jl}+MTTR_{jl}})\tau_j^{'2}} - 1 \tag{5}$$

$$\hat{\lambda}_b = \frac{m_{j-1}}{\tau_{j-1}^{'}} \times \rho_{j-1} \tag{6}$$

$$\hat{\tau}_b^{'} = \frac{\tau_j^{'} \times b}{m_j} \tag{7}$$

$$C_{s_b}^{2'} = \frac{Var(\hat{\tau}_b^{'})}{E(\hat{\tau}_b^{'})^2} = \frac{(\frac{b}{m_j})^2 Var(\tau_j^{'})}{(\frac{b}{m_j})^2 E(\tau_j^{'})^2} = C_{sj}^{2'} \tag{8}$$

$$C_{a_j}^2 = \alpha + \sum_{i=1}^{n}\beta C_{a_i}^2 \tag{9}$$

Where α and β are defined in Whitt [20].

(C) Expected waiting time calculation

After the calculation of various system parameters is completed, the expected waiting time of the product can be obtained. Based on the processing type, the factors of the product waiting time and the calculation logic will be different. Therefore, we must use different calculation methods according to their processing patterns. The calculation methods of expected waiting time are described as follows:

(a) Sequential process

$$EW_j = \phi(\lambda_j, C_{aj}^2, \tau_j^{'}, C_{sj}^{2'}, m_j) \times \frac{C_{aj}^2 + C_{sj}^{2'}}{2} \times \frac{\tau_j^{'}(\rho_j^{\sqrt{2m_j+1}-1})}{m_j(1-\rho_j)} \tag{10}$$

(please refer Whitt [20] to obtain the equations for f())

(b) Batch process

Besides the waiting for processing, products must also wait for batches in batch workstations. Therefore, the waiting time of the product must be added to this part of the batch time. Among them, the waiting time of the batched products waiting for processing is the same as the calculation logic of the sequential station, and the complete expected waiting time of the batch processing station can be calculated by the following formula:

$$EW_j = EW_{j(Q)} + EW_{j(B)} \quad (11)$$

$$EW_{j(Q)} = \phi(\lambda_j, C_{aj}^2, \tau_j^{'}, C_{sj}^{2'}, m_j) \times \frac{C_{aj}^2 + C_{sj}^{2'}}{2} \times \frac{\tau_j^{'}(\rho_j^{\sqrt{2m_j+1}-1})}{m_j(1-\rho_j)} \quad (12)$$

$$EW_{j(B)} = \sum_f \frac{\lambda_{fj}}{\lambda_j} \frac{(b-1)}{2\lambda_{fj}} = \frac{(b-1)}{2\lambda_j} \times p \quad (13)$$

$$\lambda_{fj} = \lambda_{of} \times V_{fj} \quad (14)$$

(c) Sequential process in un-batch

In the amended model, the concept of fiction machine is used to simplify the unbatched workstation after batch workstation into a GI/G/1 queueing subsystem. Similarly, in addition to the waiting time for batch products to be processed by the fiction machine, the complete product waiting time also includes the un-batching time. Therefore, the waiting time of un-batching should be calculated independently to grab the complete product waiting time. The calculation method is as follows:

$$EW_j = EW_{j(Q)} + EW_{j(U)} \quad (15)$$

$$EW_{j(Q)} = \left(\frac{C_{aj}^2 + \hat{C}_{sb}^2}{2}\right) \times \left(\frac{\hat{\tau}_b \times \hat{\rho}_b}{1-\hat{\rho}_b}\right) \quad (16)$$

$$\hat{\rho}_b = \hat{\lambda}_b \times \hat{\tau}_b \quad (17)$$

$$EW_{j(U)} = \frac{\hat{\tau}_b}{2} \quad (18)$$

(d) *Calculation of the* cycle *time impact*

Managers can select a suitable combination of workstations according to the location of the high-throughput workstation and estimate the waiting time of the product at the workstation according to the above waiting time calculation formulas, and the cycle time is the waiting time plus the processing time. Eventually, the processing time will not change because of capacity fluctuations, the impact of capacity fluctuations on cycle time is the difference in waiting time between the affected segment before and after the fluctuations multiplied by the number of reentries. Calculated as follows:

$$\Delta CT = \left(\sum_{i=1}^{n} NEW_i - \sum_{i=1}^{n} OEW_i\right) \times N \quad (19)$$

ESTIMATION OF THE IMPACT ON THROUGHPUT

In a production environment, the overall throughput will decrease under the situation that the throughput of the bottleneck workstation be affected. Due to the characteristic of capital intensive in wafer fabrication, bundle of workstations are quite close to the bottleneck. Bottleneck will shift randomly when the availability of these workstation fluctuates greatly. Based on the concepts, the impact on the overall throughput will

depend on the location of these high-throughput machines and whether it will instantly become a bottleneck due to capacity fluctuations. If the high-throughput workstation is a bottleneck, the production capacity fluctuation should have been taken into account in production plan. Overall throughput will not be affected if other non-bottleneck workstations can fully cooperate. Nevertheless, the throughput must have an impact on the bottleneck and then affect the overall throughput when the high-throughput workstation is non-bottleneck and before the bottleneck. When the high-throughput workstation is not a bottleneck and behind the bottleneck, the impact of overall throughput will depend on whether the high-throughput workstation will instantly become a bottleneck due to capacity fluctuations. Hence, the overall throughput is affected only when the high-throughput workstation is a non-bottleneck workstation under capacity fluctuation, and the magnitude of the impact will be estimated by Little's Law as follows:

$$\Delta Output = \frac{WIP}{\sum_{i=1}^{n}(NEW_i + PT_i)} - \frac{WIP}{\sum_{i=1}^{n}(OEW_i + PT_i)} \quad (20)$$

CONCLUSIONS

In small and medium fabs, the production performance will be affected seriously by the capacity fluctuations of high-throughput workstations. It is almost impossible to know the impact will be or how to set the reasonable production targets in advance. This study proposes an evaluation model to explore the impact of high-throughput workstation capacity fluctuations on production performance in small and medium-sized fabs. In this evaluation model, the queueing theory is applied to construct a mathematical model to calculate the relationship between the production capacity fluctuation factor and the waiting time, and further calculate the impact on the product cycle time. As for the impact on throughput, Little's Law is used to calculate. It is hoped that through the evaluation model, managers can understand the impact of high-throughput workstation capacity fluctuations on production performance, so that small and medium-sized semiconductor factories can make a better plan to avoid the performance loss when the high-throughput machines introduced by the rapid upgrade of process generations.

ACKNOWLEDGEMENT

The author would like to thank the National Science Council of the Republic of China for financially supporting this research under Contract No. MOST 109-2221-E-216-005

REFERENCES

[1] B. Chen, K. Liu, Z. Fan, "Research on the Capacity Fluctuation Analysis of Compressor Blade Production Line". International Asia Conference on Industrial Engineering and Management Innovation (IEMI2012) Proceedings. Springer, Berlin, Heidelberg, P765-775, 2013.

[2] N. Geng, Z. Jiang, F. Chen, "Stochastic programming based capacity planning for semiconductor wafer fab with uncertain demand and capacity." European Journal of Operational Research, 198(3), 899-908, 2009.

[3] Y. Iwata, K. Taji and H. Tamura, "Multi-objective capacity planning for agile semiconductor manufacturing." Production Planning & Control, 14(3), 244-254, 2003.

[4] A. K. Walid and A. Gharbi, "Capacity estimation of a multi-product unreliable production line." International Journal of Production Research, 40(18), 4815-4834, 2002.

[5] Y.C. Chou and R. C. You, "A Resource Portfolio Planning Methodology for Semiconductor Wafer Manufacturing, International Journal of Advanced Manufacturing Technology, 18, 12-19, 2001.

[6] M. Romauch & R. F. Hartl, "Capacity Planning for Cluster Tools in the Semiconductor Industry." arXiv preprint arXiv:1605.00914, 2016.

[7] A. M. Ortner, "Capacity planning for cluster tools - a mathematical model that reflects scheduling for parallel processing with two loadlocks." Master thesis, Alpen-Adria-University at Klagenfurt, Klangenfurt, Austria. 2008.

[8] S. C. Wood, S. Tripathi & F. Moghadam, "A generic model for cluster tool throughput time and capacity." In Advanced Semiconductor Manufacturing Conference and Workshop. 1994 IEEE/SEMI (pp. 194-199). IEEE, 1994.

[9] C. F. Chien, R. Dou & W. Fu, "Strategic capacity planning for smart production: Decision modeling under demand uncertainty." Applied Soft Computing, 68, 900-909, 2018.

[10] J. M. Swaminathan, "Tool capacity planning for semiconductor fabrication facilities under demand uncertainty." European Journal of Operational Research, 120(3), 545-558, 2000.

[11] S. J. Hood, S. Bermon & F. Barahona, "Capacity planning under demand uncertainty for semiconductor manufacturing." IEEE Transactions on Semiconductor Manufacturing, 16(2), 273-280, 2003.

[12] Y.C. Chou, C. T. Cheng, F. C. Yang, & Y. Y. Liang, "Evaluating alternative capacity strategies in semiconductor manufacturing under uncertain demand and price scenarios." International Journal of Production Economics, 105(2), 591-606, 2007.

[13] M. Cakanyıldırım and R. O. Roundy, "Optimal Capacity Expansion and Contraction under Demand Uncertainty." Working Paper, 2002.

[14] Y.F. Hung and R. C. Leachman, "A production planning methodology for semiconductor manufacturing based on iterative simulation and linear programming calculations." IEEE Transactions on Semiconductor Manufacturing, 9(2), 257-269, 1996.

[15] J. Mula, R. Poler, J. P. García-Sabater and F. C. Lario "Models for production planning under uncertainty: A review." International Journal of Production Economics, 103(1), 271-285, 2006.

[16] R. Uzsoy, C. Y. Lee, and L. A. Martin-Vega, "A Review of Production Planning and Scheduling Models in Semiconductor Industry Part I: System Characteristics, Performance Evaluation and Production Planning, IIE Transactions, 24(4), 47-60, 1992.

[17] L. C. Wang, P. C. Chu, and S. Y. Lin, "Impact of capacity fluctuation on throughput performance for semiconductor wafer fabrication." Robotics and Computer-Integrated Manufacturing55, 208-216, 2019.

[18] Y. M. Tu, and H. N. Chen, "Capacity planning with sequential two-level time constraints in the back-end process of wafer fabrication." International Journal of Production Research, 47(24), 6967-6979, 2009.

[19] Y. M. Tu, and C. L. Chen, "Model to determine the capacity of wafer fabrications for batch-serial processes with time constraints." International Journal of Production Research, vol. 49, no. 10, pp. 2907-2923, 2011.

[20] W. Whitt, "Approximations for the GI/G/m Queue. Production and Operations Management, 2(2), 114-161, 1993.